WAVES

In a purely mathematical approach to the phenomenology of *waves*, we might choose to start with the WAVE EQUATION, a differential equation describing the qualitative features of wave propagation in the same way that *SHM* is characterized by $\ddot{x} = -\omega^2 x$. The advantage of such an approach is that one gains confidence that any phenomenon that can be shown to obey the WAVE EQUATION will *automatically* exhibit *all* the characteristic properties of *wave motion*. This is a very economical way of looking at things.

Unfortunately, the phenomenology of wave motion is not very familiar to most beginners — at least not in the mathematical form we will need here; so in this instance I will adopt the approach used in most first year Physics textbooks for almost everything: I will *start with the answer* (the simplest *solution* to the WAVE EQUATION) and explore its *properties* before proceeding to show that it is indeed a solution of the WAVE EQUATION — or, for that matter, before explaining what the WAVE EQUATION *is*.

14.1 Wave Phenomena

We can visualize a vivid example for the sake of illustration: suppose the "amplitude" A is the height of the water's surface in the ocean (measured from $A = 0$ at "sea level") and x is the distance toward the East, in which direction waves are moving across the ocean's surface.[1] Now imagine that we stand on a skinny piling and watch what happens to the water level on its sides as the wave passes: it goes up and down at a regular frequency, executing *SHM* as a function of time. Next we stand at a big picture window in the port side of a submarine pointed East, partly submerged so that the wave is at the same level as the window; we take a flash photograph of the wave at a given instant and analyze the result: the wave looks instantaneously just like the graph of *SHM* except the horizontal axis is *distance* instead of *time*. These two images are displayed in Fig. 14.1.

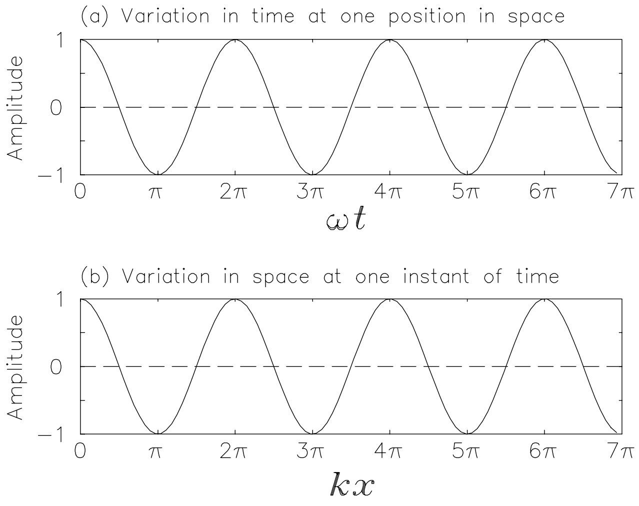

Figure 14.1 Two views of a wave.

14.1.1 Traveling Waves

How do we represent this behaviour mathematically? Well, A is a function of position $\vec{r}$ and time t: $A(\vec{r}, t)$. *At any fixed position $\vec{r}$, A oscillates in time* at a frequency ω. We can describe this statement mathematically by saying that the entire *time dependence* of A is contained in [the real part of] a factor $e^{-i\omega t}$ (that is, the amplitude *at any fixed position* obeys *SHM*).[2]

The oscillation with respect to *position $\vec{r}$ at any instant of time t* is given by the analo-

[1] Technically speaking, I couldn't have picked a *worse* example, since water waves do *not* behave like our idealized example — a cork in the water does *not* move straight up and down as a wave passes, but rather in a vertical *circle*. Nevertheless I will use the example for illustration because it is the most familiar sort of easily visualized wave for most people and you have to watch closely to notice the difference anyway!

[2] Note that $e^{+i\omega t}$ would have worked just as well, since the real part is the same as for $e^{-i\omega t}$. The choice of sign does matter, however, when we write down the *combined* time and space dependence in Eq. (4), which see.

gous factor $e^{i\vec{k}\cdot\vec{r}}$ where $\vec{k}$ is the *wave vector*;[3] it points in the *direction of propagation* of the wave and has a magnitude (called the "wavenumber") k given by

$$k = \frac{2\pi}{\lambda} \qquad (1)$$

where λ is the *wavelength*. Note the analogy between k and

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

where T is the *period* of the oscillation in time at a given point. You should think of λ as the "period in space."

We may simplify the above description by *choosing our coordinate system* so that the x axis is *in the direction of* $\vec{k}$, so that[4] $\vec{k}\cdot\vec{r} = k\,x$. Then the amplitude A no longer depends on y or z, only on x and t.

We are now ready to give a full description of the function describing this wave:

$$A(x,t) = A_0\, e^{ikx}\cdot e^{-i\omega t}$$

or, recalling the multiplicative property of the exponential function, $e^a \cdot e^b = e^{(a+b)}$,

$$A(x,t) = A_0\, e^{i(kx-\omega t)}. \qquad (3)$$

To achieve complete generality we can restore the vector version:

$$\boxed{A(x,t) = A_0\, e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)}} \qquad (4)$$

This is the preferred form for a general description of a PLANE WAVE, but for present purposes the scalar version (3) suffices. Using Eqs. (1) and (2) we can also write the plane wave function in the form

$$A(x,t) = A_0\, \exp\left[2\pi i\left(\frac{x}{\lambda} - \frac{t}{T}\right)\right] \qquad (5)$$

but you should strive to become completely comfortable with k and ω — we will be seeing a lot of them in Physics!

14.1.2 Speed of Propagation

Neither of the images in Fig. 14.1 captures the most important qualitative feature of the wave: namely, that it *propagates* — *i.e.* moves steadily along in the direction of $\vec{k}$. If we were to let the *snapshot* in Fig. 14.1b become a *movie*, so that the time dependence could be seen vividly, what we would see would be the same wave pattern *sliding along the graph to the right* at a steady rate. *What* rate? Well, the answer is most easily given in simple qualitative terms:

The wave has a distance λ (one *wavelength*) between "crests." Every *period* T, one full wavelength passes a fixed position. Therefore a given crest travels a distance λ in a time T so the *velocity of propagation* of the wave is just

$$c = \frac{\lambda}{T} \qquad \text{or} \qquad c = \frac{\omega}{k} \qquad (6)$$

where I have used c as the symbol for the propagation velocity even though this is a completely *general* relationship between the frequency ω, the wave vector magnitude k and the propagation velocity c of *any* sort of wave, not just electromagnetic waves (for which c has its most familiar meaning, namely the speed of light).

This result can be obtained more easily by noting that A is a function *only* of the *phase* θ of the oscillation,

$$\theta \equiv kx - \omega t \qquad (7)$$

[3] The name "wave vector" is both apt and inadequate — apt because the term *vector* explicitly reminds us that its direction defines the direction of propagation of the wave; inadequate because the essential inverse relationship between k and the *wavelength* λ [see Eq. (1)] is not suggested by the name. Too bad. It is at least a little more descriptive than the name given to the *magnitude* k of $\vec{k}$, namely the "wavenumber."

[4] In general $\vec{k}\cdot\vec{r} = xk_x + yk_y + zk_z$. If $\vec{k} = k\,\hat{\imath}$ then $k_x = k$ and $k_y = k_z = 0$, giving $\vec{k}\cdot\vec{r} = k\,x$.

and that the criterion for "seeing the same waveform" is θ = constant or $d\theta = 0$. If we take the differential of Eq. (7) and set it equal to zero, we get

$$d\theta \;=\; k\,dx \;-\; \omega\,dt \;=\; 0 \quad \text{or} \quad k\,dx \;=\; \omega\,dt$$

$$\text{or} \quad \frac{dx}{dt} = \frac{\omega}{k}.$$

But $dx/dt = c$, the propagation velocity of the waveform. Thus we reproduce Eq. (6). This treatment also shows why we chose $e^{-i\omega t}$ for the time dependence so that Eq. (7) would describe the phase: if we used $e^{+i\omega t}$ then the phase would be $\theta \equiv kx + \omega t$ which gives $dx/dt = -c$, — *i.e.* a waveform propagating in the negative x direction (to the *left* as drawn).

If we use the relationship (6) to write $(kx - \omega t) = k(x - ct)$, so that Eq. (4) becomes

$$A(x,t) \;=\; A_0\, e^{ik(x-ct)},$$

we can extend the above argument to waveforms that are not of the ideal sinusoidal shape shown in Fig. 14.1; in fact it is more vivid if one imagines some special shape like (for instance) a *pulse* propagating down a string at velocity c. As long as $A(x,t)$ *is a function only of* $x' = x - ct$, *no matter what its shape*, it will be *static in time* when viewed by an observer traveling along with the wave[5] at velocity c. This doesn't require any elaborate derivation; x' is just the position measured in such an observer's reference frame!

14.2 The Wave Equation

This is a bogus "derivation" in that we start with a *solution* to the WAVE EQUATION and then show what sort of differential equation it satisfies. Of course, once we have the equation we can work in the other direction, so this is not *so* bad. . . .

Suppose we know that we have a *traveling wave* $A(x,t) = A_0 \cos(kx - \omega t)$.

At a *fixed position* (x = const) we see *SHM* in time:

$$\left(\frac{\partial^2 A}{\partial t^2}\right)_x \;=\; -\,\omega^2 A \qquad (8)$$

(Read: "The second partial derivative of A with respect to time [*i.e.* the *acceleration* of A] with x held fixed is equal to $-\omega^2$ times A itself.") *I.e.* we must have a *linear restoring force.*

Similarly, if we take a "snapshot" (hold t fixed) and look at the *spatial* variation of A, we find the oscillatory behaviour analogous to *SHM*,

$$\left(\frac{\partial^2 A}{\partial x^2}\right)_t \;=\; -\,k^2 A \qquad (9)$$

(Read: "The second partial derivative of A with respect to position [*i.e.* the *curvature* of A] with t held fixed is equal to $-k^2$ times A itself.")

Thus

$$A \;=\; -\frac{1}{\omega^2}\left(\frac{\partial^2 A}{\partial t^2}\right)_x \;=\; -\frac{1}{k^2}\left(\frac{\partial^2 A}{\partial x^2}\right)_t .$$

If we multiply both sides by $-k^2$, we get

$$\frac{k^2}{\omega^2}\left(\frac{\partial^2 A}{\partial t^2}\right)_x \;=\; \left(\frac{\partial^2 A}{\partial x^2}\right)_t .$$

But $\omega = ck$ so $\dfrac{k^2}{\omega^2} = \dfrac{1}{c^2}$, giving the WAVE EQUATION:

$$\boxed{\frac{\partial^2 A}{\partial x^2} \;-\; \frac{1}{c^2}\,\frac{\partial^2 A}{\partial t^2} \;=\; 0} \qquad (10)$$

In words, the *curvature* of A is equal to $1/c^2$ times the *acceleration* of A at any (x,t) point (what we call an *event* in *spacetime*).

[5]Don't try this with an electromagnetic wave! The argument shown here is explicitly *nonrelativistic*, although a more mathematical proof reaches the same conclusion without such restrictions.

Whenever you see this differential equation governing some quantity A, *i.e.* where the acceleration of A is proportional to its curvature, you know that $A(x,t)$ will exhibit wave motion!

14.3 Wavy Strings

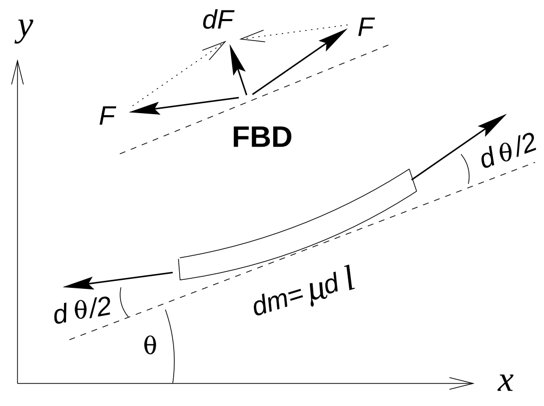

Figure 14.2 A small segment of a taut string.

One system that exhibits wave motion is the *taut string*. Picture a string with a uniform mass per unit length μ under tension F. Ignoring any effects of gravity, the undisturbed string will of course follow a straight line which we label the x axis. There are actually two ways we can "perturb" the quiescent string: with a "longitudinal" compression/stretch displacement (basically a sound wave in the string) or with a "transverse" displacement in a direction perpendicular to the x axis, which we will label the y direction.

The sketch in Fig. 14.2 shows a small string segment of length $d\ell$ and mass $dm = \mu\, d\ell$ which makes an average angle θ with respect to the x axis. The angle actually changes from $\theta - d\theta/2$ at the left end of the segment to $\theta + d\theta/2$ at the right end. For small displacements $\theta \ll 1$ [the large θ shown in the sketch is just for visual clarity] and we can use the SMALL-ANGLE APPROXIMATIONS

$$dx = d\ell\, \cos\theta \approx d\ell$$

$$dy = d\ell\, \sin\theta \approx \theta\, d\ell$$

$$\frac{dy}{dx} = \tan\theta \approx \theta \tag{11}$$

Furthermore, for small θ the net force

$$dF = 2F\, \sin(d\theta/2) \approx 2F\,(d\theta/2) = F\, d\theta \tag{12}$$

acting on the string segment is essentially in the y direction, so we can use Newton's SECOND LAW on the segment at a fixed x location on the string:

$$dF \approx dm\, a_y = \ddot{y}\, dm \qquad \text{or}$$

$$F\, d\theta \approx \ddot{y}\, \mu\, d\ell \qquad \text{or}$$

$$\left(\frac{\partial^2 y}{\partial t^2}\right)_x \approx \frac{F\, d\theta}{\mu\, d\ell} \approx \frac{F}{\mu}\left(\frac{d\theta}{dx}\right). \tag{13}$$

Referring now back to Eq. (11) we can use $\theta \approx dy/dx$ to set

$$\left(\frac{d\theta}{dx}\right) \approx \left(\frac{\partial^2 y}{dx^2}\right)_t \tag{14}$$

— *i.e.* the *curvature* of the string at time t. Plugging Eq. (14) back into Eq. (13) gives

$$\left(\frac{\partial^2 y}{\partial t^2}\right)_x - \frac{F}{\mu}\left(\frac{\partial^2 y}{dx^2}\right)_t \approx 0 \tag{15}$$

which is the WAVE EQUATION with

$$\frac{1}{c^2} = \frac{\mu}{F} \quad \text{or} \quad c = \sqrt{\frac{F}{\mu}}. \tag{16}$$

We may therefore jump right to the conclusion that *waves* will propagate down a taut string at this velocity.

14.3.1 Polarization

One nice feature of waves in a taut string is that they explicitly illustrate the phenomenon of *polarization*: if we change our notation slightly to label the string's equilibrium direction (and therefore the direction of propagation of a wave in the string) as z, then there are two orthogonal choices of "transverse" direction: x or y. We can set the string "wiggling" in either transverse direction, which we call the two orthogonal *polarization* directions.

Of course, one can choose an infinite number of transverse polarization directions, but these correspond to simple *superpositions* of x- and y-polarized waves with the same phase.

One can also superimpose x- and y-polarized waves of the same frequency and wavelength but with phases differing by $\pm\pi/2$. This gives left- and right-*circularly polarized* waves; I will leave the mathematical description of such waves (and the mulling over of its physical meaning) as an "exercise for the student...."

14.4 Linear Superposition

The above derivation relied heavily on the SMALL-ANGLE APPROXIMATIONS which are valid only for *small displacements* of the string from its equilibrium position ($y = 0$ for all x). This almost always true: the simple description of a wave given here is only strictly valid in the limit of small displacements from equilibrium; for large displacements we usually pick up "anharmonic" terms corresponding to *nonlinear restoring forces*. But as long as the restoring force stays linear we have an important consequence: *several different waves can propagate independently through the same medium.* (*E.g.* down the same string.) The displacement at any given time and place is just the *linear sum* of the displacements due to each of the simultaneously propagating waves. This is known as the PRINCIPLE OF LINEAR SUPERPOSITION, and it is essential to our understanding of wave phenomena.

In general the overall displacement $A(x,t)$ resulting from the linear superposition of two waves $A_1 e^{i(k_1 x - \omega_1 t)}$ and $A_2 e^{i(k_2 x - \omega_2 t)}$ is given by

$$A(x,t) \;=\; A_1 e^{i(k_1 x - \omega_1 t)} + A_2 e^{i(k_2 x - \omega_2 t)}. \quad (17)$$

Let's look at a few simple examples.

14.4.1 Standing Waves

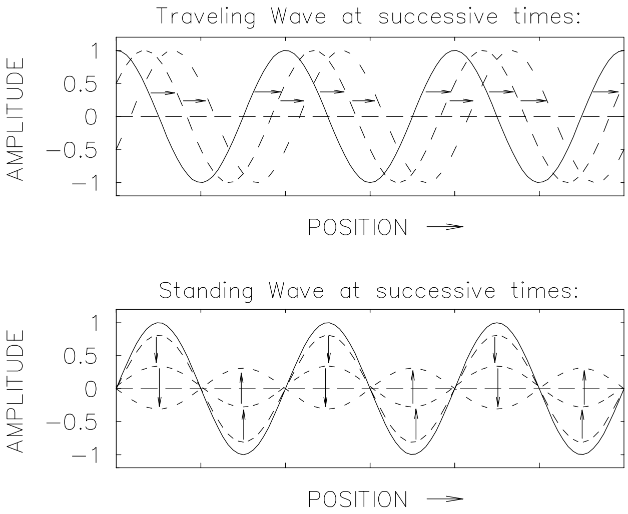

Figure 14.3 Traveling *vs.* standing waves.

A particularly interesting example of superposition is provided by the case where $A_1 = A_2 = A_0$, $k_1 = k_2 = k$ and $\omega_1 = -\omega_2 = \omega$. That is, two otherwise identical waves *propagating in opposite directions*. The algebra is simple:

$$\begin{aligned} A(x,t) &= A_0 \left[e^{i(kx - \omega t)} + e^{i(kx + \omega t)} \right] \\ &= A_0 e^{ikx} \left[e^{-i\omega t} + e^{+i\omega t} \right] \\ &= A_0 e^{ikx} [\cos(\omega t) - i \sin(\omega t) \\ &\qquad + \cos(\omega t) + i \sin(\omega t)] \\ &= 2A_0 \cos(\omega t) e^{ikx}. \end{aligned} \quad (18)$$

The real part of this (which is all we ever actually use) describes a sinusoidal waveform of wavelength $\lambda = 2\pi/k$ whose *amplitude* $2A_0 \cos(\omega t)$ oscillates in time but which does *not* propagate in the x direction — *i.e.* the lower half of Fig. 14.3. Standing waves are very common, especially in situations where a traveling wave is *reflected* from a boundary, since this automatically creates a second wave of similar amplitude and wavelength propagating back in the opposite direction — the very condition assumed at the beginning of this discussion.

14.4.2 Classical Quantization

None of the foregoing discussion allows us to *uniquely specify* any wavelike solution to the WAVE EQUATION, because nowhere have we given any BOUNDARY CONDITIONS forcing the wave to have any particular behaviour at any particular point. This is not a problem for the general phenomenology discussed so far, but if you want to actually describe one particular wave you have to know this stuff.

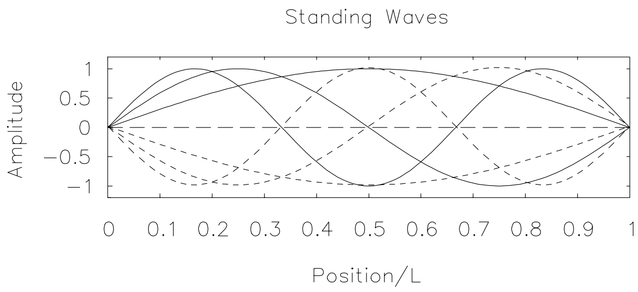

Figure 14.4 The first three allowed standing waves in a "closed box" (*e.g.* on a string with fixed ends).

Boundary conditions are probably easiest to illustrate with the system of a taut string of length L with fixed ends, as shown in Fig. 14.4.[6] Fixing the ends forces the wave function $A(x,t)$ to have *nodes* (positions where the amplitude is always zero) at those positions. This immediately rules out traveling waves and restricts the simple sinusoidal "modes" to *standing waves* for which L is an integer number of half-wavelengths:[7]

$$\lambda_n = \frac{2L}{n}, \quad n = 1, 2, 3, \cdots \tag{19}$$

[6]The Figure could also describe standing *sound waves* in an *organ pipe* closed at both ends, or the electric field strength in a resonant cavity, or the probability amplitude of an electron confined to a one-dimensional "box" of length L.

[7]Note that the $n^{\rm th}$ *mode* has $(n-1)$ *nodes* in addition to the two at the ends.

Assuming that $c = \omega/k = \lambda\nu = \text{const}$, the frequency ν [in *cycles per second* or *Hertz* (*Hz*)] of the $n^{\rm th}$ mode is given by $\nu_n = c/\lambda_n$ or

$$\nu_n = n\,\frac{c}{2L}, \quad n = 1, 2, 3, \cdots \tag{20}$$

For a string of linear mass density μ under tension F we can use Eq. (16 to write what one might frivolously describe as THE GUITAR TUNER'S EQUATION:

$$\nu_n = \frac{n}{2L}\sqrt{\frac{F}{\mu}}, \quad n = 1, 2, 3, \cdots \tag{21}$$

Note that a given string of a given length L under a given tension F has in principle an infinite number of modes (resonant frequencies); the guitarist can choose which modes to excite by plucking the string at the position of an *antinode* (position of *maximum* amplitude) for the desired mode(s). For the first few modes these antinodes are at quite different places, as evident from Fig. 14.4. As another "exercise for the student" try deducing the relationship between modes with a *common antinode* — these will all be excited as "harmonics" when the string is plucked at that position.

Exactly the same formulae apply to *sound waves* in *organ pipes* if they are *closed at both ends*. An organ pipe *open* at one end must however have an *antinode* at that end; this leads to a slightly different scheme for enumerating modes, but one that you can easily deduce by a similar sequence of logic.

This sort of restriction of the allowed modes of a system to a discrete set of values is known as QUANTIZATION. However, most people are not accustomed to using that term to describe macroscopic classical systems like taut strings; we have a tendency to think of quantization as something that only happens in QUANTUM MECHANICS. In reality, quantization is an ubiquitous phenomenon wherever *wave motion* runs up against *fixed boundary conditions*.

14.5 Energy Density

Consider again our little element of string at position x. We have shown that (for fixed x) the mass element will execute *SHM* as a function of time t. Therefore there is an effective LINEAR RESTORING FORCE in the y direction acting on the mass element $dm = \mu dx$: $dF = F\,d\theta = F\,(\partial^2 y/\partial x^2)\,dx$. But for a simple traveling wave we have[8] $y(x,t) = y_0 \cos(kx - \omega t)$ so $(\partial^2 y/\partial x^2) = -k^2 y$, giving $dF = -\left[k^2\,F\,dx\right]\,y$. In other words, the *effective spring constant* for an element of string dx long is $\kappa_{\rm eff} = k^2\,F\,dx$ where I have used the unconventional notation κ for the effective spring constant to avoid confusing it with the *wavenumber* k, which is something completely different. Applying our knowledge of the potential energy stored in a stretched spring, $dU = \frac{1}{2}\kappa_{\rm eff}\,y^2$, we have the *elastic potential energy stored in the string per unit length*, $dU/dx = \frac{1}{2}k^2\,F\,y^2$ or, plugging in $y(x,t)$,

$$\frac{dU}{dx} = \frac{1}{2}k^2\,F\,y_0^2\,\cos^2(kx - \omega t) \tag{22}$$

— that is, *the potential energy density is proportional to the amplitude squared.*

What about *kinetic* energy? From *SHM* we expect the energy to be shared between potential and kinetic energy as each mass element oscillates through its period. Well, the kinetic energy dK of our little element of string is just $dK = \frac{1}{2}dm\,v_y^2$. Again $dm = \mu\,dx$ and now we must evaluate v_y. Working from $y(x,t) = y_0\,\cos(kx - \omega t)$ we have $v_y = -\omega\,y_0\,\sin(kx - \omega t)$, from which we can write

$$\frac{dK}{dx} = \frac{1}{2}\mu\,\omega^2\,y_0^2\,\sin^2(kx - \omega t). \tag{23}$$

The total energy density is of course the sum of these two:

$$\frac{dE}{dx} = \frac{dU}{dx} + \frac{dK}{dx} \qquad \text{or}$$

$$\frac{dE}{dx} = \frac{1}{2}y_0^2\left[k^2\,F\,\cos^2\theta + \mu\,\omega^2\,\sin^2\theta\right]$$

where $\theta \equiv kx - \omega t$. Using $c = \omega/k = \sqrt{F/\mu}$ we can write this as

$$\frac{dE}{dx} = \frac{1}{2}y_0^2\left[\mu\,\omega^2\,\cos^2\theta + \mu\,\omega^2\,\sin^2\theta\right] \quad \text{or}$$

$$\frac{dE}{dx} = \frac{1}{2}\mu\,\omega^2\,y_0^2. \tag{24}$$

You can use $F\,k^2$ in place of $\mu\,\omega^2$ if you like, since they are equal. [Exercise for the student.]

Note that the net energy density (potential plus kinetic) is constant in time and space for such a uniform traveling wave. It just switches back and forth between potential and kinetic energy *twice* every cycle. Since the *average* of either $\cos^2\theta$ or $\sin^2\theta$ is 1/2, the energy density is *on average* shared equally between kinetic and potential energy.

If we want to know the energy per unit time (*power* P) transported past a certain point x by the wave, we just multiply dE/dx by $c = dx/dt$ to get

$$P \equiv \frac{dE}{dt} = \frac{1}{2}\mu\,\omega^2\,c\,y_0^2. \tag{25}$$

Again, you can play around with the constants; instead of $\mu\,\omega^2\,c$ you can use $\omega^2\,\sqrt{F\mu}$ and so on.

Note that while the wave does not transport any *mass* down the string (all physical motion is *transverse*) it does transport *energy*. This is an ubiquitous property of waves, lucky for us!

14.6 Water Waves

Although all sorts of waves are ubiquitous in our lives,[9] our most familiar "wave experiences" are probably with *water waves*, which

[8] I have avoided complex exponentials here to avoid confusion when I get around to calculating the transverse *speed* of the string element, v_y. The *acceleration* is the same as for the complex version.

[9] Indeed, we are *made* of waves, as QUANTUM MECHANICS has taught us!

are unfortunately one of the *least simple* types of waves. Therefore, although water waves are routinely used for illustration, they are rarely discussed in great depth (heh, heh) in introductory Physics texts. They do, however, serve to illustrate one important feature of waves, namely that *not all waves obey* the simple relationship $c = \omega/k$ for their *propagation velocity* c.

Let's restrict ourselves to *deep ocean* waves, where the "restoring force" is simply *gravity*. (When a wave reaches shallow water, the bottom provides an immobile boundary that complicates matters severely, as anyone knows who has ever watched surf breaking on a beach!) The motion of an "element" of water in such a wave is *not* simply "up and down" as we pretended at the beginning of this chapter, but a *superposition* of "up and down" with "back and forth" in the direction of wave propagation. A cork floating on the surface of such a wave executes *circular* motion, or so I am told. (It is actually quite difficult to confirm this assertion experimentally since it requires a fixed reference that is *not* moving with the water — a hard thing to arrange in practice without disturbing the wave itself.) More importantly, the *propagation velocity* of such waves is *higher* for *longer wavelength*.

14.6.1 Phase *vs.* Group Velocity

The precise relationship between angular frequency ω and wavenumber k for deep-water waves is

$$\omega = \sqrt{\frac{g\,k}{2}} \tag{26}$$

where g has its usual meaning. Such a functional relationship $\omega(k)$ between frequency and wavenumber is known as the DISPERSION RELATION for waves in the medium in question, for reasons that will be clear shortly.

If we have a simple traveling plane wave $A(x,t) = A_0 \exp[i(kx - \omega t)]$, with no beginning and no end, the rate of propagation of a point of constant phase (known as the PHASE VELOCITY $v_{\rm ph}$) is still given by Eq. (6):

$$\boxed{v_{\rm ph} \equiv \frac{\omega}{k}} \tag{27}$$

However, by combining Eq. (27) with Eq. (26) we find that the phase velocity is *higher* for *smaller* k (longer λ):

$$v_{\rm ph} = \sqrt{\frac{g}{2k}}. \tag{28}$$

Moreover, such a wave *carries no information*. It has been passing by forever and will continue to do so forever; it is the same amplitude everywhere; and so on. Obviously our PLANE WAVE is a bit of an oversimplification. If we want to send a *signal* with a wave, we have to turn it on and off in some pattern; we have to make wave *pulses* (or, anticipating the terminology of QUANTUM MECHANICS, "WAVE PACKETS"). And when we do that with water waves, we notice something odd: *the wave packets propagate slower than the "wavelets" in them!*

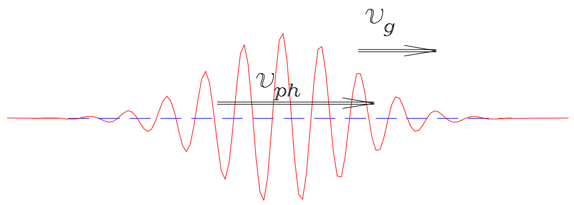

Figure 14.5 A WAVE PACKET moving at $v_{\rm g}$ with "wavelets" moving through it at $v_{\rm ph}$.

Such a packet is a superposition of waves with different wavelengths; the k-dependence of $v_{\rm ph}$ causes a phenomenon known as DISPERSION, in which waves of different wavelength, initially moving together in phase, will drift apart as the packet propagates, making it "broader" in both space and time. (Obviously such a DISPERSIVE MEDIUM is undesirable for the transmission of information!) But how do we determine the effective speed of transmission of said information — *i.e.* the propagation velocity *of* *the packet itself*, called the GROUP VELOCITY $v_{\rm g}$?

Allow me to defer an explanation of the following result until a later section. The *general* definition of the group velocity (the speed of transmission of information and/or energy in a wave packet) is

$$\boxed{v_{\rm g} \equiv \frac{\partial \omega}{\partial k}} . \tag{29}$$

For the particular case of deep-water waves, Eq. (29) combined with Eq. (26) gives

$$v_{\rm g} = \frac{1}{2}\sqrt{\frac{g}{2k}}. \tag{30}$$

That is, the *packet* propagates at *half* the speed of the "wavelets" within it. This behaviour can actually be observed in the wake of a large vessel on the ocean, seen from high above (*e.g.* from an airliner).

Such exotic-seeming wave phenomena are ubiquitous in all dispersive media, which are anything but rare. However, in the following chapters we will restrict ourselves to waves propagating through simple non-dispersive media, for which the DISPERSION RELATION is just $\omega = c\,k$ with c constant, for which $v_{\rm ph} = v_{\rm g} = c$.

14.7 Sound Waves

Picture a "snapshot" (holding time t fixed) of a small cylindrical section of an elastic medium, shown in Fig. 14.6: the cross-sectional area is A and the length is dx. An excess pressure P (over and above the ambient pressure existing in the medium at equilibrium) is exerted on the left side and a slightly different pressure $P + dP$ on the right. The resulting volume element $dV = A\,dx$ has a mass $dm = \rho\,dV = \rho\,A\,dx$, where ρ is the mass density of the medium. If we choose the positive x direction to the right, the net

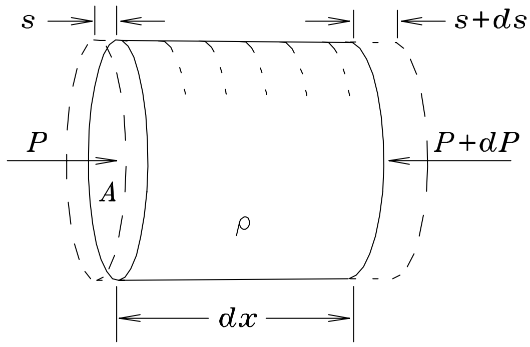

Figure 14.6 Cylindrical element of a compressible medium.

force acting on dm in the x direction is $dF_x = PA - (P + dP)A = -A\,dP$.

Now let s denote the *displacement* of particles of the medium from their equilibrium positions. (I didn't use A here because I am using that symbol for the *area*. This may also differ between one end of the cylindrical element and the other: s on the left *vs.* $s + ds$ on the right. We assume the displacements to be in the x direction but *very small* compared to dx, which is itself no great shakes.[10]

The *fractional change in volume* dV/V of the cylinder due to the *difference* between the displacements at the two ends is

$$\frac{dV}{V} = \frac{(s + ds)\,A - s\,A}{A\,dx} = \frac{ds}{dx}$$
$$= \left(\frac{\partial s}{\partial x}\right)_t \tag{31}$$

where the rightmost expression reminds us explicitly that this description is being constructed around a "snapshot" with t held fixed.

Now, any elastic medium is by definition compressible but "fights back" when compressed ($dV < 0$) by exerting a pressure in the direction of increasing volume. The BULK MODULUS B is a constant characterizing how hard the medium fights back — a sort of

[10]Note also that any of s, ds, P or dP can be either positive or negative; we merely illustrate the math using an example in which they are all positive.

3-dimensional analogue of the SPRING CONSTANT. It is defined by

$$P = -B \frac{dV}{V}. \tag{32}$$

Combining Eqs. (31) and (32) gives

$$P = -B \left(\frac{\partial s}{\partial x} \right)_t \tag{33}$$

so that the *difference* in pressure between the two ends is

$$dP = \left(\frac{\partial P}{\partial x} \right)_t dx = -B \left(\frac{\partial^2 s}{\partial x^2} \right)_t dx. \tag{34}$$

We now use $\sum F_x = m\, a_x$ on the mass element, giving

$$-A\, dP = AB \left(\frac{\partial^2 s}{\partial x^2} \right)_t dx$$

$$= dm\, a_x = \rho\, A\, dx \left(\frac{\partial^2 s}{\partial t^2} \right)_x \tag{35}$$

where we have noted that the acceleration of all the particles in the volume element (assuming $ds \ll s$) is just $a_x \equiv (\partial^2 s / \partial t^2)_x$.

If we cancel $A\, dx$ out of Eq. (35), divide through by B and collect terms, we get

$$\left(\frac{\partial^2 s}{\partial x^2} \right)_t - \frac{\rho}{B} \left(\frac{\partial^2 s}{\partial t^2} \right)_x = 0 \qquad \text{or}$$

$$\left(\frac{\partial^2 s}{\partial x^2} \right)_t - \frac{1}{c^2} \left(\frac{\partial^2 s}{\partial t^2} \right)_x = 0 \tag{36}$$

which the acute reader will recognize as the WAVE EQUATION in one dimension (x), provided

$$c = \sqrt{\frac{B}{\rho}} \tag{37}$$

is the velocity of propagation.

The fact that disturbances in an elastic medium obey the WAVE EQUATION *guarantees* that such disturbances will propagate as simple *waves* with phase velocity c given by Eq. (37).

We have now progressed from the strictly one-dimensional propagation of a wave in a taut string to the two-dimensional propagation of waves on the surface of water to the three-dimensional propagation of pressure waves in an elastic medium (*i.e.* sound waves); yet we have continued to pretend that the only *simple* type of traveling wave is a *plane wave* with constant $\vec{\boldsymbol{k}}$. This will never do; we will need to treat all sorts of wave phenomena, and although in general we can treat most types of waves as *local approximations to plane waves* (in the same way that we treat the Earth's surface as a flat plane in most mechanics problems), it is important to recognize the most important features of at least one other common idealization — the SPHERICAL WAVE.

14.8 Spherical Waves

The utility of thinking of $\vec{\boldsymbol{k}}$ as a "ray" becomes even more obvious when we get away from plane waves and start thinking of waves with *curved* wavefronts. The simplest such wave is the type that is emitted when a pebble is tossed into a still pool — an example of the "point source" that radiates waves isotropically in all directions. The wavefronts are then *circles* in two dimensions (the surface of the pool) or *spheres* in three dimensions (as for sound waves) separated by one wavelength λ and heading outward from the source at the propagation velocity c. In this case the "rays" k point along the radius vector $\hat{\boldsymbol{r}}$ from the source at any position and we can once again write down a rather simple formula for the "wave function" (displacement A as a function of position) that depends only on the time t and the *scalar* distance r from the source.

A plausible first guess would be just $A(x,t) = A_0\, e^{i(kr - \omega t)}$, but this cannot be right! Why not? Because it violates energy conservation. The energy density stored in a wave is proportional to the square of its amplitude; in the trial solution above, the amplitude of the outgoing spherical wavefront is constant as a function or r, but the *area* of that wavefront increases as r^2. Thus the energy in the wavefront increases as r^2? I think not. We can get rid of this effect by just dividing the amplitude by r (which divides the energy density by r^2). Thus a trial solution is

$$\boxed{A(x,t) \;=\; A_0 \, \frac{e^{i(kr-\omega t)}}{r}.} \qquad (38)$$

which is, as usual, correct.[11] The factor of $1/r$ accounts for the conservation of energy in the outgoing wave: since the spherical "wave front" distributes the wave's energy over a surface area $4\pi r^2$ and the flux of energy per unit area through a spherical surface of radius r is proportional to the *square* of the wave amplitude at that radius, the integral of $|f|^2$ over the entire sphere (*i.e.* the total outgoing *power*) is independent of r, as it must be.

We won't use this equation for anything right now, but it is interesting to know that it does accurately describe an outgoing[12] spherical wave.

The perceptive reader will have noticed by now that Eq. (38) is not a solution to the WAVE EQUATION as represented in one dimension by Eq. (10). That is hardly surprising, since the spherical wave solution is an intrinsically 3-dimensional beast; what happened to y and z? The correct *vector* form of the WAVE EQUATION is

$$\boxed{\nabla^2 A \;-\; \frac{1}{c^2}\,\frac{\partial^2 A}{\partial t^2} \;=\; 0} \qquad (39)$$

where the LAPLACIAN operator ∇^2 can be expressed in Cartesian[13] coordinates (x, y, z) as[14]

$$\nabla^2 \;=\; \frac{\partial^2}{\partial x^2} \;+\; \frac{\partial^2}{\partial y^2} \;+\; \frac{\partial^2}{\partial z^2}. \qquad (40)$$

With a little patient effort you can show that Eq. (38) does indeed satisfy Eq. (39), if you remember that $r = \sqrt{x^2+y^2+z^2}$. Or you can just take my word for it....

[11] I should probably show you a few wrong guesses first, just to avoid giving the false impression that we always guess right the first time in Physics; but it would use up a lot of space for little purpose; and besides, "knowing the answer" is always the most powerful problem-solving technique!

[12] One can also have "incoming" spherical waves, for which Eq. (38) becomes

$$A(x,t) \;=\; A_0 \, \frac{e^{i(kr+\omega t)}}{r}.$$

[13] The LAPLACIAN operator can also be represented in other coordinate systems such as spherical (r, θ, ϕ) or cylindrical (ρ, θ, z) coordinates, but I won't get carried away here.

[14] The LAPLACIAN operator can also be thought of as the inner (scalar or "dot") product of the GRADIENT operator $\vec{\boldsymbol{\nabla}}$ with itself: $\nabla^2 = \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{\nabla}}$, where

$$\vec{\boldsymbol{\nabla}} \;=\; \hat{\imath}\frac{\partial}{\partial x} \;+\; \hat{\jmath}\frac{\partial}{\partial y} \;+\; \hat{\boldsymbol{k}}\frac{\partial}{\partial z}$$

in Cartesian coordinates. This VECTOR CALCULUS stuff is really elegant — you should check it out sometime — but it is usually regarded to be beyond the scope of an introductory presentation like this.

14.9 Electromagnetic Waves

We have some difficulty visualizing a *wave* consisting only of electric and magnetic *fields*. However, if we plot the strength of $\vec{\boldsymbol{E}}$ along one axis and the strength of $\vec{\boldsymbol{B}}$ along another (perpendicular) axis, as in Fig. 14.7, then the direction of propagation $\hat{\boldsymbol{k}}$ will be perpendicular to both $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$, as shown.

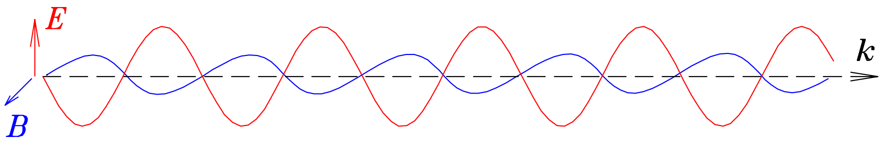

Figure 14.7 A linearly polarized electromagnetic wave. The electric and magnetic fields $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$ are mutually perpendicular and both are perpendicular to the direction of propagation $\hat{\boldsymbol{k}}$ ($\vec{\boldsymbol{k}}$ is the *wave vector*).

14.9.1 Polarization

The case shown in Fig. 14.7 is *linearly polarized*, which means simply that the $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$ fields are in specific fixed directions. Of course, the directions of $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$ could be interchanged, giving the "opposite" polarization. Polaroid sunglasses transmit the light waves with $\vec{\boldsymbol{E}}$ vertical (which are not reflected efficiently off horizontal surfaces) and absorb the light waves with $\vec{\boldsymbol{E}}$ horizontal (which are), thus reducing "glare" (reflected light from horizontal surfaces) without blocking out all light.

There is another possibility, namely that the two linear polarizations be *superimposed* so that both the $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$ vectors *rotate* around the direction of propagation $\hat{\boldsymbol{k}}$, remaining always perpendicular to $\hat{\boldsymbol{k}}$ and to each other. This is known as *circular polarization*. It too comes in two versions, *right* circular polarization and *left* circular polarization, referring to the hand whose fingers curl in the direction of the rotation if the thumb points along $\hat{\boldsymbol{k}}$.

14.9.2 The Electromagnetic Spectrum

We have special names for electromagnetic (EM) waves of different wavelengths and frequencies.[15] We call EM waves with $\lambda \gtrsim 1$ m "radio waves," which are subdivided into various ranges or "bands" like "short wave" (same thing as high frequency), VHF (very high frequency), UHF (ultra high frequency) and so on.[16] The dividing line between "radar" and "microwave" bands (for example) is determined by arbitrary convention, if at all, but the rule of thumb is that if the wavelength fits inside a *very* small appliance it is "microwave." Somewhere towards the short end of the microwave spectrum is the beginning of "far infrared," which of course becomes "near infrared" as the wavelength gets still shorter. The name "infrared" is meant to suggest

[15] If the wavelength λ *in*creases (so that the wavenumber $k = 2\pi/\lambda$ *de*creases), then the frequency ω must *de*crease to match, since the ratio ω/k must always be equal to the same propagation velocity c.

[16] One can detect a history of proponents of different bands claiming ever higher (and therefore presumably "better") frequency ranges. . . .

frequencies *below* those of the *red* end of the *visible light* spectrum of *EM* waves, which extends (depending on the individual eye) from a wavelength of roughly 500 nm (5000 Å) for red light through orange, yellow, green and blue to roughly 200 nm (2000 Å) for violet light. Beyond that we lost sight of the shorter wavelengths (so to speak) and the next range is called "near ultraviolet," the etymology of which is obvious. Next comes "far ultraviolet" which fades into "soft x-rays" and in turn "hard x-rays" and finally "gamma rays" as the frequency increases and the wavelength gets shorter. Note all the different kinds of "rays" that are all just other forms of *light* — *i.e.* *EM* waves — with different wavelengths!

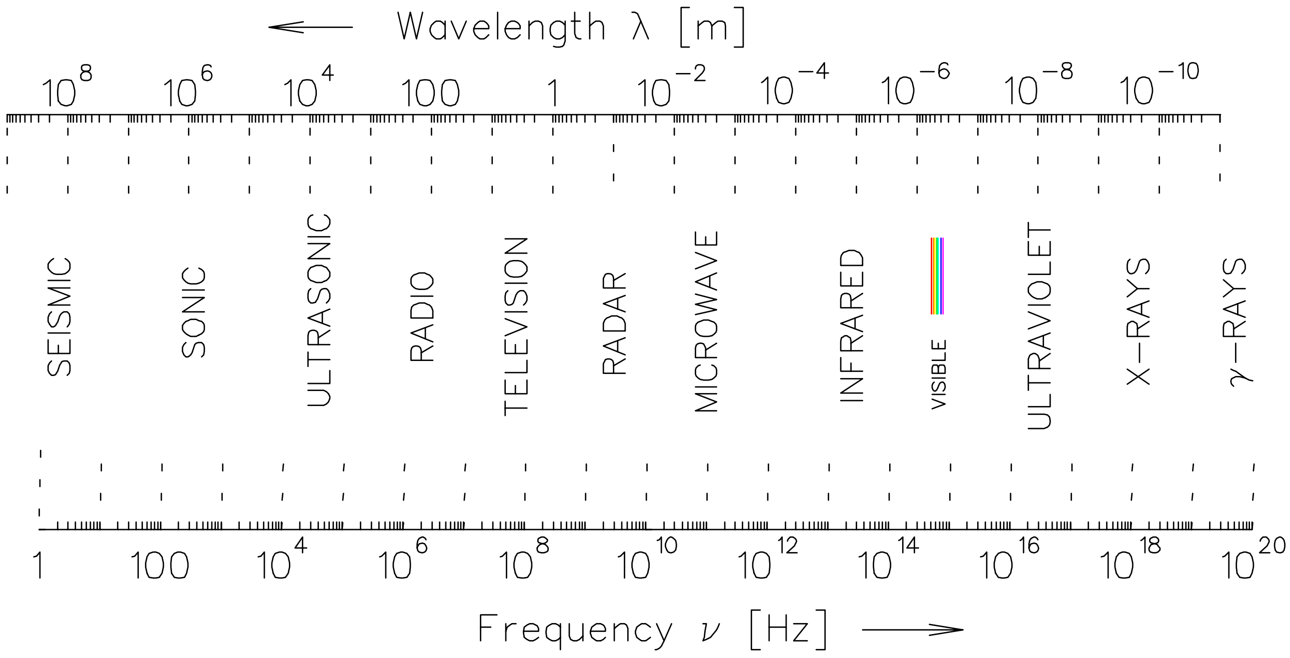

Figure 14.8 The electromagnetic spectrum. Note logarithmic wavelength and frequency scales.

14.10 Reflection

The simplest thing waves do is to REFLECT off flat surfaces. Since billiard balls do the same thing quite nicely, this is not a particularly distinctive behaviour of waves — which was probably one of the reasons why Newton was convinced that light consisted of *particles*.[17] The reflection of waves looks something like Fig. 14.9.

The incoming wave vector $\vec{\boldsymbol{k}}$ makes the same angle with the surface (or, equivalently, with the direction *normal* to the surface) as the outgoing wavevector $\vec{\boldsymbol{k}}'$:

$$\theta = \theta' \qquad (41)$$

This is the most important property of reflection, and it can be stated in words thus:

$\boxed{\text{The } \textit{incident} \text{ [incoming] angle is equal to the } \textit{reflected} \text{ [outgoing] angle.}}$

[17] He was actually correct, but it is equally true that light consists of *waves*. If you are hoping that these apparently contradictory statements will be reconciled with common sense by the Chapter on QUANTUM MECHANICS, you are in for a disappointment. Common sense will have to be beaten into submission by the utterly implausible facts.

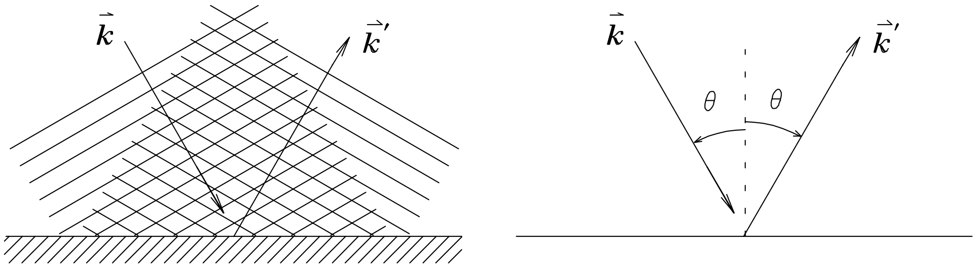

Figure 14.9 Reflection of a wave from a flat surface.

14.11 Refraction

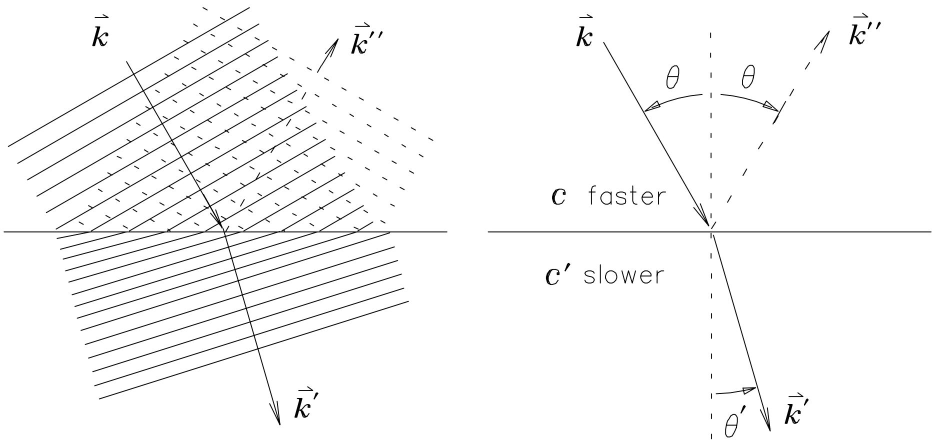

Figure 14.10 Refraction of a wave at a boundary between two media where the propagation velocity (c) of the wave in the first medium is *greater* than that (c') in the second medium. The diagram on the left shows the *wavefronts* ("crests" of the waves) and the corresponding perpendicular wavevectors $\vec{k}$ (incoming wave), $\vec{k}'$ (transmitted wave) and $\vec{k}$" (reflected wave). The diagram on the right shows the angles between the wavevectors and the normal to the interface.

When a wave crosses a boundary between two regions in which its velocity of propagation has different values, it "bends" toward the region with the *slower* propagation velocity. The following mnemonic image can help you remember the qualitative sense of this phenomenon, which is known as REFRACTION: picture the wave front approaching the boundary as a *yardstick* moving through some *fluid* in a direction perpendicular to its length. If one end runs into a *thicker* fluid first, it will "drag" that end a little so that the trailing end gets ahead of it, changing the direction of motion gradually until the whole meter stick is in the thicker fluid where it will move more slowly.[18]

Conversely, if one end emerges first into a *thinner* fluid (where it can move faster) it will pick up speed and the trailing end will fall behind. This picture also explains why there is no "bending" if the wave hits the interface *normally* (at right angles). The details are revealed mathematically

[18]Boy, is this ever Aristotelian!

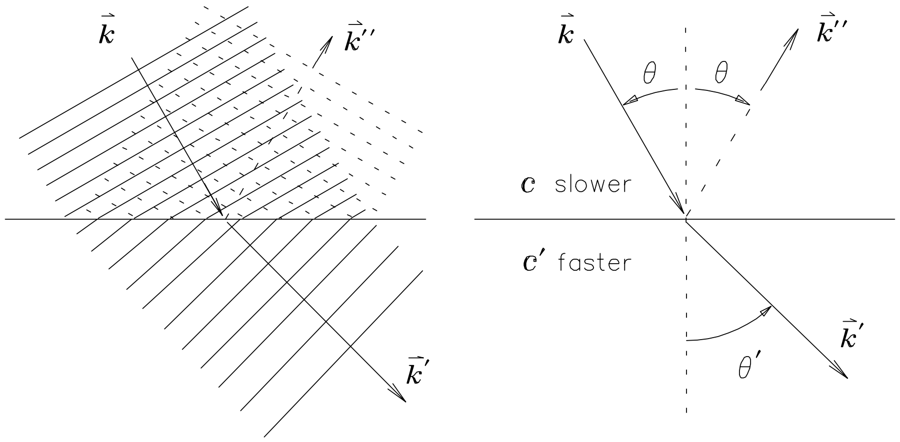

Figure 14.11 Refraction of a wave at a boundary between two media where the propagation velocity (c) of the wave in the first medium is *less* than that (c') in the second medium.

(of course) in SNELL'S LAW:[19]

$$\frac{\sin(\theta)}{\sin(\theta')} = \frac{c}{c'} \tag{42}$$

where θ is the angle of incidence of the incoming wave (the angle that $\vec{\boldsymbol{k}}$ makes with the normal to the interface), θ' is the angle that the refracted wavevector $\vec{\boldsymbol{k}}'$ makes with the same normal, c is the propagation velocity of the wave in the first medium and c' is the propagation velocity of the wave in the second medium.

[19]SNELL'S LAW is normally expressed in terms of the INDEX OF REFRACTION n in each medium:

$$n \, \sin(\theta) \;=\; n' \, \sin(\theta'),$$

where (we now know) the INDEX OF REFRACTION is the ratio of the speed of light in vacuum to the speed of light in the medium:

$$n \;\equiv\; \frac{c_0}{c}.$$

The reason for inventing such a semicircular definition was that when Willebrord Snell discovered this empirical relationship in 1621 he had no idea what n was, only that every medium had its own special value of n. (This is typical of anything that gets the name "index.") I see no pedagogical reason to even define the dumb thing.

Another semi-obvious consequence of the fact that the "crests" of the waves remain continuous[20] is that the wavelength gets shorter as the wave enters the "thicker" medium or longer as it enters a "thinner" medium. Another way of putting this is that *the frequency stays the same* (and therefore so does the *period* T) as the wave crosses the boundary. Since $c = \lambda/T$ this means that if the velocity decreases, so does the wavelength. One can follow this argument a bit further to *derive* SNELL'S LAW from a combination of geometry and logic. I haven't done this, but you might want to....

There is also always a *reflected* wave at any interface, though it may be weak. The reflected wave is shown as dotted lines in Figs. 14.10 and 14.11, where its wavevector is denoted $\vec{k}''$. This phenomenon is familiar as a source of annoyance to anyone who has tried to watch television in a room with a sunny window facing the TV screen. However, it does have some redeeming features, as can be deduced from a thoughtful analysis of Eq. (42). For instance, if the wave is emerging from a "thick" medium into a "thin" medium as in Fig. 14.11 (like light emerging from glass into air), then there is some incoming angle θ_c, called the CRITICAL ANGLE, for which the *refracted* wave will actually be *parallel to the interface* — *i.e.* $\theta' = \pi/2$ (90°). This implies $\sin(\theta') = 1$ so that SNELL'S LAW reads

$$\sin(\theta_c) \;=\; \frac{c}{c'} \tag{43}$$

which has a solution only if $c' > c$ — *i.e.* for emergence into a "thinner" medium with a higher wave propagation velocity, as specified earlier.

What happens, qualitatively, is that as θ gets larger and larger (closer and closer to "grazing incidence") the *amplitude* (strength) of the transmitted wave gets weaker and weaker, while the amplitude of the *reflected* wave gets stronger and stronger, until for incoming angles $\theta \geq \theta_c$ there *is no* transmitted wave and the wave is *entirely reflected*. This phenomenon is known as TOTAL INTERNAL REFLECTION and has quite a few practical consequences.

Because of total internal reflection, a fish cannot see out of the water except for a limited "cone" of vision overhead bounded by the critical angle for water, which is about $\sin^{-1}(1/1.33)$ or 49°. Lest this lend reckless abandon to fishermen, it should be kept in mind that the light "rays" which appear to come from just under 49° from the vertical are actually coming from just across the water's surface, so the fish has a pretty good view of the surrounding environment — it just looks a bit distorted. To observe this phenomenon with your own eyes, put on a good diving mask, carefully slip into a still pool and hold your breath until the surface is perfectly calm again. Looking up at the surface, you will see the world from the fish's perspective (except that the fish is probably a good deal less anoxic) — inside a cone of about 49° from the vertical, you can see out of the water; but outside that cone, the surface forms a perfect mirror!

How total *is* total internal reflection? Total! If the surface has no scratches *etc.*, the light is *perfectly* reflected back into the denser medium. This is how "light pipes" work — light put into one end of a long Lucite rod will follow the rod through bends and twists (as long as they are "gentle" so that the light never hits the surface at less than the critical angle) and emerge at the other end attenuated only by the absorption in the Lucite itself. Even better transmission is achieved in FIBER OPTICS, where fine threads of special glass are prepared with extremely low absorption for the wavelengths of light that are used to send signals down them. A faint pulse of light sent into one end of a fiber optic transmission line will emerge many kilometers down the line with nothing "leaking out" in between. (This feature is especially attractive to those who don't want their conversations bugged, or so I

[20] A "crest" doesn't turn into a "trough" just because the propagation velocity changes!

am told.) Another application was invented by Lorne Whitehead while he was a UBC Physics graduate student: by an ingenious trick he was able to make a large-diameter *hollow* LIGHT PIPE [trademark] which avoids even the small losses in the Lucite itself! Using this trick he is able to "pipe" large amounts of light from single (efficient) light sources [including rooftop solar collectors] into other areas [like the interiors of office buildings] using strictly passive components that do not wear out. He founded a company called TIR — see if you can guess what the acronym stand for!

14.12 Huygens' Principle

At the beginning of this chapter we pictured only PLANE WAVES, in which the wavefronts ("crests" of the waves) form long straight lines (or, in space, flat planes) moving along together in parallel (separated by one wavelength λ) in a common direction $\hat{\boldsymbol{k}}$. One good reason for sticking to this description for as long as possible (and returning to it every chance we get) is that it is so *simple* — we can write down an explicit formula for the amplitude of a plane wave as a function of time and space whose qualitative features are readily apparent (with a little effort). Another good reason has to do with the fact that *all* waves *look pretty much like plane waves* when they are *far from their origin*.[21] We will come back to this shortly. A final reason for our love of plane waves is that they are so easily related to the idea of "RAYS."

In GEOMETRICAL OPTICS it is convenient to picture the wavevector $\vec{\boldsymbol{k}}$ as a "ray" of light (though we can adopt the same notion for any kind of wave) that propagates along a straight line like a billiard ball. In fact, the analogy between $\vec{\boldsymbol{k}}$ and the *momentum* $\vec{\boldsymbol{p}}$ of a *particle* is more than just a metaphor, as we shall see later. However, for now it will suffice to borrow this imagery from Newton and company, who used it very effectively in describing the *corpuscular* theory of light.[22]

[21] This is sort of like the mathematical assertion that all lines look straight if we look at them through a powerful enough microscope.

However, *near* any *localized source* of waves the outgoing wavefronts are nothing like plane waves; if the dimensions of the source are *small compared to the wavelength* then the outgoing waves look pretty much like SPHERICAL WAVES. For sources similar in size to λ, things can get very complicated.

Christian Huygens (1629-1695) invented the following gimmick for constructing actual wavefronts from spherical waves:

HUYGENS' PRINCIPLE:

> "All points on a wavefront can be considered as *point sources* for the production of *spherical secondary wavelets*. At a later time, the new position of the wavefront will be the *surface of tangency* to these secondary wavelets."

This may be seen to make some sense (try it yourself) but its profound importance to our qualitative understanding of the behaviour of light was really brought home by Fresnel (1788-1827), who used it to explain the phenomenon of *diffraction*, which we will discuss shortly. But first, let's familiarize ourselves with the simpler phenomena of *interference*.

14.13 Interference

To get more quantitative about this "addition of amplitudes," we make the following assumption, which is crucial for the arguments to follow and is even *valid* for the most important

[22] "Corpuscles" are hypothetical *particles* of light that follow trajectories Newton called "rays," thus starting a long tradition of naming every new form or radiation a "ray."

kinds of waves, namely EM waves, under all but the most extreme conditions:

Linear Superposition of Waves:

As several waves pass the same point in space, the total *amplitude* at that point at any instant is simply the *sum* of the amplitudes of the individual waves.

For water waves this is not perfectly true (water waves are very peculiar in many ways) but to a moderately good approximation the amplitude (height) of the surface disturbance at a given position and time is just the sum of the heights of all the different waves passing that point at any instant. This has some alarming implications for sailors! If you are sailing along a coastline with steep cliffs, the incoming swells are apt to be *reflected* back out to sea with some efficiency; if the reflected waves from many parts of the shoreline happen to interfere constructively with the incoming swells at the position of your boat, you can encounter "freak waves" many times higher than the mean swell height. Experienced sailors stay well out from the coastline to avoid such unpredictable interference maxima.

14.13.1 Interference in Time

Suppose we add together two *equal amplitude* waves with slightly different *frequencies*

$$\omega_1 = \bar{\omega} + \delta/2 \quad \text{and} \quad \omega_2 = \bar{\omega} - \delta/2 \qquad (44)$$

where $\bar{\omega}$ is the average frequency and δ is the difference between the two frequencies. If we measure the combined amplitude at a fixed point in space, a little algebra reveals the phenomenon of BEATS. This is usually done with sin or cos functions and a lot of trigonometric identities; let's use the complex notation instead — I find it more self-evident, at least algebraically:

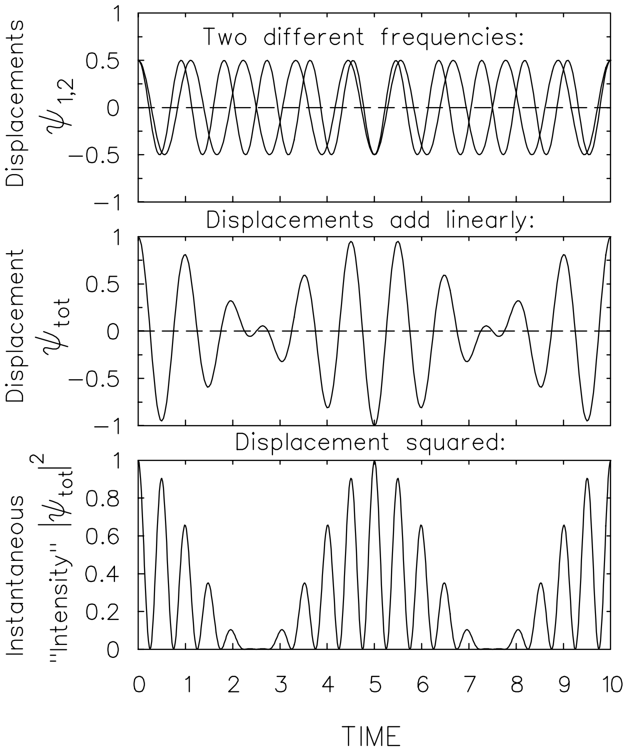

Figure 14.12 Beats.

$$\begin{aligned}
\psi(z,t) &= \psi_0 \left[e^{i\omega_1 t} + e^{i\omega_2 t}\right] \\
&= \psi_0 \left[e^{i(\bar{\omega}+\delta/2)t} + e^{i(\bar{\omega}-\delta/2)t}\right] \\
&= \psi_0 \, e^{i\bar{\omega}t} \left[e^{+i(\delta/2)t} + e^{-i(\delta/2)t}\right] \\
&= 2\,\psi_0 \, e^{i\bar{\omega}t} \cos[(\delta/2)t] \qquad (45)
\end{aligned}$$

That is, the combined signal consists of an oscillation at the *average* frequency, *modulated* by an oscillation at one-half the *difference* frequency. This phenomenon of "BEATS" is familiar to any musician, automotive mechanic or pilot of a twin engine aircraft.

One seemingly counterintuitive feature of BEATS is that the "envelope function" $\cos[(\delta/2)t]$ has only half the angular frequency of the difference between the two original frequencies. What we *hear* when two frequencies interfere is the variation of the sound INTENSITY with time; and the *intensity* is proportional to the *square* of the displacement.[23] Squaring the envelope effectively doubles its frequency (see Fig. 14.12) and so the detected BEAT FREQUENCY is the full frequency difference $\delta = \omega_1 - \omega_2$.

This is a universal feature of waves and interference: the detected signal is the *average intensity*, which is proportional to the *square* of the *amplitude* of the displacement oscillations; and it is the *displacements* themselves that add linearly to form the interference pattern. Be sure to keep this straight.

14.13.2 Interference in Space

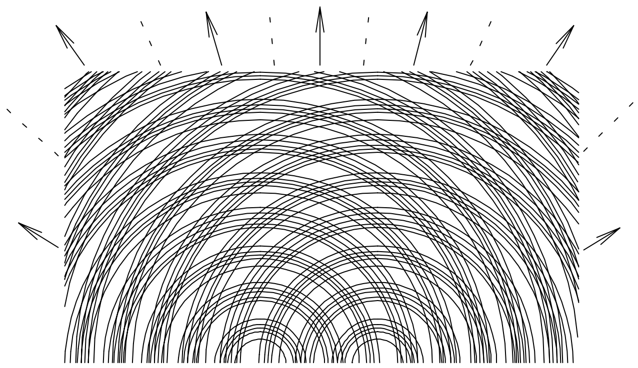

Figure 14.13 A replica of Thomas Young's original drawing (1803) showing the interference pattern created by two similar waves being emitted "in phase" (going up and down simultaneously) from two sources separated by a small distance. The arrows point along lines of constructive interference (crests on top of crests and troughs underneath troughs) and the dotted lines indicate "lines of nodes" where the crests and troughs cancel.

Suppose spherical waves emanate from two *point sources* oscillating *in phase* (one goes "up" at the same time as the other goes "up") at the same frequency, so that the two wave-generators are like synchronized swimmers in water ballet.[24] Each will produce outgoing *spherical waves* that will *interfere* wherever they meet.

The qualitative situation is pictured in Fig.14.13, which shows a "snapshot" of two outgoing spherical[25] waves and the "rays" ($\vec{\boldsymbol{k}}$ directions) along which their peaks and valleys (or crests and troughs, whatever) coincide, giving *constructive interference*. This diagram accompanied an experimental observation by Young of "interference fringes"" (a pattern of intensity maxima and minima on a screen some distance from the two sources) that is generally regarded as the final proof of the wave nature of light.[26]

If we want to precisely locate the angles at which constructive interference occurs ("interference maxima") then it is most convenient to think in terms of "rays" ($\vec{\boldsymbol{k}}$ vectors) as pictured in Fig. 14.14.

The mathematical criterion for constructive interference is simply a statement that the dif-

[23] Actually the INTENSITY is defined in terms of the *average* of the square of the displacement over times long compared with the average frequency $\bar{\omega}$. This makes sense as long as the beat frequency $\delta \ll \bar{\omega}$; but if ω_1 and ω_2 differ by an amount $\delta \sim \bar{\omega}$ then it is hard to define what is meant by a "time average". We will just duck this issue.

[24] This notion of being "in phase" or "out of phase" is one of the most archetypal metaphors in Physics. It is so compelling that most Physicists incorporate it into their thinking about virtually everything. A Physicist at a cocktail party may be heard to say, "Yeah, we were 90° out of phase on everything. Eventually we called it quits." This is slightly more subtle than, "... we were 180° out of phase..." meaning diametrically opposed, opposite, cancelling each other, *destructively interfering*. To be "90° out of phase" means to be moving at top speed when the other is sitting still (in SHM, this would mean to have all your energy in *kinetic* energy when the other has it all in *potential* energy) and *vice versa*. The $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$ fields in a linearly polarized EM wave are 90° out of phase, as are the "push" and the "swing" when a *resonance* is being driven (like pushing a kid on a swing) at maximum effect, so in the right circumstances "90° out of phase" can be productive.... Just remember, "in phase" at the point of interest means *constructive interference* (maximum amplitude) and "180° out of phase" at the point of interest means *destructive interference* (minimum amplitude — zero, in fact, if the two waves have equal amplitude).

[25] OK, they are *circular* waves, not spherical waves. *You* try drawing a picture of spherical waves!

[26] Young's classic experiment is in fact the archetype for all subsequent demonstrations of wave properties, as shall be seen in the Chapter(s) on QUANTUM MECHANICS.

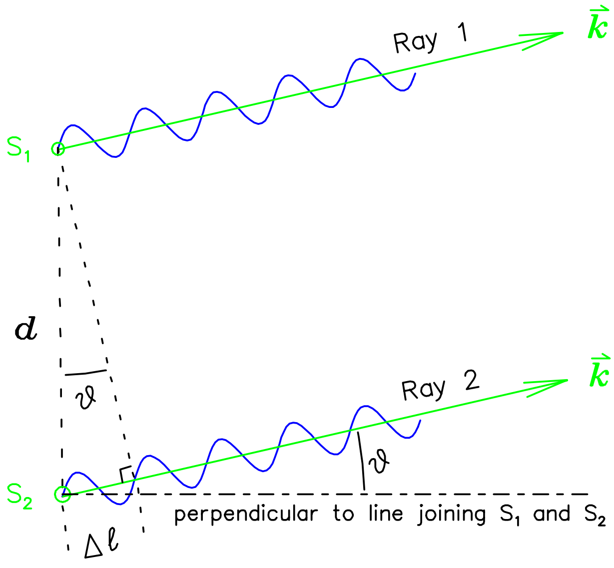

Figure 14.14 Diagram showing the condition for *constructive interference* of two "rays" of the same frequency and wavelength λ emitted in phase from two sources separated by a distance d. At angles for which the difference in path length $\Delta\ell$ is an integer number (m) of wavelengths, $m\lambda$, the two rays arrive at a distant detector in phase so that their amplitudes add constructively, maximizing the intensity. The case shown is for $m = 1$.

ference in path length, $\Delta\ell = d \sin\vartheta_m$, for the two "rays" is an integer number m of wavelengths λ, where the $_m$ subscript on ϑ_m is a reminder that this will be a different angle for each value of m:

$$\boxed{d \sin\vartheta_m \;=\; m\,\lambda} \, . \tag{46}$$

(criterion for CONSTRUCTIVE INTERFERENCE)

Conversely, if the path length difference is a *half-integer* number of wavelengths, the two waves will arrive at the distant detector exactly *out of phase* and cancel each other out. The angles at which this happens are given by

$$\boxed{d \sin\vartheta_m^{\text{destr}} \;=\; \left(m + \frac{1}{2}\right)\,\lambda} \, . \tag{47}$$

(criterion for DESTRUCTIVE INTERFERENCE)

Phasors

What happens when coherent light comes through more than two slits, all equally spaced a distance d apart, in a line parallel to the incoming wave fronts? The same criterion (46) still holds for completely *con*structive interference (what we will now refer to as the PRINCIPAL MAXIMA) but (47) is no longer a reliable criterion for *de*structive interference: each successive slit's contribution cancels out that of the adjacent slit, but if there are an *odd number of slits*, there is still one left over and the combined amplitude is not zero.

Does this mean there are *no* angles where the intensity goes to zero? Not at all; but it is not quite so simple to locate them. One way of making this calculation easier to visualize (albeit in a rather abstract way) is with the geometrical aid of PHASORS: A single wave

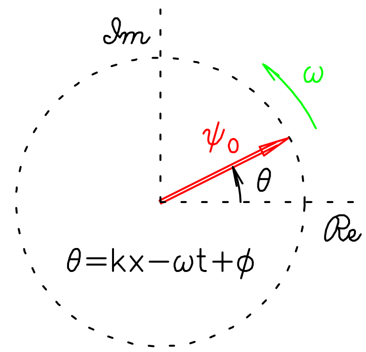

Figure 14.15 A single "PHASOR" of length ψ_0 (the wave amplitude) precessing at a frequency ω in the complex plane.

can be expressed as $\psi(x,t) = \psi_0 e^{i\theta}$ where $\theta = kx - \omega t + \phi$ is the *phase* of the wave at a fixed position x at a given time t. (As usual, ϕ is the "initial" phase at $x = 0$ and $t = 0$. At this stage it is usually ignored; I just retained it one last time for completeness.) If we focus our attention on one particular location in space, this single wave's "displacement" ψ at that location can be represented geometrically as a vector of length ψ_0 (the wave amplitude) in the complex plane called a "PHASOR" As time passes, the "direction" of the phasor rotates at an angular frequency ω in that abstract plane.

There is not much advantage to this geometrical description for a *single* wave (except perhaps that it engages the right hemisphere of the brain a little more than the algebraic expression) but when one goes to "add together" two or more waves with *different phases*, it helps a lot! For example, two waves of equal

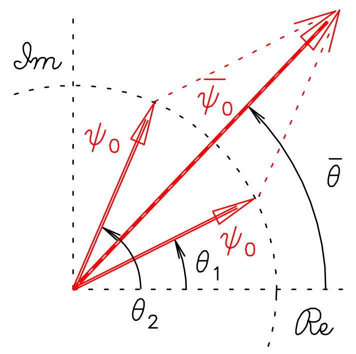

Figure 14.16 Two waves of equal amplitude ψ_0 but different phases θ_1 and θ_2 are represented as PHASORS in the complex plane. Their vector sum has the resultant amplitude $\bar{\psi}_0$ and the average phase $\bar{\theta}$.

amplitude but different phases can be added together algebraically as in Eq. (45)

$$\begin{aligned} \bar{\psi} &= \psi_0 \left[e^{i\theta_1} + e^{i\theta_2} \right] \\ &= 2\,\psi_0\, e^{i\bar{\theta}} \cos(\delta/2) \\ &= \bar{\psi}_0\, e^{i\bar{\theta}} \end{aligned} \tag{48}$$

where

$$\begin{aligned} \bar{\psi}_0 &= 2\,\psi_0 \cos(\delta/2) \\ \bar{\theta} &\equiv \tfrac{1}{2}(\theta_1 + \theta_2) \\ \delta &\equiv \theta_2 - \theta_1\,. \end{aligned} \tag{49}$$

That is, the combined amplitude $\bar{\psi}_0$ can be obtained by adding the phasors "tip-to-tail" like ordinary vectors. Like the original components, the whole thing continues to precess in the complex plane at the common frequency ω.

We are now ready to use PHASORS to find the amplitude of an arbitrary number of waves of arbitrary amplitudes and phases but a common frequency and wavelength interfering at a given position. This is illustrated in Fig. 14.17 for 5 phasors. In practice, we rarely attempt

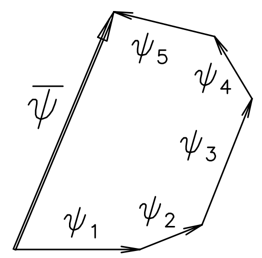

Figure 14.17 The net amplitude of a wave produced by the interference of an arbitrary number of other waves of the same frequency of arbitrary amplitudes ψ_j and phases θ_j can in principle be calculated geometrically by "tip-to-tail" vector addition of the individual PHASORS in the complex plane.

such an arbitrary calculation, since it cannot be simplified algebraically.

Instead, we concentrate on simple combinations of waves of equal amplitude with well defined phase differences, such as those produced by a regular array of parallel slits with an equal spacing between adjacent slits. Figure 14.18 shows an example using 6 identical slits with a spacing $d = 100\lambda$. The angular width of the interference pattern from such widely spaced slits is quite narrow, only 10 mrad (10^{-2} radians) between principal maxima where all 6 rays are in phase. In between the principal maxima there are 5 minima and 4 secondary maxima; this can be generalized:

> The interference pattern for N equally spaced slits exhibits $(N-1)$ *minima* and $(N-2)$ *secondary maxima* between each pair of principal maxima.

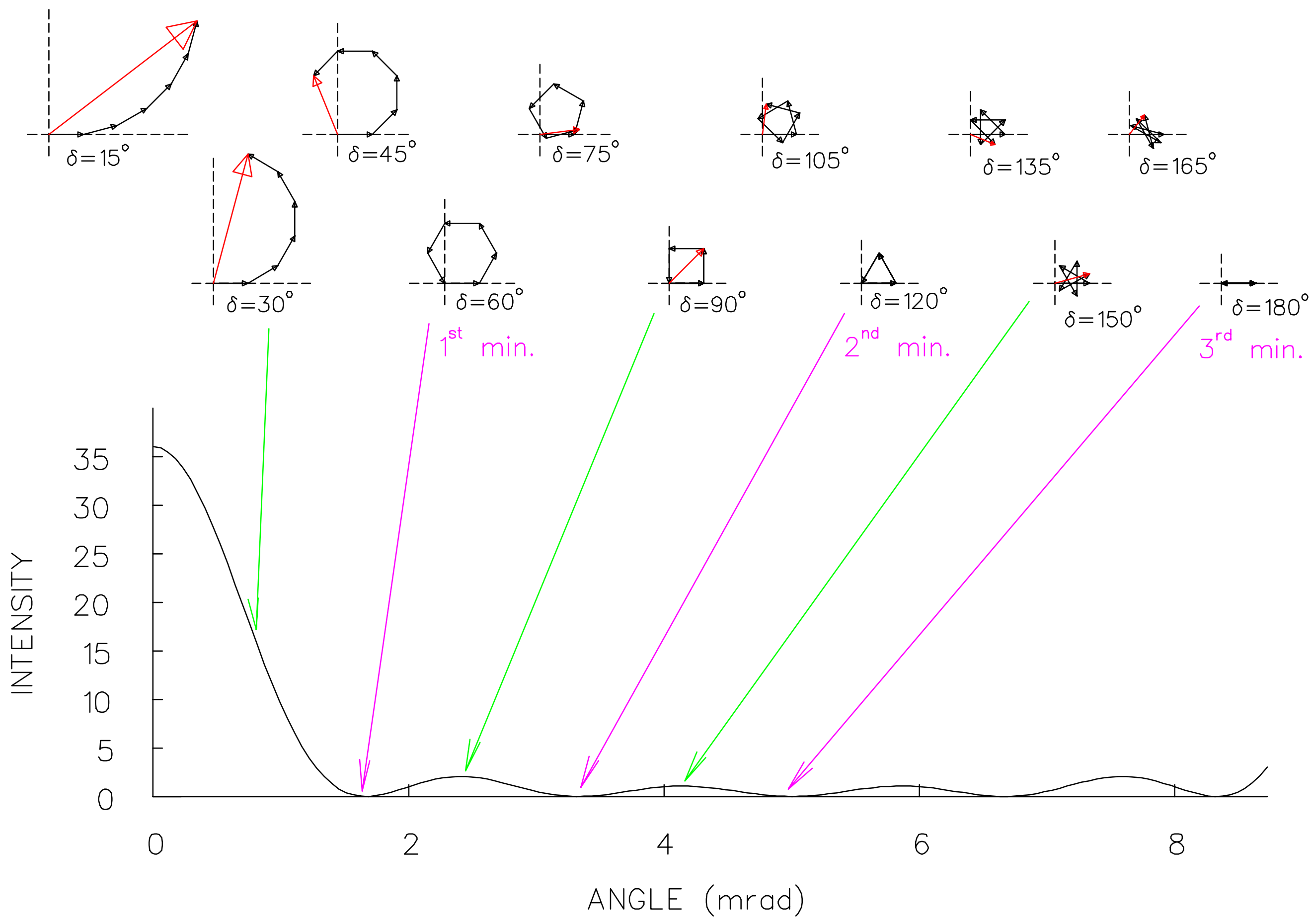

Figure 14.18 The intensity pattern produced by the interference of coherent light passing through six parallel slits 100 wavelengths apart. PHASOR DIAGRAMS are shown for selected angles. Note that, while the *phase* angle difference δ between rays from adjacent slits is a monotonically increasing function of the angle ϑ (plotted horizontally) that the rays make with the "forward" direction, the latter is a real geometrical angle in space while the former is a pure abstraction in "phase space". The exact relationship is $\delta/2\pi = (d/\lambda)\,\sin\vartheta \approx (d/\lambda)\,\vartheta$ for very small ϑ. Note the symmetry about the $3^{\rm rd}$ minimum at $\vartheta \approx 5$ mrad. At $\vartheta \approx 10$ mrad the intensity is back up to the same value it had in the central maximum at $\vartheta = 0$; this is called the first PRINCIPAL MAXIMUM. Then the whole pattern repeats. . . .

It may be conceptually helpful to show the geometrical explanation of the 6-slit interference pattern in Fig. 14.18 in terms of phasor diagrams, but clearly the smooth curve shown there is not the result of an infinte number of geometrical constructions. It comes from an algebraic formula that we can derive for an arbitrary angle ϑ and a corresponding phase difference $\delta = (2\pi d/\lambda)\sin\vartheta$ between rays from adjacent slits. The formula itself is obtained by analysis of a geometrical construction like that illustrated in Fig. 14.19 for 7 slits, each of which contributes a wave of amplitude a, with a phase difference of δ between adjacent slits.

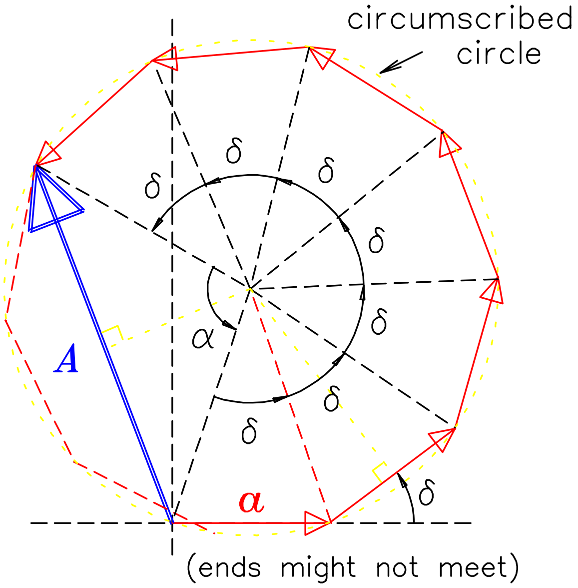

Figure 14.19 PHASOR DIAGRAM for calculating the intensity pattern produced by the interference of coherent light passing through 7 parallel, equally spaced slits.

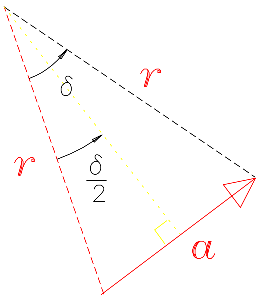

Figure 14.20 Blowup of one of the isosceles triangles formed by a single phasor and two radii from the center of the circumscribed circle to the tip and tail of the phasor.

After adding all 7 equal-length phasors in Fig. 14.19 "tip-to-tail", we can draw a vector from the starting point to the tip of the final phasor. This vector has a length A (the net amplitude) and makes a chord of the circumscribed circle, intercepting an angle

$$\alpha = 2\pi - N\,\delta\ , \tag{50}$$

where in this case $N = 7$. The radius r of the circumscribed circle is given by

$$\frac{a}{2} = r\sin\left(\frac{\delta}{2}\right)\ , \tag{51}$$

as can be seen from the blowup in Fig. 14.20; this can be combined with the analogous

$$\frac{A}{2} = r\sin\left(\frac{\alpha}{2}\right) \tag{52}$$

to give the net amplitude

$$A = a\left[\frac{\sin\left(\frac{\alpha}{2}\right)}{\sin\left(\frac{\delta}{2}\right)}\right]\ . \tag{53}$$

From Eq. (50) we know that $\alpha/2 = \pi - N\,\delta/2$, and in general $\sin(\pi - \theta) = \sin\theta$, so

$$\boxed{A = a\left[\frac{\sin\left(N\frac{\delta}{2}\right)}{\sin\left(\frac{\delta}{2}\right)}\right]} \tag{54}$$

where

$$\boxed{\delta = 2\pi\left(\frac{d}{\lambda}\right)\sin\vartheta} \tag{55}$$

Although the drawing shows $N = 7$ phasors, this result is valid for an arbitrary number N of equally spaced and evenly illuminated slits.