UCD: PHYSICS 9B WAVES, SOUND, OPTICS, THERMODYNAMICS, AND FLUIDS

Tom Weideman
University of California, Davis

LibreTexts™

University of California, Davis

UCD: Physics 9B Waves, Sound, Optics, Thermodynamics, and Fluids

Tom Weideman

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

TABLE OF CONTENTS

7: Fluid Mechanics

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

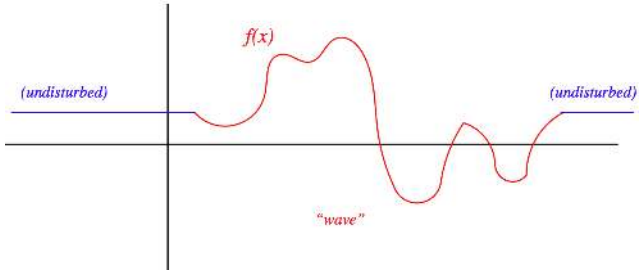
1: Waves

1.1: Wave Mathematics

Definition of a Wave

A *wave* is a disturbance that propagates through space at a constant speed. With one notable exception that we will encounter later, this "disturbance" consists of a fluctuation in the ambient condition of a medium. Waves in one dimension maintain a consistent *waveform* as they propagate (later we will see why this is not so for waves in two and three dimensions). Let's see how we can model this mathematically. We'll start with a localized disturbance frozen in time (think of it as a snapshot) that we describe with a function $f(x)$:

Figure 1.1.1 – Snapshot of a Wave



This is the waveform, but to be a wave, it needs to be propagating along the x-axis, which would make it a function of both x and t. To turn it into such a function, we first have to think about how a function can be shifted along the x-axis. This is accomplished by replacing x in the argument of the function with the sum or difference of x and the value of the shift. If one wishes to shift the function $f(x)$ in the $+x$ direction by a distance a, then the proper change is to the function $f(x-a)$. Note that *subtracting* a in the argument shifts the function in the *positive* x direction, and adding the constant shifts it in the negative x direction. We insist that the wave moves at a constant speed, so we want the wave form to shift by the same distance every time the same time interval passes. We therefore have that the general form of a wave function is:

$$f(x,t) = f(x \pm vt) \tag{1.1.1}$$

This represents a waveform $f(x)$ propagating in the $\mp x$ direction with a speed v.

There are countless functions of x and t that we can come up with, but not all can be written in the form described above. When faced with an arbitrary function of x and t, it can be challenging to determine whether the function represents a wave.

Example 1.1.1

Determine which (if any) of the functions below represent a traveling wave. For those that do, determine the direction of their propagation, and their speed. In every case the constants α and β are positive numbers.

a. $f(x,t) = (1-\alpha x+\beta t)^3 + (2+\alpha x-\beta t)^5$
b. $f(x,t) = \sin\left[(\alpha x)^2 - (\beta t)^2\right]$
c. $f(x,t) = e^{-\alpha x}e^{-\beta t}$

Solution

The idea here is to do whatever algebra that is necessary to get the function into the form $f(x \pm vt)$ ...

a. If we factor $-\alpha^3$ out of the first term, and α^5 out of the second term, we have:

$$f(x,t) = -\alpha^3\left(-\frac{1}{\alpha}+\left(x-\frac{\beta}{\alpha}t\right)\right)^3 + \alpha^5\left(\frac{2}{\alpha}+\left(x-\frac{\beta}{\alpha}t\right)\right)^5$$

We can see that this is purely a function of $\left(x-\frac{\beta}{\alpha}t\right)$. In such problems, it might help to substitute z for $(x \pm vt)$ and show that there are no x's or t's left over. The resulting function $f(z)$ is in fact the waveform. For this case:

$$f(z) = (1-\alpha z)^3 + (2+\alpha z)^5$$

This is therefore a traveling wave moving in the $+x$ direction (because of the opposite signs of x and t), and the speed must be $\frac{\beta}{\alpha}$.

b. At first glance this might appear to be the function of a traveling wave, but we can show that in fact it cannot be written in the correct form. Writing the difference of two squares as a product gives:

$$f(x,t) = \sin[(\alpha x+\beta t)(\alpha x-\beta t)]$$

Each of the factors has the right form (once a factor of α is divided out), but if we substitute z for one of them, we cannot similarly eliminate the second factor. Put another way, one of the factors indicates the wave is moving in the $+x$ direction, while the other indicates it is moving the opposite way. It can't be doing both, so this is not the equation of a traveling wave.

c. Combining the exponentials gives:

$$f(x,t) = e^{-\alpha x - \beta t} = e^{-\alpha\left(x + \frac{\beta}{\alpha}t\right)} \quad \Rightarrow \quad f(z) = e^{-\alpha z}$$

Clearly this represents a wave propagating in the $-x$ direction with a speed of $\frac{\beta}{\alpha}$.

The Wave Equation

It seems like there has to be an easier way to determine if a function of x and t represents a wave. It turns out that there is! To see this, let's start with the basic definition above. If we define $g_{\pm}(x,t) \equiv x \pm vt$, then we can write the wave function as $f(g_{\pm})$. Now we can write derivatives of the function with respect to x and t in terms of derivatives with respect to $g_{\pm}$ using the chain rule. Note that these are functions of more than one variable, so we need to use *partial* derivatives. These work precisely like ordinary derivatives, except that when the derivative is taken with respect to one variable, all the other variables are treated as constants.

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{\partial f}{\partial g_{\pm}}\frac{\partial g_{\pm}}{\partial x} = \frac{\partial f}{\partial g_{\pm}} \quad &\Rightarrow \quad \frac{\partial^2 f}{\partial x^2} &= \frac{\partial^2 f}{\partial g_{\pm}^2} \\ \frac{\partial f}{\partial t} &= \frac{\partial f}{\partial g_{\pm}}\frac{\partial g_{\pm}}{\partial t} = \pm v\frac{\partial f}{\partial g_{\pm}} \quad &\Rightarrow \quad \frac{\partial^2 f}{\partial t^2} &= v^2\frac{\partial^2 f}{\partial g_{\pm}^2} \end{aligned} \tag{1.1.2}$$

Putting these together gives us a relation between second derivatives known as the *wave equation*:

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} \tag{1.1.3}$$

This second order partial differential equation holds if and only if the function behaves like a traveling wave (or a linear combination of traveling waves) with speed v.

Example 1.1.2

For the functions in the previous example, demonstrate whether they do or do not satisfy the wave equation with the proper wave speed.

Solution

We found that the formulas for cases (a) and (c) represent waves, so we plug those into the wave equation:

(a)

$$\frac{\partial^2}{\partial x^2} f(x,t) = \frac{\partial^2}{\partial x^2}\left[(1-\alpha x+\beta t)^3 + (2+\alpha x-\beta t)^5\right] = \frac{\partial}{\partial x}\left[-3\alpha(1-\alpha x+\beta t)^2 + 5\alpha(2+\alpha x-\beta t)^4\right] = 6\alpha^2(1-\alpha x+\beta t) + 20\alpha^2(2+\alpha x-\beta t)^3$$

$$\frac{\partial^2}{\partial t^2} f(x,t) = \frac{\partial^2}{\partial t^2}\left[(1-\alpha x+\beta t)^3 + (2+\alpha x-\beta t)^5\right] = \frac{\partial}{\partial t}\left[3\beta(1-\alpha x+\beta t)^2 - 5\beta(2+\alpha x-\beta t)^4\right] = 6\beta^2(1-\alpha x+\beta t) + 20\beta^2(2+\alpha x-\beta t)^3$$

From direct comparison, it is clear that these two terms are proportional, which means they satisfy the wave equation:

$$\frac{\partial^2}{\partial x^2} f(x,t) = \frac{\alpha^2}{\beta^2}\frac{\partial^2}{\partial t^2} f(x,t)$$

The constant of proportionality for the wave equation is $\frac{1}{v^2}$, so this confirms that $v = \frac{\beta}{\alpha}$.

(b)

$$\frac{\partial^2}{\partial x^2} f(x,t) = \frac{\partial^2}{\partial x^2}\sin\left[(\alpha x)^2 - (\beta t)^2\right] = \frac{\partial}{\partial x}\left\{2\alpha^2 x\cos\left[(\alpha x)^2 - (\beta t)^2\right]\right\} = 2\alpha^2\cos\left[(\alpha x)^2 - (\beta t)^2\right] - 4\alpha^4 x^2\sin\left[(\alpha x)^2 - (\beta t)^2\right]$$

$$\frac{\partial^2}{\partial t^2} f(x,t) = \frac{\partial^2}{\partial t^2}\sin\left[(\alpha x)^2 - (\beta t)^2\right] = \frac{\partial}{\partial t}\left\{-2\beta^2 t\cos\left[(\alpha x)^2 - (\beta t)^2\right]\right\} = -2\beta^2\cos\left[(\alpha x)^2 - (\beta t)^2\right] - 4\beta^4 t^2\sin\left[(\alpha x)^2 - (\beta t)^2\right]$$

These two terms are clearly not proportional, so this function does not satisfy the wave equation.

(c)

$$\frac{\partial^2}{\partial x^2} f(x,t) = \frac{\partial^2}{\partial x^2}\left[e^{-\alpha x} e^{-\beta t}\right] = \frac{\partial}{\partial x}\left[-\alpha e^{-\alpha x} e^{-\beta t}\right] = \alpha^2 e^{-\alpha x} e^{-\beta t}$$

$$\frac{\partial^2}{\partial t^2} f(x,t) = \frac{\partial^2}{\partial t^2}\left[e^{-\alpha x} e^{-\beta t}\right] = \frac{\partial}{\partial t}\left[-\beta e^{-\alpha x} e^{-\beta t}\right] = \beta^2 e^{-\alpha x} e^{-\beta t}$$

These two terms are proportional, and the constant of proportionality gives the correct velocity once again.

Waves in Two and Three Dimensions

Consider a two-dimensional wave, such as a ripple radiating outward from a pebble dropped into a still lake. If the distance of the wave front from the source is r, following the "wave form remains unchanged" prescription, the functional form of the traveling wave is $f(r,t) = f(r - vt)$. Alternatively, we can extend the wave equation to two or three dimensions as follows:

$$\begin{aligned} &\textit{two dimensions}: && \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} \\ &\textit{three dimensions}: && \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} \end{aligned} \tag{1.1.4}$$

In the one-dimensional case, we showed above that the wave form that remains unchanged as it propagates is described by a function that satisfies the wave equation. It turns out that in the two-dimensional case, this is no longer true. We can show this by repeating the procedure outlined in Equation 1.1.2. A waveform that remains unchanged as it spreads radially outward would have the form $f(r,t) = f(r - vt)$ (we are considering an outgoing waveform, which accounts for the minus sign). Defining the function $g(r,t) \equiv r - vt$, we have for the derivative of the wave function with respect to x:

$$\frac{\partial f}{\partial x} = \left(\frac{\partial f}{\partial g}\right)\left(\frac{\partial g}{\partial x}\right) = \left(\frac{\partial f}{\partial g}\right)\left(\frac{\partial g}{\partial r}\right)\left(\frac{\partial r}{\partial x}\right) \tag{1.1.5}$$

Note that the variable r depends upon both x and y, specifically:

$$r = \sqrt{x^2 + y^2} = \left(x^2 + y^2\right)^{\frac{1}{2}} \quad \Rightarrow \quad \frac{\partial r}{\partial x} = \frac{1}{2}\left(x^2 + y^2\right)^{-\frac{1}{2}}(2x) = \frac{x}{r}, \quad \text{and} \quad \frac{\partial r}{\partial y} = \frac{y}{r} \tag{1.1.6}$$

Plugging this and $\dfrac{\partial g}{\partial r} = 1$ in above gives:

$$\begin{aligned} \frac{\partial f}{\partial x} = \frac{\partial f}{\partial g}\left(\frac{x}{r}\right) \quad &\Rightarrow \quad \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial g^2}\left(\frac{x^2}{r^2}\right) + \frac{\partial f}{\partial g}\left(\frac{1}{r}\right)\left(1 - \frac{x^2}{r^2}\right) \\ \frac{\partial f}{\partial y} = \frac{\partial f}{\partial g}\left(\frac{y}{r}\right) \quad &\Rightarrow \quad \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 f}{\partial g^2}\left(\frac{y^2}{r^2}\right) + \frac{\partial f}{\partial g}\left(\frac{1}{r}\right)\left(1 - \frac{y^2}{r^2}\right) \end{aligned} \tag{1.1.7}$$

The dependence on t is the same as in the one-dimensional case:

$$\frac{\partial f}{\partial g} = -\frac{1}{v}\frac{\partial f}{\partial t} \quad \Rightarrow \quad \frac{\partial^2 f}{\partial g^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} \tag{1.1.8}$$

Plugging this into the previous equations, and adding them together gives:

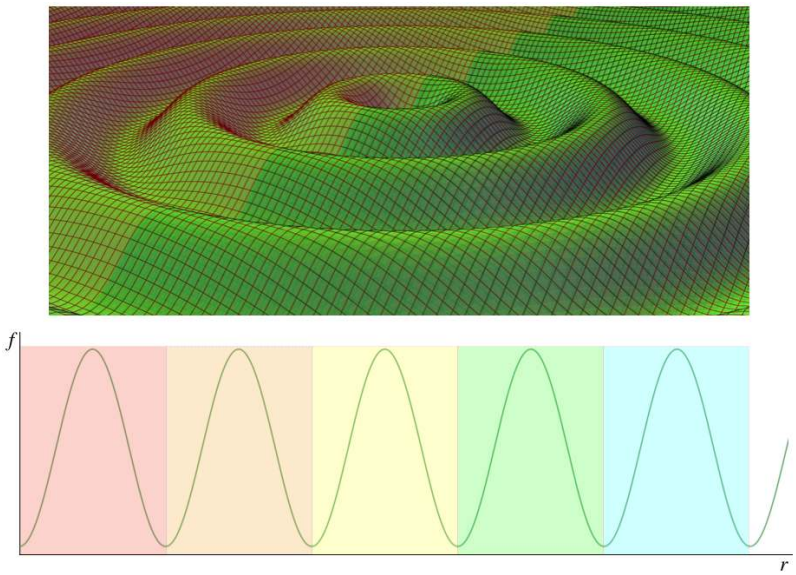
$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}\left(\frac{x^2}{r^2} + \frac{y^2}{r^2}\right) - \frac{1}{v}\frac{\partial f}{\partial t}\left(\frac{1}{r}\right)\left(2 - \frac{x^2}{r^2} - \frac{y^2}{r^2}\right) \tag{1.1.9}$$

Noting that $x^2 + y^2 = r^2$, we finally get:

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} - \frac{1}{v}\frac{\partial f}{\partial t}\left(\frac{1}{r}\right) \tag{1.1.10}$$

The second term on the right hand side of the equation clearly makes this differential equation look different from the one-dimensional wave equation. So the question is, which is the correct way of describing a two-dimensional wave? Does it maintain its wave form as it propagates outward, or does it satisfy our previous wave equation extended to two dimensions? The figures below display the two possibilities we are talking about.

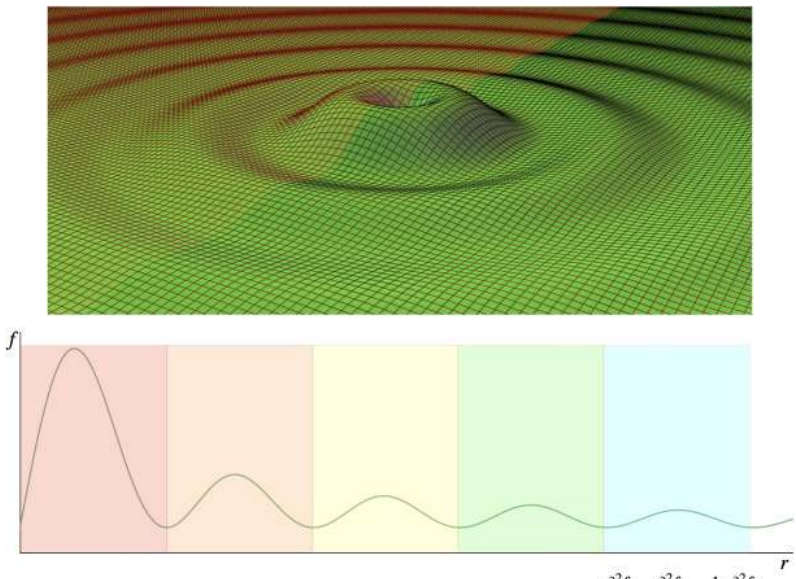
Figure 1.1.2 – Two-Dimensional Circularly-Radiating Function with Unchanging Waveform



wave form remains the same as wave moves away from center: $f(r,t) = f(r - vt)$

The graph shows a cross-sectional snapshot of the wave – the waveform repeats as a function of r.

Figure 1.1.3 – Two-Dimensional Circularly-Radiating Function Satisfying 2-D Wave Equation



wave form diminishes in height as wave moves away from center: $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}$

The graph shows a cross-sectional snapshot of the wave – the waveform does not repeat as a function of r.

We can't answer this question purely mathematically – we have to observe what actually happens in nature. As we will see in Section 1.3, conservation of energy will require that it is in fact the extended two-dimensional wave equation that gives the correct answer – two and three-dimensional waveforms *do not remain fixed* as they radiate outward. The wave displacements diminish with distance from the source, as shown in Figure 1.1.3.

Finally, it should be mentioned that a shorthand notation is commonly used for the sum of the three-dimensional spatial double derivatives, which looks like:

$$\nabla^2 f = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} \tag{1.1.11}$$

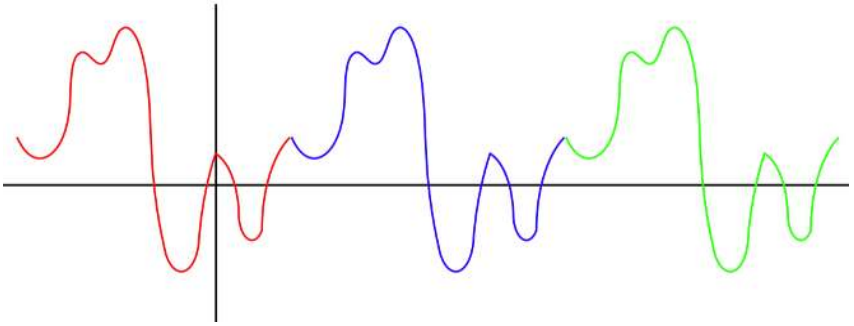
The differential equation given earlier is in cartesian coordinates (x, y, z), but it could just as easily be written in cylindrical (r, θ, z) or spherical (r, θ, ϕ) polar coordinates (though we won't do so here). This shorthand form avoids committing to a coordinate system altogether.

1.2: Wave Properties

Periodic Waves

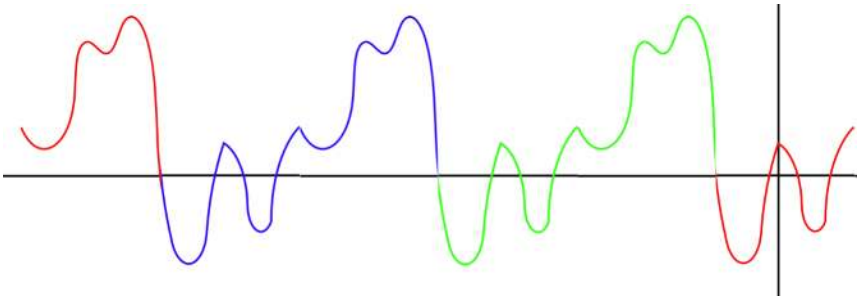
There are qualities that are not required of general waves which are nonetheless common features of waves encountered in nature. The most common special characteristic of a wave is when it continually repeats a specific waveform as it propagates. Such a wave is said to be *periodic*. There are a couple ways to determine if a wave is periodic. The first is to take a snapshot of the wave, and see if its waveform is repeated in space:

Figure 1.2.1a – Snapshot of a Periodic Wave



It should be noted that the starting point of each waveform in the diagram above was chosen arbitrarily. That is, if we look at the same snapshot of the wave as above, we could just as easily demonstrate its periodic nature with different segments:

Figure 1.2.1b – Snapshot of a Periodic Wave



The second way to determine if a wave is periodic is mathematical. The function repeats itself upon translation by a certain distance in the $\pm x$ direction. That is:

$$f\left(x \pm vt\right) = f\left(x \pm vt \pm n\lambda\right), \qquad n = 0,\ 1,\ 2, \ldots \tag{1.2.1}$$

The quantity λ is the length of the repeating waveform, and is called the *wavelength* of the wave. A glance at the two diagrams above should make it clear that the wavelength is a universal feature of that particular wave, and does not depend upon where we choose the starting point to be.

The snapshot of the wave tells us something about its spatial features, but the wave is moving, so if we want to know something about its time-dependence, we need to select a specific point in space, and observe the displacement of the medium as the wave goes by. The wave moves at a constant speed, and the length of each repeating waveform is the same, so the time span required for a single waveform to go by is a constant for the entire wave, called the *period* of the wave. An alternative way of measuring the temporal feature of the wave is the rate at which medium displacements repeat, called *frequency*. Frequency is measured in units of cycles per second, a unit known as *hertz* (*Hz*). Since 1 period is the time required for one cycle, there is a simple relationship between these quantities:

$$f = \frac{1}{T} \tag{1.2.2}$$

We can make another association of periodic wave properties. If we pick a specific point on a waveform (called a *point of fixed phase* for the wave), and follow its motion, it should be clear that it travels a full wavelength in the time of one period. We therefore can relate the wave speed, wavelength, and period (or frequency):

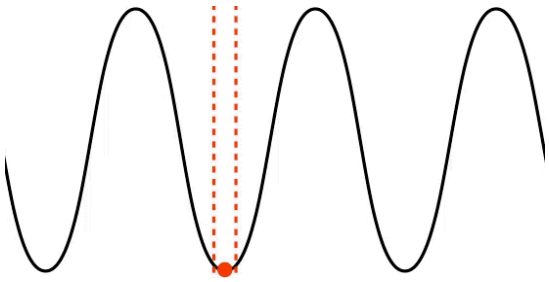
$$v = \frac{\lambda}{T} = \lambda f \tag{1.2.3}$$

Wave Polarization

While the disturbance is not always a displacement of a medium, it always has a directional element to it. A wave that actually displaces a medium has an obvious direction: that of the displacement. Other waves have directional gradients that signify a direction. The direction in which the pressure is changing fastest (the pressure gradient direction) defines a direction for sound waves, and the direction of the electric field vectors defines a direction for light. This directional aspect of waves is also given a name: *polarization*. Generally the direction of medium displacement or gradient is compared to the direction of the wave's motion. There are two special cases that we will encounter for polarization of a wave:

transverse polarization: the medium's displacement or gradient is perpendicular to the wave's direction of motion

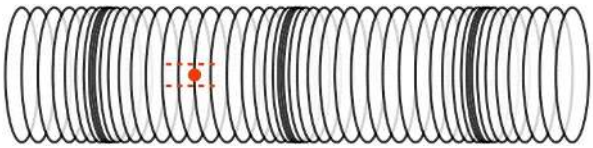
Figure 1.2.2 – Transverse Wave



Note that the displacement of a single point in the medium (depicted by the red dot) is moving only vertically, while the wave moves horizontally. That these two motions are perpendicular to each other is the defining characteristic of a transversely polarized wave. Waves on strings and surface water waves are examples of this kind of wave. As noted earlier, not all waves involve the medium displacing (we will see some examples where this is the case later), but whatever fluctuation is occurring has a direction that can be compared with the direction of the wave's motion.

longitudinal polarization: the medium's displacement or gradient is parallel to the wave's direction of motion.

Figure 1.2.3 – Longitudinal Wave



This time the displacement of a single point in the medium is parallel to the direction of the motion of the wave, the defining characteristic of a longitudinally polarized wave. Notice that like any other wave, the medium is not traveling with the wave, it is moving back-and forth. Physically these are waves induced by *compressions* (regions where the medium is more dense) and *rarefactions* (regions where the medium is less dense). These kinds of waves can be created in springs (as depicted above), but the most common physical example of this kind of wave is sound. Any medium (solid, liquid, or gas) will react to compression, and will therefore exhibit this kind of wave.

Alert

Snapshot graphs of waves of both kinds of polarization are sketched graphically with the displacement on the vertical axis and the position on the horizontal axis. When this is done, it "looks like" a transverse wave, but it is important to keep in mind that such a graph is not a picture of the wave. The vertical axis measures the displacement of the medium from the equilibrium point, which in the case of the red dot on the spring coil for the longitudinal wave in Figure 1.2.3 is the center of the horizontal dotted red lines.

Harmonic Waves

In the category of periodic waves, the easiest to work with mathematically are *harmonic waves*. The word "harmonic" is basically synonymous with "sinusoidal." For a one-dimensional wave, one might therefore assume that a harmonic wave function looks like:

$$f(x,t) = A\cos(x \pm vt) \tag{1.2.4}$$

While this comes close, it has a problem with units. The *total phase* of the wave function (the part in parentheses that is the argument of the cosine) cannot have any physical units, and this function has a phase with units of length. We can therefore repair this problem by dividing the phase by a constant of the wave that has units of length. The obvious such constant is the wavelength. So now our candidate wave function is:

$$f(x,t) = A\cos\left[\frac{1}{\lambda}(x \pm vt)\right] \tag{1.2.5}$$

This gets close, but if we are using radians as the measurement of phase, there is one more change we must add. If we consider a snapshot of this wave at $t = 0$, we would find that the sinusoidal waveform should repeat itself every time the value of x is displaced by λ. If we are using radians as our angular measure, then this requires multiplying the phase by 2π. Then every change of x by λ will result in a change in the phase by 2π, and the function repeats itself properly. So we now have:

$$f(x,t) = A\cos\left[\frac{2\pi}{\lambda}(x \pm vt)\right] \tag{1.2.6}$$

There is one final addition to the phase that we need to make. Suppose we take a snapshot of the wave at $t = 0$ and look at the origin, $x = 0$. This function tells us that the value of the wave's displacement must be its maximum: A. This is not a very general wave! To account for the possibility that the wave might have a different initial condition at the origin, we need to include a *phase constant*, ϕ. Distributing the factor of $\frac{2\pi}{\lambda}$, and using Equation 1.2.3, we get the final form of the wave function of a 1-dimensional harmonic wave:

$$f(x,t) = A\cos\left(\frac{2\pi}{\lambda}x \pm \frac{2\pi}{T}t + \phi\right) \tag{1.2.7}$$

It is common to write this wave function in more compact ways. The first involves the definition of the *wave number* k, and *angular frequency* ω:

$$k \equiv \frac{2\pi}{\lambda}, \quad \omega \equiv 2\pi f = \frac{2\pi}{T} \quad \Rightarrow \quad f(x,t) = A\cos(kx \pm \omega t + \phi) \tag{1.2.8}$$

Another definition that saves even more space is lumping the total phase of the wave into a single function variable: $\Phi(x,t)$. It is clearly linear in the variables x and t. That is:

$$f(x,t) = A\cos(\Phi), \qquad \Phi(x,t) = \frac{2\pi}{\lambda}x \pm \frac{2\pi}{T}t + \phi = kx \pm \omega t + \phi \tag{1.2.9}$$

Finally, it should be noted that although the cosine function was arbitrarily chosen here, we could have just as easily chose a sine function. The only difference between representing the wave with these two functions is the phase constant. That is, we can change from one function to the other if we change the phase constant by $\frac{\pi}{2}$:

$$\cos(\Phi) = \sin\left(\Phi + \frac{\pi}{2}\right) \quad \Rightarrow \quad \phi \to \phi + \frac{\pi}{2} \tag{1.2.10}$$

Wave Graphs

The important thing to take away from the harmonic wave function in Equation 1.2.7 is that the wave has four *constants of the motion* that completely define it. Besides the wavelength, period, and phase constant, there is the *amplitude*, A. All of these remain fixed in time, completely defining the wave that evolves thanks to its x and t dependence. It is often useful to use these constants to analyze a wave in parts. For example, we have already discussed analyzing the spatial features of the wave by taking a "snapshot" – a frozen moment in time. This amounts to choosing a value for t (often zero, but not always), so that the wave function now becomes only a function of x.

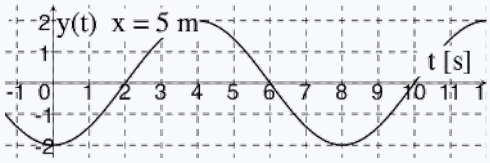
We can also isolate the time variable. In this case, we pick a specific position x, and graph the time dependence of the displacement of that point in the wave. For harmonic waves, these displacement-vs-time graphs represent harmonic oscillation. It should be noted that like the spatial graph, the time graph is a cosine (or sine) function, and this can lead to confusion, as it "looks like" a wave.

What links these two graphs is the motion of the wave. The speed of the wave is related to the wavelength (which can be read off the position graph) and the period (which can be read off the time graph). The direction of the wave's motion gives us the relative

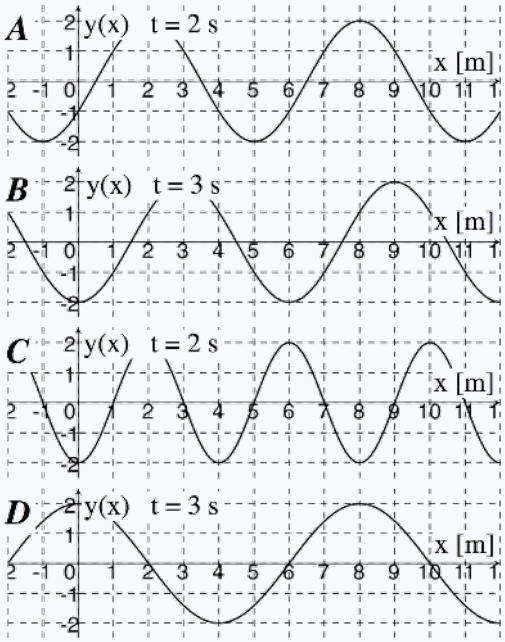
signs of the position and time variables in the wave function (recall that opposite signs means it is moving in the $+x$-direction). Determining the phase constant requires understanding the meaning of "total phase," and some simple algebra. Let's see how this all fits together with an example.

Example 1.2.1

The figure below is a graph of the simple harmonic motion of a particle of string through which an harmonic transverse wave is passing (the displacement is parallel to the y-axis, and the motion is along the x-axis). The position (x-value) of the oscillating particle is $5m$, as indicated on the graph.



a. *One of the four position graphs given below represents a snap-shot of the wave at a given instant in time (the moment in time for each case is indicated on the graph). Find the y-vs-x graph below that belongs to this wave.*



b. *Find the amplitude, wavelength and period of this wave.*
c. *Find the speed at which this wave is traveling.*
d. *Find the direction ($\pm x$) in which this wave is moving.*
e. *Find the phase constant in the range [0, 2π] for the wave function of this wave.*

Solution

a. The one thing that both graphs have in common is the displacement of the particle of string. The harmonic motion graph occurs at the position $x = 5m$, so we can look at what displacement the four prospective graphs give for that position at a common time. Let's start with graph A. The value of the displacement at the position $x = 5m$ of the particle is $y = -2$, and this occurs at time $t = 2s$. Looking back at the harmonic motion graph, we see that the displacement of the particle at $t = 2s$ is $y = 0$, so graph A cannot represent the same wave as the harmonic motion graph. Graph B has a displacement of $y = -1$ at time $t = 3s$. Looking back at the harmonic motion graph, we see that at $t = 3s$ the displacement is between $y = +1$ and $y = +2$, so graph B also does not work. Graph C has a displacement at $x = 5m$, $t = 2s$ that matches that of the harmonic motion graph (i.e. $y = 0$). It's easy to confirm using the logic shown above that graph D does not work, so the answer is graph C. It is important to note that graph C is by no means unique, it is simply the only one that works from the four choices given. But now that we know graph C represents this wave, we know significantly more about it, and we will use this information for the remaining parts of this example.

b. The maximum displacement of the string is the amplitude of the wave, and both graphs must agree on this. They do, and it is a value of 2 (units are not given). We know that graph C is a proper snapshot graph of this wave, which makes the distance between peaks equal to the wavelength, so $\lambda = 4m$. The time interval between peaks on the harmonic motion graph is the period of one oscillation, so $T = 8s$.

c. With the wavelength and period, we can immediately compute the wave speed: $v = \frac{\lambda}{T} = 0.5\frac{m}{s}$.

d. Now things start to get a bit tricky, and a little visualization & logic are required. The one point we know about from these two graphs is $x = 5m, t = 2s$. Now let's consider what happens to that particle of string a short time later. On the harmonic motion graph, we see that if we move to a slightly larger value of t, the displacement becomes positive. The wave must therefore moving in a direction such that a short time later the displacement will rise. Looking at the snapshot of the wave in graph C, we see that the direction we need to shift the wave form such that the displacement immediately starts going up is the $-x$-direction. The wave must therefore be moving in that direction. Note that if the common point between the two graphs happened to be a maximum or minimum of the wave, then it would not be possible to determine the direction of the wave velocity, because the displacement a short time later would be in the same direction no matter which way the wave was moving.

*e. First, it should be noted that there are an infinite number of phase constants, since one can always add or subtract 2π to get a new phase constant that will also work. So our task here is to find any phase constant that works, then add or subtract an appropriate number of units of 2π to get a value that falls within the range required. To solve for a phase constant, one must first understand what the total phase of the wave is. Given that we are using a cosine function, we know that the peak of the wave occurs when the argument of the cosine (i.e. the total phase) is an integer multiplied by 2π. We will make it easy on ourselves by just choosing zero. So we need to find a position and time for which the displacement of the wave is a maximum – **any choice will work**, and as an exercise, the reader should try more points to prove this. For this solution, we will use the peak on the harmonic motion graph at $x = 5m, t = 4s$. Note that we must choose the x and t terms to have the same sign, as we know the wave is moving in the $-x$-direction. We will choose both to have positive signs, but both negative will also work.*

$$\Phi(x,t) = \frac{2\pi}{\lambda}x \pm \frac{2\pi}{T}t + \phi \quad \Rightarrow \quad 0 = \frac{2\pi}{4m}(5m) + \frac{2\pi}{8s}(4s) + \phi \quad \Rightarrow \quad \phi = -\frac{7}{2}\pi$$

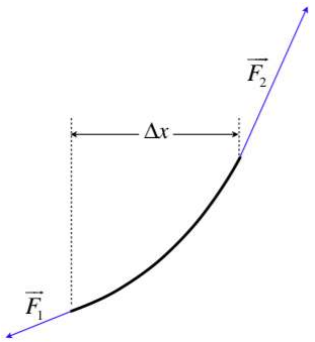
This phase constant doesn't fall within the desired [0, 2π] range, so we add 4π to it and find $\phi = \frac{\pi}{2}$.

Wave Velocity

Up until now, we have simply stated that waves have fixed velocities. What we have not yet considered are the physical conditions that determine the speed that a wave will have. Clearly these conditions must depend only upon the type of wave it is (string, slinky, light, sound, etc.) and the medium through which it is passing. We will use our tools from classical mechanics to look at the simple physical physical system of a transverse wave traveling through a taut string.

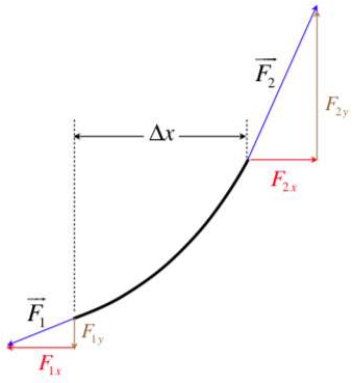
The wave function describes the displacement of a single particle of the string, so we will start with a small segment. The wave form curves the string, so the pulls of tension from each end of an infinitesimal segment of the string are not directly opposite to each other. A free-body diagram of such a segment of length Δx (the bend is exaggerated for the purpose of illustration) looks like this (note that we are ignoring gravity here):

Figure 1.2.4a – Free-Body Diagram of a Segment of String



The two forces can now be broken into horizontal and vertical components.

Figure 1.2.4b – Free-Body Diagram of a Segment of String



Remember that this is a transverse wave, which means that this segment only accelerates vertically. The horizontal components of the forces therefore must cancel, and remain fixed – the magnitude of the horizontal components must therefore be the constant horizontal tension applied to the string.

$$\text{tension} = F = F_{1x} = F_{2x} \tag{1.2.11}$$

Now apply Newton's 2nd law in the vertical direction. Setting up as the positive direction, we have:

$$F_{2y} - F_{1y} = ma_y \tag{1.2.12}$$

The mass of this segment of string is the linear mass density of the string (which we will call μ) multiplied by the segment's length Δx. The vertical acceleration is the second derivative of the y position with respect to time. Putting these into the equation gives:

$$F_{2y} - F_{1y} = (\mu \Delta x)\left(\frac{\partial^2 y}{\partial t^2}\right) \tag{1.2.13}$$

The two forces $\vec{F}_1$ and $\vec{F}_2$ are pulling directly through the string, so their directions are tangent to the curve made by the string on each end. The slope of the curve made by the string is the first derivative of the displacement with respect to x, and this slope is also the ratio of the vertical force to the horizontal force, so:

$$\left.\begin{array}{ll} \text{slope at bottom of segment:} & \left(\frac{\partial y}{\partial x}\right)_1 = \frac{F_{1y}}{F_{1x}} = \frac{F_{1y}}{F} \\ \text{slope at btop of segment:} & \left(\frac{\partial y}{\partial x}\right)_2 = \frac{F_{2y}}{F_{2x}} = \frac{F_{2y}}{F} \end{array}\right\} \quad \Rightarrow \quad F_{2y} - F_{1y} = F\left[\left(\frac{\partial y}{\partial x}\right)_2 - \left(\frac{\partial y}{\partial x}\right)_1\right] \tag{1.2.14}$$

Plugging this back into Equation 1.2.12, we get:

$$F\left[\left(\frac{\partial y}{\partial x}\right)_2 - \left(\frac{\partial y}{\partial x}\right)_1\right] = \mu \Delta x \left(\frac{\partial^2 y}{\partial t^2}\right) \quad \Rightarrow \quad \frac{\left(\frac{\partial y}{\partial x}\right)_2 - \left(\frac{\partial y}{\partial x}\right)_1}{\Delta x} = \frac{\mu}{F}\left(\frac{\partial^2 y}{\partial t^2}\right) \tag{1.2.15}$$

When we require that Δx be very small, the left-hand side of the equation becomes a derivative. There is already on derivative, so the result is a second derivative, giving:

$$\frac{\partial^2 y}{\partial x^2} = \left(\frac{\mu}{F}\right)\frac{\partial^2 y}{\partial t^2} \tag{1.2.16}$$

But this is precisely the wave equation! So we can just read-off the velocity of the wave (recall from Equation 1.1.3 that the coefficient of the second derivative in time is $\frac{1}{v^2}$):

$$\frac{1}{v^2} = \frac{\mu}{F} \quad \Rightarrow \quad v = \sqrt{\frac{F}{\mu}} \tag{1.2.17}$$

Mechanical Wave Speeds in Other Media

Although we derived the result for the very specific case of a taut string, its general features applies to all mechanical waves – there is always an element of the restoring force in the medium (in the string case, the tension), and the inertial of the medium (in the string case, the linear density), and the square root dependence comes out to be universal as well!

We said in Physics 9A that particles near each other (whether a solid, liquid or gas) exert restoring forces on each other (repel when they get closer, attract when they separate). This force ultimately is what is responsible for the wave speed for which we found the simplest case above, described in terms of tension. If such a wave (which, unlike a transverse string wave, is actually a longitudinal compression wave, and goes by the more generic name of "sound") is traveling through a fluid (defined as either a gas or a liquid), then the Van der Waals restoring forces between the particles are summarized in terms of a quantity called the *bulk modulus*. The "inertia" part of the wave speed equation changes from the linear mass density of the (one-dimensional) string to the volume mass density of the (three-dimensional) fluid, making the velocity relation look like:

$$v_{in\ fluid} = \sqrt{\frac{B}{\rho}} \tag{1.2.18}$$

It turns out that this bulk modulus depends not only on the makeup of the fluid, but also its temperature. We won't go into any more detail than this about this physical property of fluids.

These compression waves can also travel through solids, and the Van der Waals restoring forces between particles manifest slightly differently than for fluids (as you can imagine, the "spring constants" between nearby particles in a solid are significantly higher than between particles in a fluid). In this case, the physical property that accounts for the restoring force between particles is called *Young's modulus*. But the resulting wave speed formula looks the same:

$$v_{in\ solid} = \sqrt{\frac{Y}{\rho}} \tag{1.2.19}$$

We will not explore the exact nature of the bulk and Young's moduli, other than to note that they must obviously have the same units. Simply knowing that they play the same role for fluids and solids respectively as the tension plays for a transverse wave on a string will suffice for our purposes.

Digression: Tsunamis

When one reads above "waves in a fluid", one might be inclined to visualize transverse waves, as one sees on the surface of the ocean. But those waves have as their restoring forces gravity and upward pressure. In fact, such surfable waves are known as "gravity waves" (not to be confused with "gravitational waves" which are themselves very different). The waves-through-fluid for which the equation above applies are longitudinal – like a wave on a slinky that compresses and expands the rings. Such waves in the ocean are significantly faster than their gravity wave counterparts. In cases where an underwater earthquake strikes, a great deal of energy can be transmitted into this type of wave, and its high rate of speed can correspond to quite a long wavelength. When such a wave reaches a beach, the longitudinal displacement of the water can therefore be quite large, vacating the water just offshore for several hundred yards, followed by an expansion onshore a similar distance, inundating the land near the beach. These waves are what are known as tsunamis. These are sometimes mistakenly referred to as "tidal waves", but gravitational effects from the sun and moon are responsible for the tides, and these don't play a role in this type of wave. Also, some visualize tsunamis as very tall waves that crash on shore, but again, tsunamis are longitudinal, so their large displacements are horizontal, not vertical.

Summary of Wave Attributes

To conclude this section, we will recap the wave attributes we have seen so far. For a wave to occur, two things are needed – a medium through which the wave passes, and a "driver" which is the initial source of the energy carried by the wave. Let's look at how each of the wave attributes links to these ingredients.

- **wave speed** – We just showed above that the speed of the wave depends entirely upon the medium alone. There is no way to generate the wave at its origin such that it will move faster. While we have only shown this for transverse string waves, it is true in general for all waves.
- **amplitude** – The maximum displacement of an harmonic wave depends upon how much the driver displaces the medium at the source of the wave. We also found that the amplitude can change (get smaller) when the wave is radiating outward from a point

source in 2 or 3 dimensions.

- **period (or frequency)** – The period of the wave is determined by how long it takes the wave's generator to complete a full cycle. The wave's frequency will exactly mimic the frequency at which it is driven.
- **wavelength** – The length of a wave is determined by a combination of its frequency and wave speed. So we cannot point to either the driver or the medium as the direct cause of wavelength. In some sense, wavelength is a "dependent variable," while the independent variables that we can control from outside are wave speed and frequency.
- **polarization** – The displacement of the medium can be perpendicular to the wave motion (transverse) or parallel to it (longitudinal).

Example 1.2.2

The transverse harmonic waves carried on a taut string can be adjusted in three ways:

- *the tension in the string can be increased or decreased*
- *the maximum displacement of the driver can be made larger or smaller*
- *the frequency of oscillation of the driver can be increased or decreased*

After a lot of careful adjustments, a wave is created which has the property that the speed of the wave equals the maximum speed that a tiny segment of the string has as the wave passes through it.

a. *Describe what kinds of changes to the output of the driver (if any) will maintain the equal-speed property above.*
b. *Find the ratio of the amplitude of the wave and its wavelength.*

Solution

a. We start with the wave function:

$$f(x,t) = A\cos\left(\frac{2\pi}{\lambda}x \pm \frac{2\pi}{T}t + \phi\right)$$

This describes the vertical position of a segment of string as a function of the horizontal position and time. The speed of this segment is therefore the first derivative of this function with respect to time:

$$v_{string}(x,t) = \frac{\partial}{\partial t}f(x,t) = \frac{\partial}{\partial t}A\cos\left(\frac{2\pi}{\lambda}x \pm \frac{2\pi}{T}t + \phi\right) = \mp\frac{2\pi}{T}A\sin\left(\frac{2\pi}{\lambda}x \pm \frac{2\pi}{T}t + \phi\right)$$

The maximum value of the sine function is 1, so the maximum speed of the string is $\frac{2\pi}{T}A$. In terms of the frequency, this maximum speed is $2\pi fA$. We can adjust both the frequency and the amplitude of the driver, and this won't change the speed of the wave (that can only be changed by altering the tension), so for the max speed of the string to remain unchanged (and equal to the wave speed), the amplitude and frequency have to be changed by the same factor, one of them up and the other down.

b. The wavelength can be determined from the speed of the wave and the frequency. The speed of the wave equals the maximum speed of the string, so:

$$\lambda f = v_{wave} = v_{max\ of\ string} = 2\pi fA \quad \Rightarrow \quad \frac{A}{\lambda} = \frac{1}{2\pi}$$

1.3: Energy Transmission

1-Dimensional Waves

While we will be interested in energy transmission in all kinds of waves, we will start with 1-dimensional harmonic waves as our model, as we are already familiar with the harmonic motion exhibited by the medium as such a wave passes. In particular, we have an idea of how to deal with the *energy* of a single oscillating particle, so for now we will also restrict ourselves to mechanical waves, where the particle comprising the medium are actually oscillating (i.e. think about a transverse wave on a string).

The energy of a single oscillating particle comes in two forms: kinetic and elastic potential (we'll maintain the convention that the particle is displacing in the y direction as the wave moves in the $\pm x$ direction):

$$E_{tot} = KE + PE_{elastic} = \frac{1}{2}mv^2 + \frac{1}{2}ky^2 \tag{1.3.1}$$

Of course both the speed of the particle and its displacement are changing with time, so it's more useful to express the energy of this particle in terms of one of the constants of the motion. When the particle reaches its maximum displacement, it stops moving, so its kinetic energy goes to zero and all of the energy is potential. But we have given this maximum displacement a name – amplitude. So the total energy of the oscillating particle is:

$$E_{tot} = \frac{1}{2}kA^2 \tag{1.3.2}$$

One might complain that there are no springs present for this kind of wave, so what are we supposed to plug into k? Well, there *is* a restoring force on every particle in the string as the wave passes, and this *behaves* like the restoring force of a spring, but we can write this expression more appropriately if we replace the spring constant with an equivalent expression in terms of the mass of the particle and the frequency of oscillation. Recall that for simple harmonic motion we have:

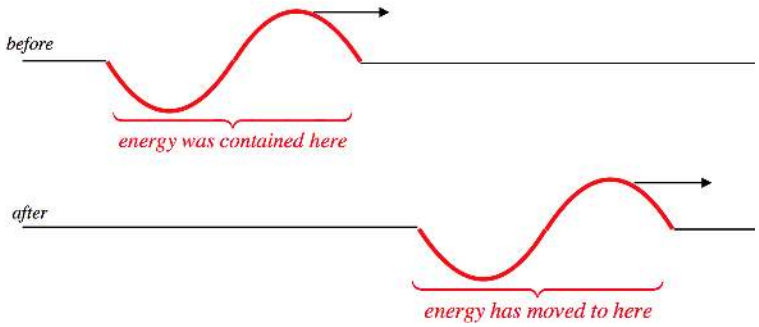
$$2\pi f = \omega = \sqrt{\frac{k}{m}} \quad \Rightarrow \quad k = m(2\pi f)^2 \tag{1.3.3}$$

Now the energy of the particle is in terms of the medium (the mass of the particle) and the wave (the amplitude and frequency):

$$E_{tot} = \frac{1}{2}m(2\pi f)^2 A^2 \tag{1.3.4}$$

We stated at the very beginning that waves carry energy from one point to another. Now that we see that a single particle in the medium carries energy, it should be clear that this is true. Consider a wave pulse that is harmonic for just one wavelength:

Figure 1.3.1 – Wave Pulse Carries Energy



Clearly the region where the particles are oscillating changes in this case, which means that the region that contains the energy is changed. The pulse transports energy across the expanse by having particles in the medium transfer energy to their nearest-neighbors, without the particles themselves having to make the trip.

Suppose we wish to know how much energy is in the whole wave, rather than what is just in a single particle. In this case, we treat the wave as continuous, with an infinite number of infinitesimal particles oscillating. The mass of these particles is very small, and can be written in terms of the mass density of the medium (again, think of this as a wave on a string), multplied by a small segment along the x direction:

$$dm = \mu dx \tag{1.3.5}$$

We can now use this mass to express the infinitesimal amount of energy possessed by that particle, by plugging this into Equation 1.3.4:

$$dE = \frac{1}{2} dm (2\pi f)^2 A^2 = \frac{1}{2} \mu dx (2\pi f)^2 A^2 \tag{1.3.6}$$

This is the energy in a single particle of the medium within the wave, so to get the full energy carried by the wave, we need only add up all these parts by performing an integral. The range of the single wave goes for one wavelength, so choosing the origin to be at one end of the wave, we have:

$$E_{in\ wave} = \int\limits_{in\ wave} dE = \int\limits_0^\lambda \frac{1}{2} \mu dx (2\pi f)^2 A^2 = \frac{1}{2} \mu (2\pi f)^2 A^2 \int\limits_0^\lambda dx = \frac{1}{2} \mu (2\pi f)^2 A^2 \lambda \tag{1.3.7}$$

Of course, if we have a full harmonic wave, as described by the wave function given in Equation 1.2.7, we have an infinite number of these single wave pulses, and the *amount* of energy in the entire wave is infinite and uninteresting. What is finite, even in the case of a full harmonic wave, is the *rate* as which energy is being transferred. To compute this, we simply need to divide how much energy a single wave pulse is carrying by the time it takes to completely cross some fixed point. Well, we know that this time interval is one period, so we have for the power of the wave:

$$P_{wave} = \frac{E_{one\ wave\ pulse}}{T} = \frac{1}{2} \mu (2\pi f)^2 A^2 \frac{\lambda}{T} = \frac{1}{2} \mu (2\pi f)^2 A^2 v, \tag{1.3.8}$$

where v is the speed of the wave. We can put this entirely in terms of the constants of motion of the wave and the medium for our case of a wave on a string by plugging in for the velocity in terms of the string tension and mass density (Equation 1.2.17):

$$P_{wave} = \frac{1}{2} \sqrt{\mu F} (2\pi f)^2 A^2 \tag{1.3.9}$$

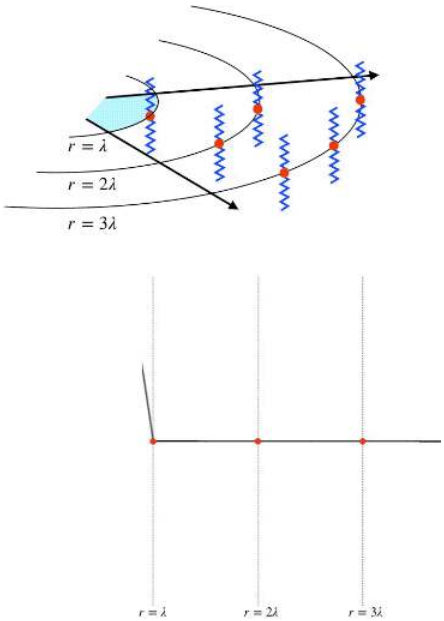
This calculation is specific to harmonic waves on strings, and we will not go into how this result changes for other types of harmonic waves (which pass through different sorts of media, may not be mechanical in nature, etc.). However, we will note that for a one-dimensional wave, the power is proportional to the square of the amplitude. As we will see, we will need to modify this result slightly for waves in two and three dimensions.

Multi-Dimensional Waves

In Section 1.1 we found that in order to satisfy the wave equation, waves that propagate out from a central source, into two or three dimensions cannot repeat their waveform. Here we will see why that is so, and get some idea of specifically how the waveform changes. Again, we will remain within the confines of our harmonic wave model for simplicity. First we need to clarify an important assumption: In our discussion we will assume that dissipative effects of the medium are negligible. That is, the particles in the medium that oscillate do so without "friction." This means we are assuming that all of the energy in the wave remains within the wave, and none of the energy is converted into thermal energy in the medium.

Consider now a wave radiating outward from a point source in two dimensions (think of a circular ripple on a pond caused by a pebble). Each position in the medium contains a particle oscillating harmonically (like a mass on a spring), and as the wave propagates outward, the number of oscillating particles increases. The particles in the medium are spaced the same everywhere, so the number of particles encountered by the circular wave is proportional to its circumference, and therefore proportional to its radius. This means that when the radius of the wave front doubles, it is oscillating twice as many particles in the medium.

Figure 1.3.2 – Circular Wave Energy Conservation



As the wave moves out, there is no energy lost, so the when the circle enlarges, the energy is distributed amongst a larger number of oscillators. The energy in each oscillator is determined by its amplitude of oscillation, so for more oscillators to have the same energy as fewer oscillators, their amplitudes must decrease. Specifically, the energy per oscillator is proportional to the *square* of the amplitude (Equation 1.3.2), which means that doubling the radius of the circle reduces the amplitude by a factor of $\sqrt{2}$, tripling the radius reduces the amplitude by a factor of $\sqrt{3}$, and so on. The figure above shows what happens to the amplitude of the wave in cross-section as it goes from a radius of 1 wavelength to 3 wavelengths.

The wave doesn't change its velocity from the inner circle to the outer circle, so the rate at which energy passes through each circle must be the same. What is different about two circles is the *density* of the energy contained in each. For the smaller circle, the energy is distributed over a smaller circumference than for the larger circle, so the energy density becomes smaller as the wave propagates outward. We can define power density in the same manner – by dividing the power of the wave (which is the same for both rings, and everywhere else) by the size of the region through which it is passing. This "power density" is called *intensity*. For our two-dimensional wave, this is the ratio of the power of the wave and the circumference of the circle through which it is passing:

$$I_{2d}(r) = \frac{P}{2\pi r} \tag{1.3.10}$$

Therefore the intensity of a two-dimensional wave radiating outward from a central point varies in inverse proportion to the distance from the central source. We find that the intensity is proportional to the square of the amplitude:

$$A \propto \frac{1}{\sqrt{r}} \quad \Rightarrow \quad I \propto A^2 \tag{1.3.11}$$

It turns out that the proportionality of intensity and square amplitude was the case for one-dimension as well. For a one-dimensional wave, the energy density does not change, because all of the energy is handed from one oscillator to another neighboring single oscillator. Therefore the power density (intensity) doesn't change, which is consistent with what we already know; the amplitude of a one-dimensional wave remains constant.

Far more common in our studies are three-dimensional waves with central sources (namely sound and light), and the power density in these cases involves dividing by a spherical surface area, rather than a circle. In this case, the intensity of the wave has units of watts per square meter (whereas the intensity of the two-dimensional wave had units of watts per meter), and we have:

$$I_{3d}(r) = \frac{P}{4\pi r^2} \tag{1.3.12}$$

Once again we find the same relationship between intensity and amplitude. The same mechanism is at work: As the wave moves outward from a central point, the number of oscillators on each spherical surface is proportional to the surface area. Doubling the

radius of a spherical surface quadruples the surface area, so the number of oscillators grows with the square of the radius. This means that the energy per oscillator drops with the square of the radius, and the amplitude is inversely-proportional to the radius:

$$A \propto \frac{1}{r} \quad \Rightarrow \quad I \propto A^2 \tag{1.3.13}$$

The relation between intensity and amplitude is therefore universal among waves, and one that we will keep in mind in the sections to come.

Note that this intensity drops faster than that of the two-dimensional wave, satisfying what's known as an *inverse-square law*: The intensity gets weaker in inverse proportion to the square of the distance from the source. Since the power of the wave is the same everywhere, we have the following relationship of intensities at two distances r_1 and r_2 from the source for waves that propagate outward from a point source:

$$I_1 r_1^2 = I_2 r_2^2 \tag{1.3.14}$$

Digression: Return to the Wave Equation

We can look to the wave equation in two and three dimensions to see if the relationship we obtain above between amplitude and radius actually holds true. For subtle reasons that we won't go into, it does hold true for three dimensions, but is only approximately true for two dimensions (the approximation improves as the wave gets farther from the source). If we accept the two-dimensional approximation, then we have a nice extension of the description of the wave function we gave in Section 1.1:

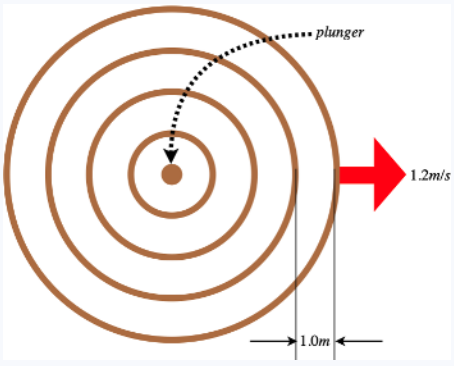
$$\begin{aligned} 1\ dimension: &\quad f(x,t) &&= f(x \pm vt) \\ 2\ dimensions: &\quad f(r,t) &&= \frac{f(r \pm vt)}{\sqrt{r}} \quad (approximate,\ for\ large\ r) \\ 3\ dimensions: &\quad f(r,t) &&= \frac{f(r \pm vt)}{r} \end{aligned}$$

To demonstrate these relations, one must write the wave equation in polar coordinates for two and three dimensions, respectively. For the reader that may be inspired to give demonstrating this a try, here are the double spatial derivatives for each case:

$$\begin{aligned} 2\ dimensions: &\quad \nabla^2 f &&= \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) \\ 3\ dimensions: &\quad \nabla^2 f &&= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) \end{aligned}$$

Example 1.3.1

At time $t = 0$, a plunger begins oscillating up-and-down at a steady rate for 6 full oscillations in a body of otherwise calm water. During this time, it puts $420J$ of energy into the surface waves it creates. The wavelength of the wave is measured to be $1.0m$, and the wave speed is measured to be $1.2m/s$.



a. *Find the power supplied by the plunger.*
b. *Find the intensity of the leading wavefront at time $t = 4.0s$.*

c. *The amplitude of the leading wavefront at $t = 4.0s$ is measured to be $5.4cm$. Find its amplitude at $t = 8.0s$.*

Solution

a. The power is the rate at which the energy is being transferred into the waves. We know how much energy is put into 6 oscillations, so if we divide that energy by the time span of 6 oscillations, we have the value of the power. The time span of 6 oscillations is 6 periods, and a single period we can calculate from the wavelength and wave speed:

$$T = \frac{\lambda}{v} = \frac{5}{6}s \quad \Rightarrow \quad \Delta t = 6T = 5.0s \quad \Rightarrow \quad P = \frac{E}{\Delta t} = \frac{420J}{5s} = 84W$$

b. To get the intensity, we need to know the circumference of the leading wavefront. We know the speed of the wave and how long it has been traveling, so:

$$r = v\Delta t = \left(1.2\frac{m}{s}\right)(4.0s) = 4.8m \quad \Rightarrow \quad I = \frac{P}{2\pi r} = \frac{84W}{2\pi 4.8m} = 2.8\frac{W}{m}$$

c. For the two-dimensional wave, the amplitude gets smaller as the radius grows, by a factor of $\frac{1}{\sqrt{r}}$. The wave's speed is unchanging, so after 8 seconds the wave has traveled twice as far from the source than after 4 seconds. Doubling the distance traveled therefore reduces the amplitude by a factor of $\sqrt{2}$, giving:

$$A\left(t = 8.0s\right) = \frac{A\left(t = 4.0s\right)}{\sqrt{2}} = \frac{5.4cm}{\sqrt{2}} = 3.8cm$$

1.4: Superposition and Interference

Combining Similar Waves

When two or more waves of the same type in the same medium coexist in the same region of space, they combine to create a new wave. The way they combine is a simple process known as *superposition*. This consists of simply adding the displacements (or whatever the wave function represents) of the two or more waves at the same place and time. For a 1-dimensional wave, this means:

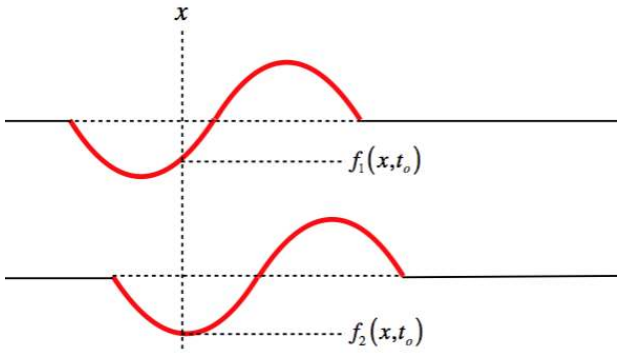
$$f_{tot}(x,t) = f_1(x,t) + f_2(x,t) \tag{1.4.1}$$

Alert

It's important to emphasize that two waves can only superpose if they are the same type. Many different kinds of waves can travel through the same medium (light, sound, and displacement waves can all travel through water in a lake, for instance), but these cannot superpose with each other.

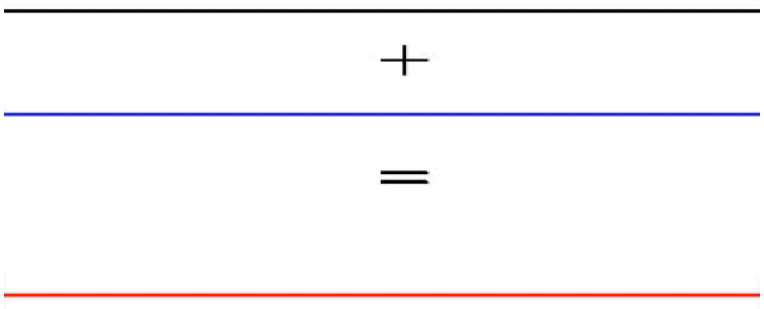
It's easy to show that if the two individual wave functions satisfy the wave equation, then so does the total wave function. It bears repeating with a diagram that this superposition sum involves adding displacements *at the same place and time*. So if we took a snapshot of two waves, we would determine the total wave by lining them up:

Figure 1.4.1 – Superposition



The composite wave is then the combination of all of the points added thus. Of course, these are traveling waves, so over time the superposition produces a composite wave that can vary with time in interesting ways. Here is a simple example of two pulses "colliding" (the "sum" of the top two waves yields the bottom wave).

Figure 1.4.2 – "Collision" of Pulses



Notice that even though the resultant wave looks very different from its "parents," the medium somehow "remembers" the original waves, and when they no longer coincide, they continue along as exactly the waves they were before the superposition. That is, the waves do not affect each other, as particles would if they collided – waves don't bounce off each other, for example. They simply

create a new wave while they occupy the same space in the medium, and when their individual motions carry them to different parts of the medium, they return to being the waves they were before.

Interference – All-or-Nothing

There are some special cases involving superposition that are particularly interesting to examine, and these involve a phenomenon known as *interference*. There are many degrees of interference possible, all of which fall between the following two extremes:

- *constructive interference*: The waves are perfectly aligned and timed so that their crests and troughs coincide, such that the total wave has the maximum possible amplitude (equal to the sum of the amplitudes of the two constituent waves).
- *destructive interference*: The waves are perfectly aligned and timed so that the crests of one wave align with the troughs of the other such that leading to a wave that has the minimum possible amplitude (equal to the difference of the amplitudes of the two constituent waves).

The phrase *total destructive interference* refers to the case of destructive interference when the resultant wave has zero amplitude, i.e. the two waves totally cancel each other. In the cases we will discuss, we will only talk about this extreme case of destructive interference, so we will typically leave out the word "total," even though we are still talking about total cancelation.

Interference – Intensity of Combined Wave

We will examine a great many examples of interference in physical phenomena in the sections to come. We therefore need to take some time to develop the mathematics behind this effect. We will do this within the same framework that we have been using – that of harmonic waves. When we look at the physical attributes of interference, what we will be examining is what happens to the intensity of the combined wave. For example, interference in sound will be exhibited in volume, and in light it will be brightness. Both of these are measures of intensity. We need a reference point for intensity, and the one we will use is that of maximal constructive interference. So what we seek is an equation that relates the intensity of two superposed, out-of-phase, but otherwise identical waves to the intensity we would see if they were in phase. That is, we want something that looks like this:

$$I\left(\Delta\Phi\right) = I_o g\left(\Delta\Phi\right) \tag{1.4.2}$$

The quantity I is the intensity of the wave as a function of the phase difference of the two (identical) parent waves. If the two waves happen to be in phase, then the combined wave's intensity is I_o when the two waves are in phase. Note that this is *four times the intensity of each individual wave*, since the constructive interference adds the amplitudes (which are equal – the waves are identical) and the intensity is proportional to the square of the amplitude.

The function needs to have the following properties:

- It has to always be non-negative, since intensity is never a negative number.
- It has to vanish when the phase difference equals π (modulo 2π), since this means the waves totally destructively interfere.
- It has to equal 1 when when the phase difference is 0 (modulo 2π), since this means the waves constructively interfere.

To find this function, we start with two wave functions that are identical except for their phases and superpose them:

$$f_{tot} = f_1 + f_2 = A\cos(\Phi_1) + A\cos(\Phi_2) = A\left[\cos(\Phi_1) + \cos(\Phi_2)\right] \tag{1.4.3}$$

We want this function to only depend upon the difference in the two phases, so we will write each total phase in terms of deviation from their average phase (which we will call simply Φ), and the difference in phase between the two waves, $\Delta\Phi$:

$$\Phi = \frac{\Phi_1 + \Phi_2}{2} \quad \Rightarrow \quad \Phi_1 = \Phi - \frac{\Delta\Phi}{2}, \quad \Phi_2 = \Phi + \frac{\Delta\Phi}{2} \tag{1.4.4}$$

Plug these into Equation 1.4.3 gives:

$$f_{tot} = A\left[\cos\left(\Phi - \frac{\Delta\Phi}{2}\right) + \cos\left(\Phi + \frac{\Delta\Phi}{2}\right)\right] \tag{1.4.5}$$

Now we can apply a trigonometric identity:

$$\cos(A-B) + \cos(A+B) = 2\cos A\cos B \quad \Rightarrow \quad f_{tot} = 2A\cos(\Phi)\cos\left(\frac{\Delta\Phi}{2}\right) \tag{1.4.6}$$

The phase difference between the two waves can be written in terms of the difference in position, time, and the phase constant, using Equation 1.2.9:

$$\Delta\Phi = \frac{2\pi}{\lambda}(x_1 - x_2) \pm \frac{2\pi}{T}(t_1 - t_2) + \phi_1 - \phi_2 = \frac{2\pi}{\lambda}\Delta x \pm \frac{2\pi}{T}\Delta t + \Delta\phi \tag{1.4.7}$$

As the waves propagate along, the values of x and t will change, but as the two waves are identical (traveling in the same direction with the same speed), the differences in x and t don't change for a given phase. Therefore the factor in Equation 1.4.6 that includes the phase difference is a constant. Putting that constant together with the $2A$ gives us the amplitude of the new conglomerate wave (with the time-varying phase being the average of the phases of the two waves):

$$f_{tot} = A_{new}\cos(\Phi), \qquad A_{new} \equiv 2A\cos\left(\frac{\Delta\Phi}{2}\right) \tag{1.4.8}$$

The intensity is proportional to the square of the amplitude, so the intensity of this combined wave is:

$$I \propto A_{new}^2 = 4A^2\cos^2\left(\frac{\Delta\Phi}{2}\right) \tag{1.4.9}$$

The intensity of each individual wave is proportional to A^2. If the waves were in phase, the total amplitude would double, which means that the total in-phase intensity I_o is proportional to $4A^2$. The intensity of the (out of phase) combined wave is therefore:

$$I = I_o\cos^2\left(\frac{\Delta\Phi}{2}\right) \tag{1.4.10}$$

Notice that this relationship between total intensity and phase difference exactly matches the three criteria we outlined above.

Example 1.4.1

Two identical harmonically-oscillating devices attached to a taut string are turned on simultaneously and in phase, vibrating at a frequency of $20Hz$. One source is located at the origin, and the other is positioned at $x = 25cm$. Each source vibrates with a maximum displacement of the string of $8.0cm$. The string has a density of $35g\cdot m^{-1}$, and is stretched with a tension of $5.2N$. The sources are allowed to vibrate long enough for their respective waves to superpose.

a. *Find the amplitude of the combined wave.*
b. *Find the power of the combined wave.*
c. *Find the distance in the $+x$ direction that the source not at the origin must be moved for the power output to be maximized, and determine this maximum power output.*

Solution

a. The amplitude of the combined wave is given by Equation 1.4.8, which requires that we calculate the phase difference $\Delta\Phi$:

$$\Delta\Phi = \frac{2\pi}{\lambda}\Delta x \pm \frac{2\pi}{T}\Delta t + \Delta\phi$$

The sources start at the same moment in time and in phase, so $\Delta t = 0$ and $\Delta\phi = 0$. Therefore the only source of phase difference is the separation of the sources, Δx. We still need the wavelength, which we can get from the frequency and wave speed. The latter we get from the string density and tension. Putting all this together gives:

$$\lambda = \frac{v}{f}, \quad v = \sqrt{\frac{F}{\mu}} \quad \Rightarrow \quad \lambda = \frac{1}{f}\sqrt{\frac{F}{\mu}} = \frac{1}{20Hz}\sqrt{\frac{5.2N}{0.035kg\cdot m^{-1}}} = 61cm$$

Plugging this in above gives:

$$\Delta\Phi = \frac{2\pi}{\lambda}\Delta x = \frac{2\pi}{61cm}(25cm) = 0.82\pi$$

The amplitude of the combined wave is therefore:

$$A_{combined} = 2A\cos\left(\frac{\Delta\Phi}{2}\right) = 2\,(8.0cm)\cos(0.41\pi) = 4.46cm$$

Notice that the amplitude of the combined wave is actually smaller than that of each individual wave. This is because the separation is such that there is some partially destructive interference going on.

b. The power of the combined wave can be found directly from Equation 1.3.9:

$$P = \frac{1}{2}\sqrt{\mu F}(2\pi f)^2 A^2 = \frac{1}{2}\sqrt{\left(0.035 kg\cdot m^{-1}\right)(5.2N)}(2\pi)^2(20Hz)^2(0.045m)^2 = 6.8W$$

c. The power of a one-dimensional wave is the same as its intensity. From Equation 1.4.10 it is clear that the maximum intensity occurs when the cosine function equals ± 1. This occurs when the argument is some integer multiplied by π. The argument in the case above is $\frac{\Delta\Phi}{2} = \frac{0.82\pi}{2}$, so the source must be shifted enough for the total phase to change by an additional 1.18π, to give: $\frac{\Delta\Phi}{2} = \frac{1.18\pi + 0.82\pi}{2} = \pi$. This phase difference is related to the change in position through the wavelength:

$$\Delta\Phi = \frac{2\pi}{\lambda}\Delta x \quad \Rightarrow \quad \Delta x = \Delta\Phi\left(\frac{\lambda}{2\pi}\right) = (1.18\pi)\frac{61cm}{2\pi} = 36cm$$

Notice that the position where the second source creates maximum constructive interference is exactly one wavelength away from the first source: $25cm + 36cm = 61cm = \lambda$. This is because with only spatial separation in play, the phase difference changes by 2π every time the spatial separation changes by λ.

The maximum possible intensity is I_o, and since we know the intensity of the combined wave and the phase difference, we can compute this maximum from Equation 1.4.10:

$$I_o = \frac{I}{\cos^2\left(\frac{\Delta\Phi}{2}\right)} = \frac{6.8W}{\cos^2\left(\frac{0.82\pi}{2}\right)} = 87W$$

A simpler approach would be to note that the intensity is proportional to the square of the amplitude, and since changing the position of the second source increases the amplitude from $4.46cm$ to $16cm$ (double the individual wave amplitude), which is an increase of a factor of 3.59, the intensity increases by a factor of $3.59^2 = 12.9$, giving the same result (except for some rounding errors along the way).

How to Create Interference

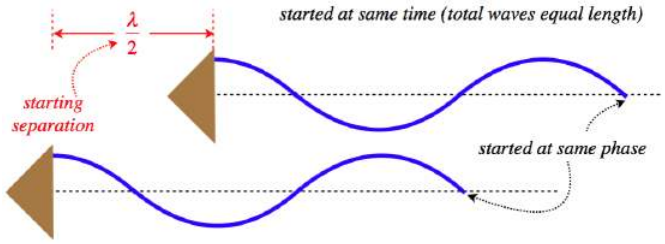
Whenever two waves interfere, whether it is constructively, destructively, or anything in-between, it's clear that the critical factor is the phase difference between the waves, $\Delta\Phi$. From Equation 1.4.5 we can see several ways in which a difference in phase can occur: The two waves can travel a different distance (Δx), they can have been traveling for a different amount of time (Δt), they could have started out of phase ($\Delta\phi$), or it could be some combination of these three differences.

To understand how these three differences can be physically manifested, it's easiest to let two of them be zero, and let only one difference occur at a time. We can do this in two ways. The first is to simply construct physical situations that assures this, and the second consists of nothing more than a change of perspective. For the sake of studying this effect, we will only consider destructive interference, but we can do the same for constructive or anything in between as well. To keep things simple, we'll interfere (approximately) 1-dimensional waves traveling in the same direction, which we can model with identical harmonic sound waves coming from two speakers pointed in the same direction. There is nothing about the result that is specific to sound, however; this is a phenomenon common to all waves.

Case 1: Different Travel Distance ($\Delta x \neq 0$, $\Delta t = 0$, $\Delta\phi = 0$)

We start with a case where the two sound waves emanate from their respective speakers such that the leading wave front for each sound wave is at the same phase (in the diagram below, if we use a cosine function to describe these waves, then both waves have a phase of $\frac{\pi}{2}$ at their leading edge). We also start the waves from their speakers at the same moment (in the diagram below, both waves have been propagating for one full oscillation plus another quarter of an oscillation, so they started at the same time). But the waves are offset in the positions where they begin by one-half wavelength, which results in the two waves occupying the same medium π radians out of phase:

Figure 1.4.3 – Destructive Interference Due to Travel Distance Only

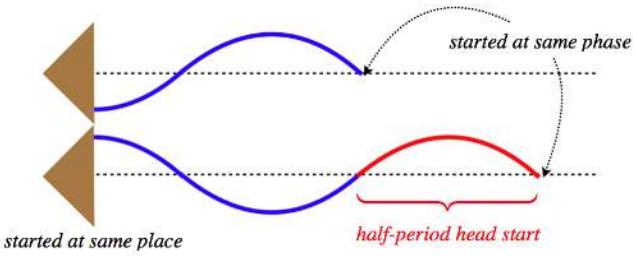


$$\Delta\Phi = 2\pi \underbrace{\Delta\left(\frac{x}{\lambda}\right)}_{\frac{1}{2}} \pm 2\pi \underbrace{\Delta\left(\frac{t}{T}\right)}_{0} + \underbrace{\Delta\phi}_{0} = \pi \quad (\textit{destructive interference})$$

Case 2: Different Time Intervals ($\Delta x = 0$, $\Delta t \neq 0$, $\Delta\phi = 0$)

Next we will place the speakers side-by-side, so that the travel distances of the waves at any position in the medium is the same. As before, the leading wave front of the waves will be in phase, but this time we will turn on one of the speakers at a time of one-half period before the other speaker. This lag between the two once again throws the two waves out of phase, resulting in destructive interference:

Figure 1.4.4 – Destructive Interference Due to Starting Time Only

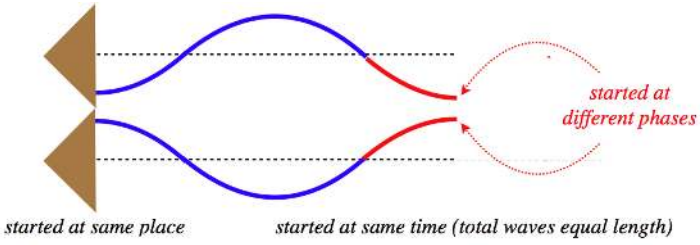


$$\Delta\Phi = 2\pi \underbrace{\Delta\left(\frac{x}{\lambda}\right)}_{0} \pm 2\pi \underbrace{\Delta\left(\frac{t}{T}\right)}_{\frac{1}{2}} + \underbrace{\Delta\phi}_{0} = \pi \quad (\textit{destructive interference})$$

Case 3: Different Starting Phases ($\Delta x = 0$, $\Delta t = 0$, $\Delta\phi \neq 0$)

Now finally, we will position speakers side-by-side, and turn them on at the same moment, but will arrange things so that the sounds they emanate leave the speakers π radians out of phase, to get the same destructive interference:

Figure 1.4.5 – Destructive Interference Due to Phase Constant Only



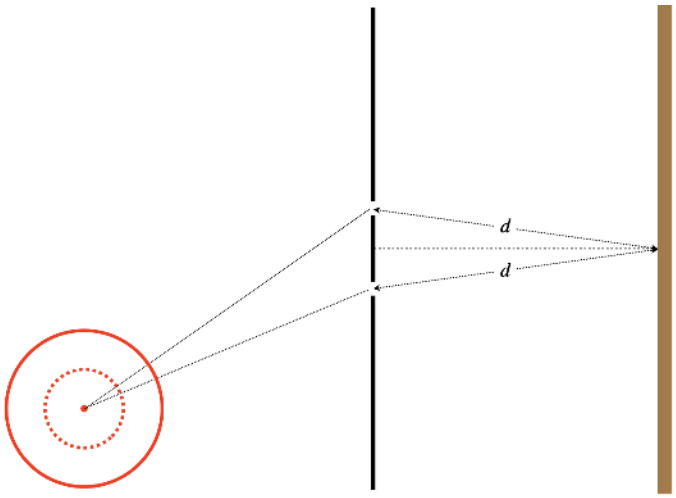
$$\Delta\Phi = 2\pi \underbrace{\Delta\left(\frac{x}{\lambda}\right)}_{0} \pm 2\pi \underbrace{\Delta\left(\frac{t}{T}\right)}_{0} + \underbrace{\Delta\phi}_{\pi} = \pi \quad (\textit{destructive interference})$$

One More Case: Different Perspectives

While these all seem distinctly different, very often the same interference effect can be described in any of the three ways, simply by changing our perspective. To see this, let's consider a two-dimensional harmonic wave coming from a single point source. This could be a wave caused by a pebble dropped into a pond, for example. This wave radiates circularly-outward from the source. We of course cannot have interference with only one wave, so we will provide a means to split it into two separate waves: The wave will strike a barrier with two small holes in it, through which the wave can pass. As we will see in a future section, these two holes themselves now act like point sources of two separate waves, with the energy of these waves coming from the original wave. It's these two waves that we will allow to interfere.

Consider the diagram below. We will be looking at the intensity of the wave when it strikes a second barrier at a position that is equal distances (labeled as '*d*' in the diagram) from the two holes. The original source of the wave is *not* the same distance from both holes, however.

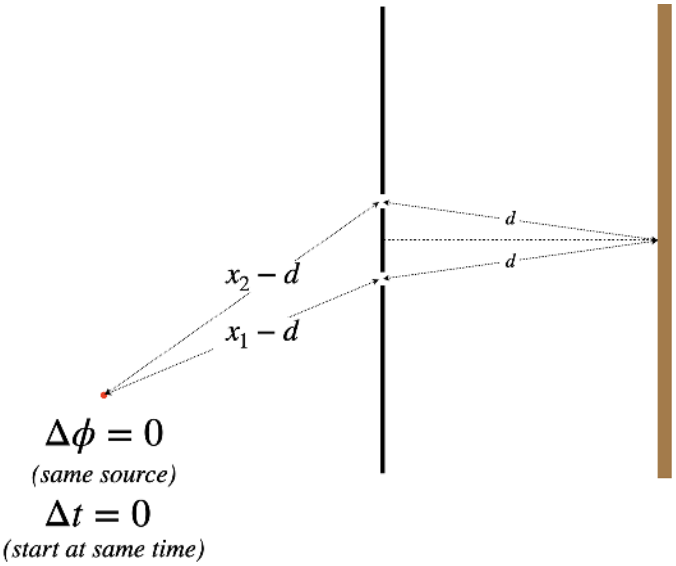
Figure 1.4.6 – Single Source Interference Setup



Suppose we find that the two split waves interfere destructively at the final destination. How can we explain this result? It turns out that we can do it in any of the three ways described above, depending on what perspective we decide to take.

First, we can note that the two waves start at the same time and with the same phase, from the original point source. But these two waves (which initially are part of the same starting wave) do *not* travel equal distances: $x_1 \neq x_2$. They travel the same distance from the holes to the screen, but the distances to the holes from the starting point are different:

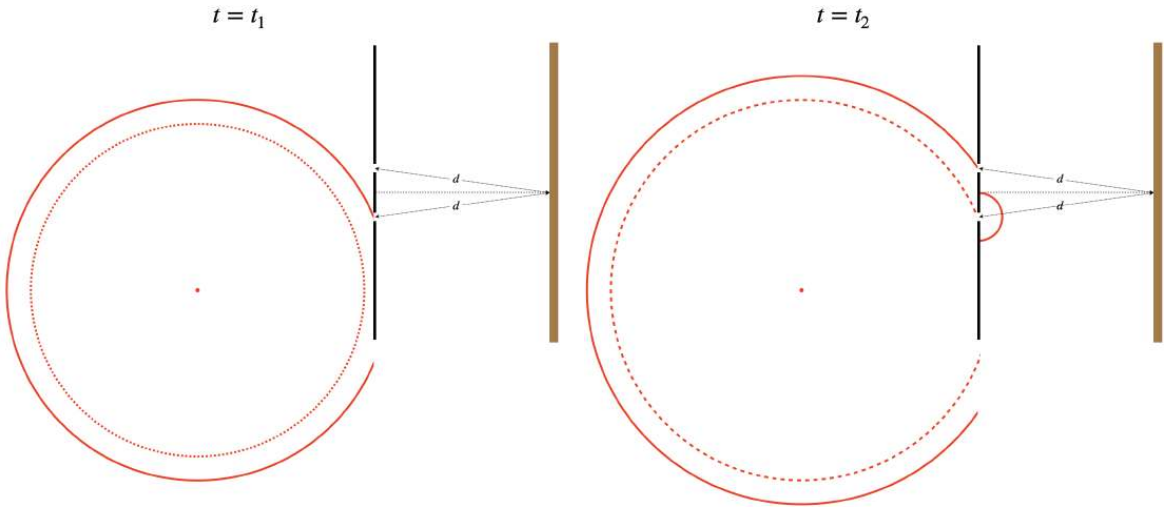
Figure 1.4.7 – Single Source Interference – Distance Traveled Explanation



We would see destructive interference if (for example), the extra distance traveled by the wave passing through one hole happens to be a half wavelength farther than the distance traveled by the wave passing through the other hole: $x_2 - x_1 = \frac{1}{2}\lambda$.

Now let's change our perspective. Suppose we are watching this wave from the right of the two holes, and know nothing at all about the single point source. As far as we are concerned, the two waves are starting from positions that are the *same distance* from the position where we are seeing the destructive interference. We would not conclude that the difference in distance traveled is the cause of this interference. But suppose we had been watching since before the first wave emerged from a hole. In this case, we would see a wave with a certain starting phase come out of one hole first, and a short time later, a wave with the same phase come from the other hole. It comes out with the same phase because it comes from the same wavefront from the original point source.

Figure 1.4.8 – Single Source Interference – Time Elapsed Explanation

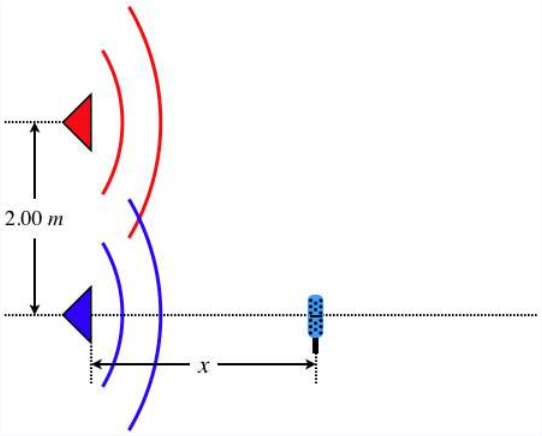


We would see destructive interference if (for example), the time elapsed between when we see the two waves emerge differs by one-half period: $t_2 - t_1 = \frac{1}{2}T$.

And finally, there is one other perspective from which we can view this. Once again, we will view the waves from the side of the holes where we can't see the point source, so that we again measure equal travel distances. But this time, let's assume we don't start viewing until after the wave has been passing through both holes for awhile. We look at our watch and note that at time $t = 0$ (when we start watching), there are waves coming from both holes that are out of phase with each other by π radians. Both waves travel the same distance and start at the same time, but start out out of phase, and therefore destructively interfere.

Example 1.4.2

Two speakers, both pointing in the $+x$ direction, are placed on the y-axis, separated from each other by a distance of $2.00\ m$. They emit the same tone, which has a frequency of $784\ Hz$, in phase with each other. A microphone is placed directly in front of and very close to one of the speakers, and is gradually moved from along the x-axis farther and farther from the speaker. Assume that the fact that the microphone is a little farther from one speaker than the other does not result in a noticeable intensity difference between the two speakers, so that the sound waves coming from the speakers have the same amplitude when they reach the microphone. The speed of sound waves in air is $344\frac{m}{s}$.



a. Find the distances from the closer speaker where the microphone detects no sound.
b. Find the distances from the closer speaker where the sound gets loudest (i.e. constructive interference).
c. Suppose the tone coming from the speakers has an adjustable frequency, and that it is gradually lowered. Find the frequency below which the microphone has no position on the x-axis where it measures total silence.

Solution

a. The starting phase and time are the same, so the only source of phase difference comes from the difference in distance traveled. From the Δx contribution to the phase difference that causes destructive interference (i.e. when the extra distance is an odd number of half-wavelengths), we therefore have:

$$\frac{2\pi}{\lambda}\Delta x = n\pi \quad \Rightarrow \quad \Delta x = \frac{1}{2}n\lambda\,, \qquad n = 1,\ 3,\ 5,\ \ldots$$

The difference in distances traveled by the two waves us found using the Pythagorean theorem, so putting this in above gives:

$$\Delta x = \sqrt{x^2 + (2.00\ m)^2} - x = \frac{1}{2}n\lambda \tag{1.4.11}$$

Calling the n^{th} value of x "x_n" and doing the algebra gives:

$$x_n = \frac{1}{n\lambda}\left(4.00m^2 - 0.250n^2\lambda^2\right)$$

We are given the frequency of the sound, so we can find its wavelength:

$$\lambda = \frac{v}{f} = \frac{344\frac{m}{s}}{784Hz} = 0.439m$$

Notice that although the value of n is not restricted, when it gets too high, the value of x_n will become negative. Plugging in all of the values of n that give positive values of x yields five possible values:

$$x_1 = 9.00m, \quad x_3 = 2.71m, \quad x_5 = 1.27m, \quad x_7 = 0.531m, \quad x_9 = 0.022m$$

b. Constructive interference occurs when the path difference is an even number of half-wavelengths (i.e. some number of full wavelengths). We can get our answer directly from part (a) simply by taking even values of n instead of odd values. Once again, the number of values of n is limited by the restriction that sign of x_n must be positive.

$$x_2 = 4.34m, \quad x_4 = 1.84m, \quad x_6 = 0.858m, \quad x_8 = 0.259m$$

c. The value of Δx clearly gets smaller as x gets larger (the hypotenuse gets closer and closer to equaling the x value as x gets larger), so the largest possible value of Δx is just the separation of the speakers. If the speakers are separated by less than one-half wavelength, then Δx can never get as big as a half wavelength, and no totally destructive interference is possible. These speakers are separated by $2.00\ m$, so the wavelength of the sound must be shorter than $4.00m$ for there to be any instance of total destructive interference. This wavelength corresponds to a frequency of:

$$f = \frac{v}{\lambda} = \frac{344 \frac{m}{s}}{4.00m} = 86Hz$$

Frequencies lower than this create wavelengths that are so long that Δx is never large enough to cause total destructive interference.

1.5: Standing Waves

Interference Patterns

We found that interference occurs between two identical waves, but we didn't mention what the source of two identical waves might be. We will find that most of the time the two waves are actually the *same* wave, where one part of it has been diverted somehow, so that it behaves like a separate wave. When we witness the interference created in such a situation, it is often in the form of an *interference pattern*. This is a recognizable pattern of intensity that repeats itself in space or in time, or in both. We will see lots of these patterns in the sections to come, but as usual we will start with a simple (but important) one-dimensional example of an interference pattern, called a *standing wave*. [*Actually, standing waves occur in 2 and 3 dimensions as well, though we will confine our discussion to those of the 1-dimensional variety.*]

Alert

The moniker "standing wave" puts yet another strain on our definition of what it means to be a wave. It does satisfy the wave equation (as does any superposition of waves), but although the wave equation yields a wave velocity, this waveform does not propagate at all. It is better to think of standing waves as what they are – interference patterns.

All interference patterns are formed from multiple identical waves, and like so many other interference patterns, this is accomplished through multiple versions of the same wave. In the case of the standing wave, these two versions are the result of wave *reflections* off two endpoints. That is, a single wave bounces back-and forth between two endpoints, and as it crosses itself during the journey, the standing wave interference pattern is formed from the superposition – the two waves that are interfering only differ in their directions of motion.

Wave Reflection & Transmission

Before we delve into the details of standing waves, we first need to look at the phenomenon that makes them possible – wave reflection. The mathematics of wave reflection can become quite involved and we will not delve into it here, but the bottom line is that waves reflect off sudden changes in the medium. We have found that the medium is best characterized by the speed of waves that pass through it, and in fact it is correct to say that a wave reflects when it encounters a region of the medium where the wave speed changes.

At this point one might ask why the wave doesn't simply continue in the direction it was going, but at a different speed. It does! But it also reflects. That is, the wave splits into two parts, called the *reflected wave* and the *transmitted wave*. Of course, energy is conserved during this schism, so the energy in the original wave is greater than the energies in either of these waves. The amount of energy that goes to each wave is determined mathematically by a process known as "matching boundary conditions" at the point of reflection, but as mentioned earlier, we will not make a close examination of this process here (this topic is explored in courses on quantum mechanics, such as Physics 9D).

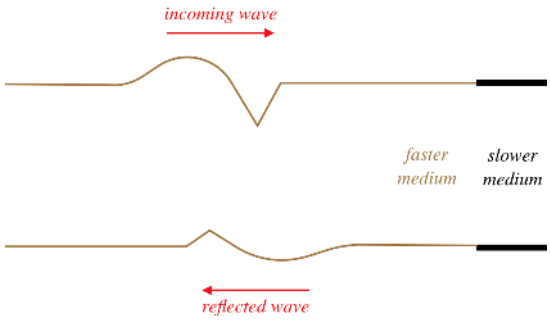
The requirement that the speed of the wave changes at the point of reflection in the medium doesn't distinguish between the wave is coming from a faster medium to a slower one, or from a slower one to a faster one. It turns out that the wave will partially transmit and partially reflect, no matter which direction it is going. But there is an observable phenomenon that distinguishes these two possibilities. Suppose a pulse of a wave on a string consists of just a single bump (like half a sine function) that lies on the top half of the string. If this wave reaches a point in the medium where it speeds up (the string's linear density goes down), then the reflected pulse remains in the top half of the string. But if the pulse encounters a point in the medium where it slows down (the string's linear density goes up), then the reflected pulse flips to the bottom half of the string. The transmitted wave never flips over.

Figure 1.5.1 – Reflection and Transmission (Slow-to-Fast Medium)

Figure 1.5.2 – Reflection and Transmission (Fast-to-Slow Medium)

It should also be noted that the reflected wave in both cases reflects its wave form along $-x \leftrightarrow +x$. This is not obvious in the case of a symmetric pulse, but if the wave is asymmetric, then it becomes apparent.

Figure 1.5.3 – Reflection of an Asymmetric Wave (Fast-to-Slow Medium)



The leading edge of the incoming waveform is the leading edge of the reflected waveform. While there is some loss of amplitude for the reflected wave compared to its incoming counterpart (some of the energy is taken by the transmitted wave), the wavelength of the reflected wave is the same as the incoming wave. This is because the velocity of the reflected wave is the same as the incoming wave. The wavelength of the transmitted wave will not match the wavelength of the incoming wave, however. The time span between the front and rear of the waveform striking the new medium is the same time as it takes for the full waveform to be transmitted, so the periods of the incoming, transmitted, and reflected waves are all the same, but since the velocity is different for the reflected and transmitted waves, the result is different wavelengths for these waves (λ will be longer in the faster medium).

Reflection without Transmission

In order to discuss standing waves, we need to completely confine the wave between two endpoints – no energy can be allowed to escape via transmission. We can make such a confinement simply by cutting off the medium at the endpoints. The wave will reflect off this sudden absence of medium, and all of the energy of the incoming wave returns in the reflected wave. But does the wave flip over or stay upright in such a case?

In fact both of these results are possible, because the edge of the medium can react in one of two ways. If the edge of the medium is held fixed (i.e. not allowed to exhibit the displacement that the wave provides every other point in the medium), then the reflected wave flips over. If the edge of the medium is free to displace, then the reflected wave does not flip over.

Figure 1.5.4 – Reflection off a Fixed End

Figure 1.5.5 – Reflection off a Free End

Explaining this result is quite tricky from a perspective of forces on the end of string, and even after figuring that out, it's hard to extend it to other types of waves (this phenomenon applies to all waves, though sometimes determining what is meant by "fixed" and "free" can be tricky). But there is a nice way to use an imaginary model to achieve this result. It goes like this:

Suppose we model a single wave hitting the end of the medium with *two* waves, moving in opposite directions through the point that is the end of the medium and passing each other. Clearly the second wave doesn't exist, since there is no medium beyond the end, but its emergence from the passing point is seen as the "reflected wave," while the other wave vanishes past the passing point. With this model, we first see that they must have the same basic waveform, and that the leading edge of one wave must correspond to the leading edge of the other. But now we ask how the imaginary wave (i.e. the second wave, before it emerges as the reflected wave) must be oriented for the passing point to be fixed or free. For the passing point to remain stationary, the superposition of the two waves at that point must result in total destructive interference. This can only happen if the second wave is inverted compared to the first wave, so when it emerges as the reflected wave, it has been flipped over. If the passing point moves freely, the two waves cannot interfere destructively, so the second wave emerges upright. Note that this analysis tells us that the free end displaces an amount equal to *twice* the amplitude, since the waves are identical and the interference is constructive.

As often as we use harmonic waves, it is useful to put the phenomenon of reflected waves in that context. When we flip over a sine or cosine wave, the result is identical to shifting that wave by a phase of π:

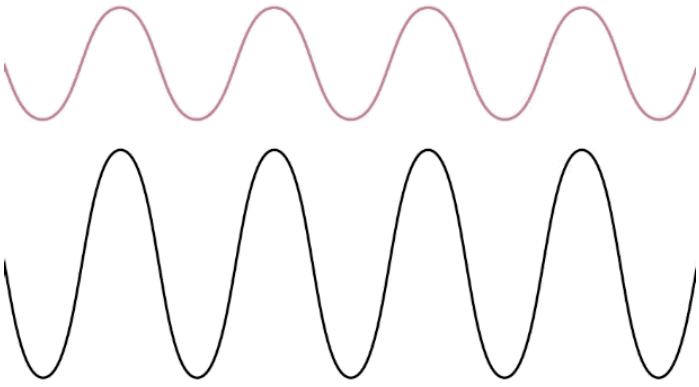
$$flip\ wave\ function: \quad A\cos\left(\frac{2\pi}{\lambda}x \pm \frac{2\pi}{T}t + \phi\right) \rightarrow -A\cos\left(\frac{2\pi}{\lambda}x \pm \frac{2\pi}{T}t + \phi\right) = A\cos\left(\frac{2\pi}{\lambda}x \pm \frac{2\pi}{T}t + \phi + \pi\right) \quad (1.5.1)$$

The inversion of a reflected wave after coming off a fixed end or a slower medium is therefore often referred to as a *phase shift of π*.

Standing Wave Mathematics

Now we know that we can get a wave to bounce back-and-forth between two ends of a medium, and the waves going each way are identical. If conditions for these waves are just right, their superposition results in a standing wave.

Figure 1.5.6 – Standing Wave



Let's see how this result occurs mathematically. This requires superposing two wave functions with the wave wavelength (wave number) and period (angular frequency) that are moving in opposite directions:

$$\begin{aligned} right-moving\ wave: &\qquad f_1(x,t) = A\cos(kx-\omega t+\phi_1) \\ left-moving\ wave: &\qquad f_2(x,t) = A\cos(kx+\omega t+\phi_2) \end{aligned} \tag{1.5.2}$$

Recall that this standing wave occurs because a single wave is bouncing back-and-forth between endpoints in the medium. The endpoints must either each be free (no phase shift) or fixed (π phase shift). For the sake of getting an easy-to-read result, we'll assume that a fixed endpoint lies at position $x=0$. Because we are talking about a position where the wave reflects, and because the point is fixed, the two waves must be out of phase by π radians. Mathematically this means that the difference in their phase constants is π:

$$\phi_2 - \phi_1 = \pi \tag{1.5.3}$$

Plugging $x=0$ and $\phi_2 = \phi_1 + \pi$ into the superposition of the two waves and setting the result equal to zero (that point remains fixed by our simplifying assumption), we get:

$$0 = f_{tot}(0,t) = A\cos(0-\omega t+\phi_1) + A\cos(0+\omega t+\phi_1+\pi) = A\cos(-\omega t+\phi_1) - A\cos(\omega t+\phi_1) \quad \Rightarrow \quad \cos(-\omega t+\phi_1) = \cos(\omega t+\phi_1) \tag{1.5.4}$$

We can solve this for ϕ_1, which comes out to be: $0, \pm 2\pi \pm 4\pi \ldots$. We'll take the simplest solution of zero, which leaves us with the following wave function:

$$f_{tot}(x,t) = A\cos(kx-\omega t) - A\cos(kx+\omega t) \tag{1.5.5}$$

We can now apply the following trigonometric identity to get a simplified form of the standing wave function:

$$\cos(X-Y) - \cos(X+Y) = 2\sin X \sin Y \quad \Rightarrow \quad f_{SW}(x,t) = 2A\sin kx \sin\omega t = 2A\sin\left(\frac{2\pi x}{\lambda}\right)\sin\left(\frac{2\pi t}{T}\right) \tag{1.5.6}$$

All harmonic waves are collections of harmonically-oscillating points in a medium, that vary in total phase from one position to the next. Traveling waves satisfy this, but the amplitudes of these oscillators are all the same in this case. When interference occurs, the *amplitude* can vary from one position to the next as well (e.g. positions of destructive interference have zero amplitude, and positions of constructive interference have very large amplitudes), and this is evident in this result for the interference pattern we call a standing wave. This formula can be written as a collection of harmonic oscillators all with the same a period (and therefore the same sine function), but with different amplitudes at different positions:

$$f_{SW}(x,t) = \left[\mathcal{A}(x)\right]\sin\left(\frac{2\pi t}{T}\right), \qquad \mathcal{A}(x) \equiv 2A\sin\left(\frac{2\pi x}{\lambda}\right) \tag{1.5.7}$$

It is not hard to visualize this wave – it is a sine function along the x-axis, which remains in place ("standing") as its displacement at various positions oscillates with time. That is, it is exactly like the standing wave depicted in Figure 1.5.6, with the left end being the origin. There are several things to note about this standing wave:

- There are fixed points that occur at specific positions on the standing wave (when the sine function of position vanishes), called *nodes*. These are separated by a distance equal to one half the wavelength of the traveling waves. We will say that the "wavelength" of the standing wave equals the wavelength of the traveling waves that are forming it.
- The maximum displacement of the standing wave only occurs at specific positions, called *antinodes*, which are also separated by a distance of one half wavelength.
- The maximum displacement of the standing wave ($2A$) occurs when the sine functions of time and position both equal 1, and it is twice the amplitude of the traveling waves that compose it. This is referred to as the "amplitude" of the standing wave.
- The period of oscillation of the standing wave (the time it takes to get back to where it started) is the same as the period of the traveling waves that compose it (T).
- This one-dimensional function cannot be written in the form $f(x \pm vt)$, but it *is* a solution of the wave equation. The reason is that the ambiguity of the sign of $\pm v$ is washed away in the square of v in the wave equation. We originally described a wave as a phenomenon that transports energy from one position to another, and a standing wave clearly does not do this, so it is probably better described as a special time-varying interference pattern.

Note that we could have insisted that the end of the medium at the origin is free rather than fixed. This would result in no phase shift for the reflected wave, and it is left as an exercise to show that this results in a standing wave function with two cosine functions replacing the two sine functions in Equation 1.5.7.

Standing Wave Harmonics

The formula for a standing wave is still rather abstract, in that it really only restricts the behavior of the standing wave at a single point (the origin), and assumes that we know the wavelength and period. Here we will consider a different restriction, one that is more useful for physics applications. We will define a distance between two endpoints, and insist that a standing wave forms between them. We also need to specify if the ends are held fixed or are free. If an end is fixed, it must be a node of the standing wave, and if it is free, it must be an

antinode, so this greatly restricts the standing waves that can be formed. In particular, it puts very specific restrictions on the possible wavelengths a standing wave can have.

Let's start with the longest possible wavelength that a standing wave can have if its two ends are separated by a distance L. There are three possibilities in terms of the node/antinode endpoints: Both ends can be fixed (nodes), both ends can be free (antinodes), or there can be one of each type at the two ends. Note that in the first two cases, the distance between the two ends must equal one-half wavelength, while in the third case the distance between the ends is one-quarter wavelength (again, we are looking specifically at the *longest possible wavelengths* to satisfy these conditions).

Figure 1.5.7 – Longest Wavelength Standing Wave – Both Ends Fixed

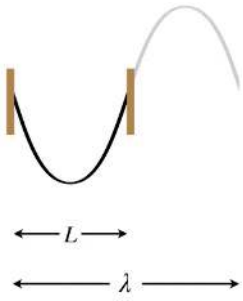


Figure 1.5.8 – Longest Wavelength Standing Wave – Both Ends Free

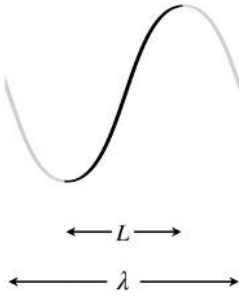


Figure 1.5.9 – Longest Wavelength Standing Wave – One End Free, One End Fixed

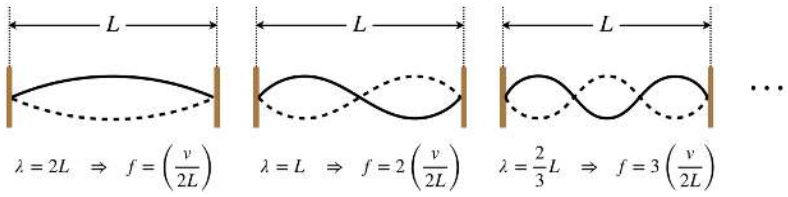
In the Figures above, the dark curves indicate the extent of the medium (i.e. that which is actually vibrating). The gray portions are only added to show the actual wavelength λ of the standing wave and how it relates to the length L of the medium.

These are not the only standing waves possible for the given length L. An infinitude of additional standing waves are possible with shorter wavelengths as well, but *only certain wavelengths will work*. We can characterize these by the number of nodes or antinodes present. The set of standing waves allowed for a given length of medium are called the *harmonics* of the system. The harmonic with the longest possible wavelength is called the *fundamental harmonic*, and the rest are numbered up from there according to frequency.

Speaking of frequency, it must be noted that the frequency of oscillation of a standing wave changes from one harmonic to the next. As we have already seen, the wavelength of the standing wave equals the wavelength of the the two opposite-moving traveling waves, and the period (or frequency) of the standing wave matches the traveling wave periods as well. If we consider a shorter-wavelength standing wave (one with more antinodes), then the wavelength of the traveling waves that make it must also be shorter. But the medium is unchanged, so the speed of those traveling waves must remain the same. This can only be true if the frequency of the traveling wave has gone up, which means the frequency of the standing wave must also go up.

We define the n^{th} harmonic as that harmonic with a frequency that is n times as great as the fundamental harmonic. Let's see what that means for the three possible endpoint conditions. We'll start with both ends fixed. For the fundamental harmonic, we found that the wavelength was double the length of the medium. The next shortest wavelength would include a single node between the two endpoints, and as allowable standing waves get shorter and shorter, we simply keep adding nodes, one at a time (this can be described as fitting an additional half-wavelength between the endpoints each time).

Figure 1.5.10 – Harmonics, Both Ends Fixed



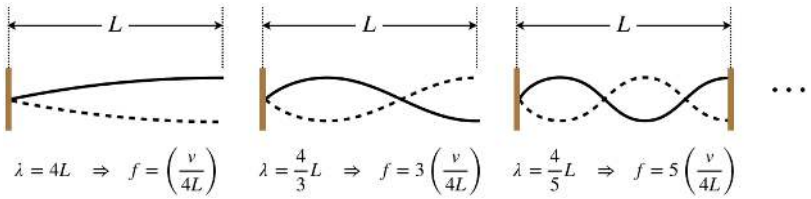
The pattern for this case is clear: The n^{th} possible standing wave has a frequency of n times the fundamental harmonic, which means that the each time we add an antinode, we get the next-highest harmonic, and the number of antinodes equals the order of the harmonic. Mathematically we summarize it this way (v is the speed of the traveling wave on the string):

$$\lambda_n = \frac{2L}{n} \quad \Rightarrow \quad f_n = n\left(\frac{v}{2L}\right), \qquad n = 1,\ 2,\ 3,\ \ldots \tag{1.5.8}$$

If we look at both ends free, we find that the same pattern emerges, which should be clear from the fact that the wavelength of the fundamental harmonic is the same when both ends are free or fixed. The only difference between the two cases are that we count the number of *nodes* to get the harmonic in the both ends free case, not antinodes, as we did for the case of both ends fixed.

When only one end is free, we get a different result when it comes to counting harmonics. We still squeeze additional half wavelength between the endpoints for the next possible wave, but the frequencies of the harmonics have a different relationship to the fundamental.

Figure 1.5.11 – Harmonics, One End Fixed, One End Free

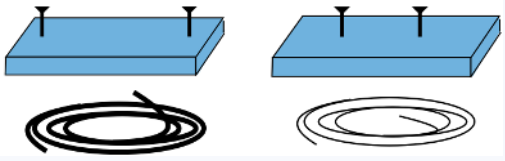


Notice that in this case each time a half-wavelength is added, the frequency jumps an amount equal to *two* fundamental harmonics. So for the case of one end fixed and the other end free, the allowed standing waves include no even-numbered harmonics. Mathematically:

$$\lambda_n = \frac{4L}{n} \quad \Rightarrow \quad f_n = n\left(\frac{v}{4L}\right), \qquad n = 1,\ 3,\ 5,\ \ldots \tag{1.5.9}$$

Example 1.5.1

Two boards with nails separated by different distances are combined with uniform strings that have different lengths and masses, to form one-string guitars.



Show that these guitars make tones of the same pitch (as determined by their fundamental harmonics) when the following quantity α is the same for both:

$$\alpha = \frac{FL}{md^2},$$

where F is the tension in the string, L is the length of the string, m is the mass of the string, and d is the distance separating the nails.

Solution

The frequencies of the fundamental harmonics must be equal, which means:

$$f_A = f_B \quad \Rightarrow \quad \frac{v_A}{\lambda_A} = \frac{v_B}{\lambda_B}$$

With both strings exhibiting their fundamental harmonics, they both have the same relationship between their wavelengths at the nail separations – in both cases the nail separation is half a wavelength:

$$\lambda_A = 2d_A\ , \quad \lambda_B = 2d_B \quad \Rightarrow \quad \frac{v_A}{d_A} = \frac{v_B}{d_B}$$

The speed of the traveling waves that create the standing wave is determined by the tension and the string density. The string is uniform, so its density is the ratio of the string's mass and its length. Therefore:

$$\mu = \frac{m}{L} \quad \Rightarrow \quad v = \sqrt{\frac{F}{\mu}} = \sqrt{\frac{FL}{m}}$$

Plugging this in above gives:

$$\frac{1}{d_A}\sqrt{\frac{F_A L_A}{m_A}} = \frac{1}{d_B}\sqrt{\frac{F_B L_B}{m_B}} \quad \Rightarrow \quad \frac{F_A L_A}{m_A d_A^2} = \frac{F_B L_B}{m_B d_B^2} = \alpha$$

Example 1.5.2

Since the time of ancient Rome, commanders of armies have known that it is prudent to have the troops break stride in their march when crossing a wooden bridge. This is because if the troops march in a synchronized cadence, they produce a periodic coordinated jolt to the bridge, which could excite one of its natural harmonic frequencies, causing a standing wave to develop in the bridge.

a. *If a marching army does create a standing wave in the bridge, what aspect of this standing wave (A, f, T, λ) would be directly responsible for causing the bridge to break apart? Explain.*
b. *Suppose a platoon comes upon a wood & rope bridge that is supported only at its two ends. The commander stops the company short of the bridge and shakes the nearest end of the bridge, testing to see if it seems strong enough to hold the troops. The bridge ripples all the way down its length, with the pulse reflecting off the other side and returning, for a round-trip time of about $2.5s$. Find the frequency of the fundamental harmonic standing wave for this bridge.*
c. *The commander decides the bridge is sturdy, and makes the tragic decision to order the company to march on. Their marching pace and spacing is such that a standing wave forms in the bridge, and the ropes break when the center of the bridge dips well below its usual point as two outer parts of the bridge surge upward. Find the marching pace of the company in steps per second.*

Solution

a. The bridge breaks apart when various components are stretched and separated so far that they can no longer hold together. This deformation of the bridge is a direct result of the amplitude of the standing wave. Put another way, the violence with which the bridge shakes is a measure of the energy put into it, and the energy in the standing wave is a function of its amplitude.

b. Call the length of the bridge L and the speed of the wave v. The time it takes the wave to travel two lengths of the bridge is given as 2.5s, and in terms of the distance traveled and speed of the wave, we have:

$$2.5s = t = \frac{2L}{v}$$

For the fundamental harmonic, the length of the bridge (which is fixed at both ends) is one-half wavelength, so plugging in a half wavelength for L gives:

$$2.5s = \frac{2\,(0.5\lambda)}{v} = \frac{\lambda}{v} = \frac{1}{f} \quad \Rightarrow \quad f = 0.40Hz$$

c. The description of the standing wave makes it clear that it has three antinodes, which means it is the 3rd harmonic. The two ends of the wave are fixed, so the third harmonic occurs at three times the fundamental frequency, or $1.2Hz$ For the footfalls to excite this harmonic, they need to match this frequency, so the marching pace is 1.2 steps per second.

Alert

If you are a musician, you likely have heard of overtones. At the simplest level (like one-dimensional standing waves with both ends fixed or free), these are synonymous with harmonics. But in the one-dimensional case when one end is free, or in the case of two-dimensional

standing waves (like those produced by a membrane on a drum), these definitions diverge. We will not go into the details of these divergences, and the rare times we refer to the "first overtone," we will mean simply the next highest allowable harmonic.

Energy In Standing Waves

Let's consider the case of a second harmonic standing wave on a string between two fixed ends. We know the following things to be true:

- Between the endpoints, there is exactly one full wave moving right and an identical full wave moving left at all times.
- Each of these waves contains an amount of energy that is proportional to the square of its amplitude.
- The standing wave has an amplitude twice as great as the amplitude of each individual traveling wave.

So the question is, doesn't doubling the amplitude mean the standing wave has *four times the energy* of an individual traveling wave? If it does mean this, where does this extra energy come from, if there are only two such waves providing energy?

This apparent paradox stems from something we discussed earlier – *it is dangerous to think of a standing wave in the same context as a traveling wave*! This is especially true in the context of energy distribution. Let's consider the energy of a single particle in a medium as a harmonic wave passes through. Such a particle is following harmonic motion, so if it happens to be at the crest or the trough of the wave, then its kinetic energy is zero, while its potential energy is a maximum. Conversely, if it is at the middle, then it has its maximum kinetic energy and no potential energy. *But no matter where it is in the phase of the wave, its energy is the same.*

Now compare that with a particle in the medium of a standing wave. If the particle is at a node, then it never moves, and is never displaced from equilibrium, so its energy is zero. A particle at an antinode, on the other hand, has lots of energy. The amplitude of its harmonic motion is twice the amplitude that a particle on one of the two traveling waves would have, if the second wave wasn't there.

The bottom line is that the standing wave, *when viewed as an interference pattern*, clearly just redistributes the energy of the two traveling waves (which themselves distribute the energy uniformly), taking energy away from some regions of the medium and giving it to others. With some clever calculus, we can show that this works out exactly.

If the string has a linear density of μ, then an infinitesimal segment of the string of length dx has a tiny mass of $dm = \mu dx$. A traveling wave has every such infinitesimal segment oscillating with the same amplitude, so every particle on the string contributes the same infinitesimal energy, and adding these contributions for a full wavelength gives:

$$E_{traveling\ wave} = \int_0^\lambda \frac{1}{2} dm\ \omega^2 A^2 = \int_0^\lambda \frac{1}{2} \mu dx\ \omega^2 A^2 \tag{1.5.10}$$

The density of the string, the frequency of oscillation, and the amplitude of oscillation are the same for every particle in the string, so they do not vary with x, which makes the integral simple to perform:

$$E_{traveling\ wave} = \frac{1}{2} \mu \lambda \omega^2 A^2 \tag{1.5.11}$$

The segments of the string for a standing wave behave differently. They all vibrate harmonically (with the nodes exhibiting zero vibration), but they reach different maximum displacements. Put another way, a standing wave is a collection of an infinite number of harmonic oscillators, all with different amplitudes. So we need to write down the energy for each particle, and add them all up. The waveform of the standing wave gives us the amplitude (which we will call $a(x)$) of particle oscillation as a function of position x, so from Equation 1.5.7, we have:

$$Amplitude\ of\ medium\ at\ x = a(x) = 2A \sin\left(\frac{2\pi x}{\lambda}\right) \tag{1.5.12}$$

Recall that A is the amplitude of the two traveling waves that are interfering. The energy of this tiny piece of the string is:

$$dE = \frac{1}{2} dm\ \omega^2 \left[a(x)\right]^2 \tag{1.5.13}$$

Putting in $dm = \mu dx$ and $a(x)$ and integrating over the full wavelength of the wave, we get:

$$E_{standing\ wave} = \int_0^\lambda \frac{1}{2} \mu dx\ \omega^2 \left[2A \sin\left(\frac{2\pi x}{\lambda}\right)\right]^2 \tag{1.5.14}$$

Making the substitution $u \equiv \frac{2\pi x}{\lambda}$ leaves an integral that is easy to look up, and gives the following answer:

$$E_{standing\ wave} = \frac{\mu\lambda\omega^2 A^2}{\pi}\int_0^{2\pi} \sin^2 u\ du = \mu\lambda\omega^2 A^2 \tag{1.5.15}$$

Comparing this with the answer for the traveling wave, we see that it is twice as much – the energy content of the standing wave equals the sum of the energies of the two traveling waves that interfere to create it. If we want to write the energy contained in a single wavelength of a standing wave in terms of the standing wave's "amplitude" (the amplitude of the harmonic motion located at an antinode), we have:

$$\mathcal{A} = 2A \quad \rightarrow \quad E_{standing\ wave} = \frac{1}{4}\mu\lambda\omega^2\mathcal{A}^2 \tag{1.5.16}$$

CHAPTER OVERVIEW

3: Physical Optics

3.1: Light as a Wave

What is "Waving"?

The jump from mechanical waves to sound was a difficult one, mainly because the "displacement" of the wave changed from matter that oscillates back-and-forth, to (in the case of sound in a gas) oscillations in pressure or density. This difficulty gets greatly magnified for the case of light. We know that light is a wave based on how it behaves – it exhibits the same properties of other waves we have examined – it interferes with itself, it follows an inverse-square law for intensity (brightness), and so on. But we also know that we can see light from the sun, moon, and stars, which means that light waves can travel through the vacuum of space. Unlike every other wave we have seen, it doesn't require any medium at all! So what do we use as the "displacement" for our wave function?

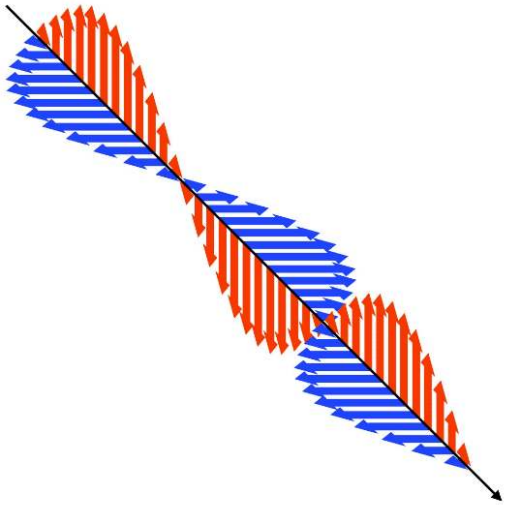
Back in the 19th century, physicists studied extensively the subjects of electricity (lightning, shocking your finger on a doorknob, balloons sticking to your hair, etc.) and magnetism (compasses, sticking things to your refrigerator, etc.). It started becoming clear that the two forces, while different, had some links. Electric currents were found to affect compass needles, and magnets moving near wires were found to create electric currents. It all came together with an amazing (for the time) effort in mathematics by a man named James Clerk Maxwell. He showed that changing electric fields could induce magnetic fields, while changing magnetic fields could in turn induce electric fields. This is a recipe for propagation of these fields, and the equation he derived for this propagation was exactly the wave equation! So he predicted, from results taken from experiments in electricity and magnetism, that an *electromagnetic wave* could be produced. The wave equation included physical constants from both electricity and magnetism, and extracting the wave speed from this equation resulted in a number Maxwell was already familiar with – the speed of light. It is traditional to denote this speed with a lower-case 'c':

$$c = 3.0 \times 10^8 \frac{m}{s} \tag{3.1.1}$$

So the "displacement" of such a wave is actually the electric and magnetic field vectors (both types of fields are waving simultaneously, with each inducing the other) in the space through which the light wave is traveling. Don't worry that this doesn't make much sense right now – it should be a bit clearer when you get to Physics 9C and study electricity & magnetism.

Okay, so for light we now have the wave speed and the "displacement." Let's address a couple other elements of light as a wave. First, a medium is not needed, as electric and magnetic field can exist in a vacuum. The presence of a medium (such as air or water) *does* effect the electric and magnetic fields, because media are made up of atoms, which are composed of positive and negative electric charges. Because of this, the speed of light within a medium is different (slower) than its speed in a vacuum. Mathematics and experiments show that light is a transverse wave – the electric and magnetic field vectors point in directions that are perpendicular to the direction of motion of the light wave (and as it turns out, they also rare always perpendicular to each other).

Figure 3.1.1 – Electromagnetic Wave



The red arrows in the figure above represent electric field vectors, and blue arrows magnetic field vectors. Specifically, this is a *plane-polarized* EM wave, which means the field vectors of a given type remain in a single plane. We will discuss plane polarization soon, but it should be noted that EM waves do not have to behave this way, so long as the electric and magnetic field vectors remain perpendicular to each other and to the direction of motion. For example, a *circularly polarized* EM wave features electric and magnetic field vectors that circulate their directions (while remaining perpendicular to each other and the direction of motion) as the wave propagates, like the hands of an analog clock, and can do so in a clockwise or counterclockwise manner.

Finally, we need to say two things about light perception. For sound, intensity (proportional to amplitude-squared) is perceived as loudness, and for light it is brightness. For sound, frequency is perceived as pitch, and for visible light it is perceived as color. The qualification "visible" must be appended because we can only see a very limits spectrum of light frequencies, the rainbow of colors often described with the acronym ROYGBIV (Red, Orange, Yellow, Green, Blue, Indigo, Violet). The red end of the visible spectrum exhibit the lowest frequencies, and the violet the highest. But of course light waves can come in frequencies much lower and much higher, and at various arbitrary cutoffs, they are given names you have probably heard before. In order of increasing frequency below the red end of the visible spectrum we have: *radio waves*, *microwaves*, and *infrared*; and above the violet end of the spectrum: *ultraviolet*, *x-rays*, and *gamma rays*.

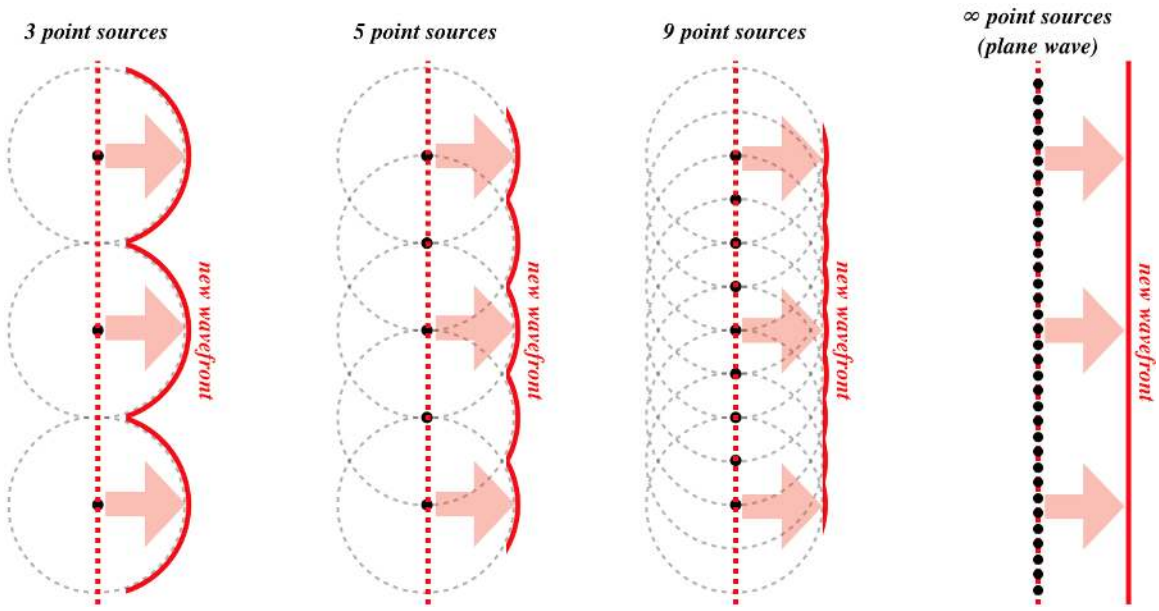
Huygens's Principle

When we discussed the case of a wave on a string, we said that the wave causes each particle on the string to vibrate up-and-down in harmonic motion. It should therefore not be surprising that if we grab the string at a single point and force it to vibrate in harmonic motion, that a wave will propagate away from that point. In fact, this gives us a way of describing how the wave propagates: The wave causes a single point to oscillate, which in turn causes a wave to be generated, which then vibrates another point, and so on. In the 17th century a Dutch scientist named Christian Huygens generalized this idea to three dimensions. The principle which now bears his name can be stated this way:

Every (3-dimensional) wave propagates by having every point on a wavefront being an independent generator of a new spherical wave, and the interference of all of those individual spherical waves results in the overall wave observed.

When we look at a single point light source, the farther away it is, the flatter the light wavefronts will be when they reach us. When the source is very far away (e.g. the sun), then the wavefronts are essentially flat. We call waves with such flat wavefronts *plane waves*, for obvious reasons. But now the question arises, “If Huygens’s principle is valid, how can plane waves occur?” After all, each point on the plane wave behaves as a point source of a spherical wave. Let's look at the spherical wave contributions of many point sources on a plane. We'll do this gradually, starting with just a few points on a plane, and filling in the spaces between them little-by-little:

Figure 3.1.2 – Plane Wave from Huygens's Principle

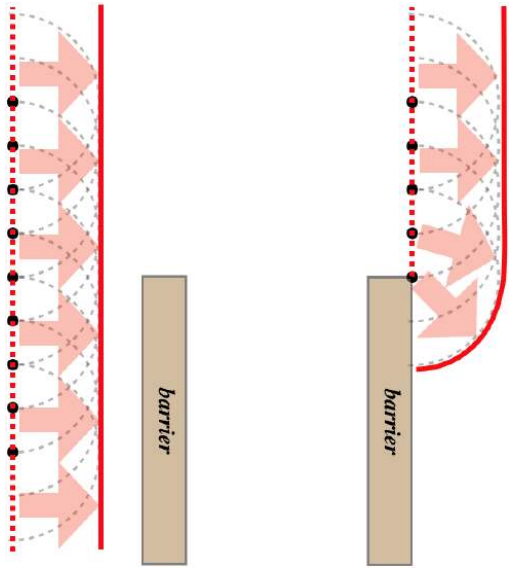


One might ask why a plane wave only propagates in a single direction. Suppose a plane wave propagating to the right. If each new wavefront becomes a source for a new wave, why don't waves come out of it in both directions? It is difficult to express in a simple diagram like the one above the effects of superposition, but the short answer is that there is destructive interference between all of the previous wavefronts and the new one, which results in zero wave energy traveling "backwards."

It should also be noted that a plane wave is a one-dimensional wave, which means that its intensity does not drop off with distance. But the intensities of the spherical wavelets do follow an inverse-square law. So if they get weaker with distance, why don't plane waves? The reason is that the farther a wavelet travels, the more other wavelets it encounters. These encounters result in constructive interference, bolstering the amplitude (and therefore the intensity) The rate at which the wavelets encounter other wavelets and constructively interfere is exactly enough to compensate for each wavelet losing its own individual intensity, maintaining the plane wave's intensity.

Where Huygens's principle becomes particularly useful is in explaining what happens when a plane wave encounters a barrier. A plane wave moves straight ahead because there is destructive interference of the wavelets in other directions. But a barrier removes a number of wavelets by either absorbing or reflecting the part of the wavefront from which those wavelets were going to spawn. The result is that the wave "bends around corners," a phenomenon known as *diffraction*.

Figure 3.1.3 – Diffraction from Huygens's Principle



Like other wave phenomena, this is not unique to light. Ocean waves diffract around barriers like reefs, peninsulas, and docks. It's certainly possible to hear a sound made from around a corner. It should be noted that the part of a wave that diffracts around a corner is no longer a plane wave, and is subject to the reduction in intensity the farther it travels. Of course reflections of waves are also responsible for their ability to change direction in the presence of barriers, but the phenomenon of diffraction in conjunction with interference leads to other important observable properties that we will deal with next.

Alert

You should be aware that diffraction is so intimately tied up with the interference effects that it causes (the subjects of the next few sections) that many physicists use the word "diffraction" to indicate the interference phenomena themselves, rather than the "going around corners" definition.

3.2: Double-Slit Interference

Splitting a Light Wave into Two Waves that Interfere

We now return to the topic of static interference patterns created from two sources, this time for light. As with sound, we first need to start with two light sources that are at the same frequency. In the case of light, we say that the sources are *monochromatic*. For sound we were able to keep track of the starting phases of sounds coming from separate speakers by connecting them to a common source, but for light it's a bit trickier. There simply isn't a way to coordinate the phases of light waves coming from two independent sources (like two light bulbs). Light waves from multiple independent sources have phases that are essentially distributed randomly, resulting in a variety of light referred to as *incoherent*. In fact, even light from a single source such as an incandescent bulb is incoherent, because the vibrations of the various electrons that create the waves are not coordinated. It turns out (for complicated reasons we won't go into) that after light travels a long distance the coherence of the waves grows (so light from the sun is highly coherent), but for experiments with light sources located here on Earth we are forced to use lasers, which do produce coherent light. Again, the reason that laser light is coherent is complicated, and outside the scope of this class.

Even with the coherence available from a single laser, we cannot coordinate the phases of two separate laser sources, so we need to somehow use the waves coming from a single laser source. We do this by directing the light from a single source through two very narrow adjacent slits, called a *double-slit apparatus*. Huygens's principle assures us that then each slit becomes a source for a spherical wave emanating from the position of each slit, and since the wavefront reaches each slit at the same time, the two sources start in phase, just like the tones coming from two speakers attached to the same source.

Okay, so to get an idea of the interference pattern created by such a device, we can map the points of constructive and destructive interference. We can do this by mapping what happens to two spherical waves that start at different positions near each other, and specifically keeping track of the crests (solid circles) and troughs (dashed circles). [*Note: The two waves shown are in different colors to make it easier to distinguish them – the actual light from both sources is all the same frequency/wavelength/color.*]

Figure 3.2.1 – Double-Slit Interference

A coherent plane wave comes into the double slit, and thanks to Huygens's principle, the slits filter-out only the point sources on the plane wave that can pass through them, turning the plane wave into two separate radial waves, which then interfere with each other. Whenever a crest meets a trough there is total destructive interference, and whenever two crests or two troughs meet, the interference is (maximally) constructive. We notice a number of things here:

- If we watch the points of total destructive and maximally constructive interference as the waves evolve, they follow approximately straight lines, all passing through the center point between the two slits.

- Because of symmetry, we see that these lines are symmetric about the horizontal line that divides the two slits, and that the center line itself is a line followed by a point of maximal constructive interference.
- These lines alternate in type as the angle increases – the central line is constructive, the lines on each side with the next-greatest angle trace points of destructive interference, the next pair of lines trace points of constructive interference, and so on.
- There are a limited number of these lines possible.

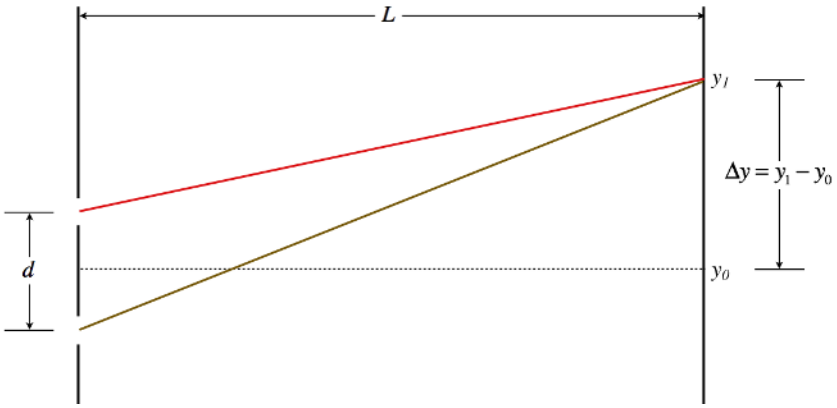
How are these effects *perceived*? Total destructive interference means zero intensity, which is the absence of any wave – darkness. Constructive interference is perceived as bright light, so if we placed a reflecting screen in the way of these light waves, we would see alternating regions of brightness and darkness, called *fringes*. It should be noted that the brightness varies continuously as one observes different positions on the screen, but we are focusing our attention on the brightest and darkest positions only. For the figure above, the screen would exhibit a *central bright fringe* directly across from the center point between the slits, then the first *dark fringes* some distance off-center, then more bright fringes outside of those. It is possible for a double-slit apparatus to produce either more or fewer fringes, depending upon the slit separation and the wavelength of the light. We will discuss the roles these variables play next.

Geometry of a Double-Slit Apparatus

Since we are (for now) only considering the brightest and darkest points, we can work with lines and geometry to get some mathematical answers. As stated above, these points only *approximately* follow straight lines from the center point, so our analysis will necessarily require some approximations. Whenever this is the case in physics, it is important to make a note of the physical features that go into determining the usefulness of the approximation as well as the tolerances we are willing to accept.

We begin by defining the slit separation (d) and the distance from the slits to a screen where the brightness interference pattern is seen (L). We also label some of the quantities related to the position on the screen in question.

Figure 3.2.2a – Double-Slit Geometry

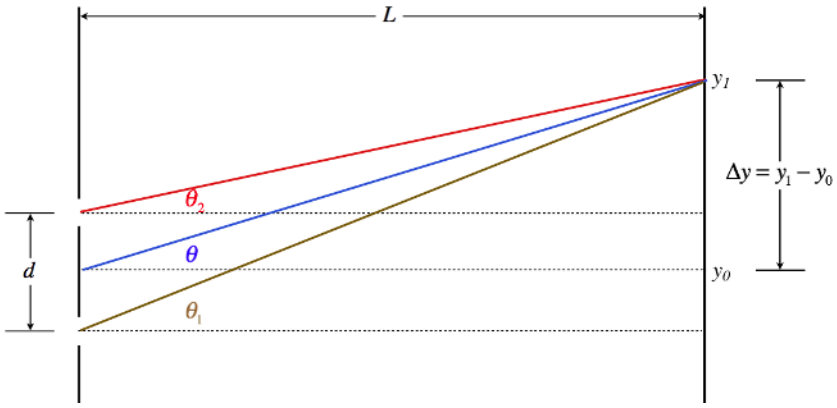


We are looking for those lines that define the destructive and constructive interference, so we want to express things in terms of a line that joins the midpoint of the two slits and the point located at y_1. In particular, we are looking for the angle θ that this line makes with the center line. We already know the center line traces a constructive interference, so our final answer should reflect this for $\theta = 0$.

The key physical argument we make here is that the wave that travels to y_1 from the upper slit has a shorter trip than the wave that gets there from the lower slit. The two waves start at the same time, and in phase, so this difference in distance traveled (Δx) accounts for the phase difference in the two waves that causes interference. So to relate the interference witnessed at y_1 to θ, we need to determine how (Δx) is related to θ.

As a start, we will draw in the line that goes from the midpoint of the slits to y_1, and label a bunch of angles:

Figure 3.2.2b – Double-Slit Geometry



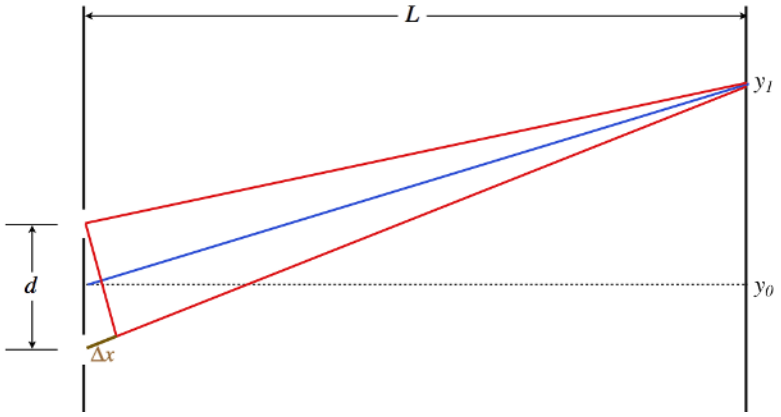
Now we need to do some math and apply some approximations. The tangents of these angles can be written in terms of the sides of the triangles they form:

$$\begin{aligned} \tan\theta_2 &= \frac{\Delta y - \frac{d}{2}}{L} \\ \tan\theta &= \frac{\Delta y}{L} \\ \tan\theta_1 &= \frac{\Delta y + \frac{d}{2}}{L} \end{aligned} \tag{3.2.1}$$

We don't actually require this math to convince us that if the slit separation is very small compared to the distance to the screen (i.e. $d \ll L$), then these three angles are all approximately equal. This is a good approximation, as this phenomenon is typically observed with slits separated by distances measured in fractions of millimeters, while distances to the screen are measured in meters. So henceforth we will make no mention of the angles θ_1 and θ_2.

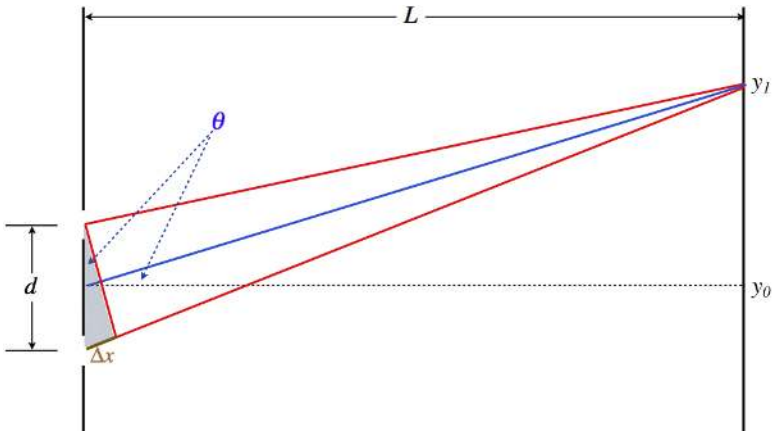
The next step is to break the lower (brown) line into two segments – one with the same length as the top (red) line that touches y_1 but doesn't quite reach the lower slit, and the other with the additional distance traveled, (Δx) that connects the first line to the lower slit. Then with the two equal-length segments, form an isosceles triangle:

Figure 3.2.2c – Double-Slit Geometry



Returning to our angle approximation where the top and bottom lines are approximately parallel, we see that this triangle has approximately two right angles at its base, which means there is a small right triangle formed by the base of the triangle, Δx, and the slit separation d. Imagine rotating the triangle clockwise. The angle at the top of this small triangle closes to zero at exactly the same moment that the blue line coincides with the center line, so this angle equals θ:

Figure 3.2.2d – Double-Slit Geometry



This gives us precisely the relationship between Δx and θ that we were looking for:

$$\Delta x = d \sin\theta \tag{3.2.2}$$

Now all we have to do is put this into the expression for total destructive and maximally-constructive interference. We know that total destructive interference occurs when the difference in distances traveled by the waves is an odd number of half-wavelengths, and constructive interference occurs when the the difference is an integer number of full wavelengths, so:

$$\begin{aligned} &\text{center of bright fringes:} && d\sin\theta = m\lambda \\ &\text{totally dark points:} && d\sin\theta = \left(m+\tfrac{1}{2}\right)\lambda \end{aligned} \qquad m = 0,\ \pm1,\ \pm2, \ldots \tag{3.2.3}$$

Let's take a moment to examine these equations, comparing what they require with the bulleted observations we made above:

- The plus-or-minus values of the integer m confirms that the fringes are symmetrically reflected across the center line.
- The case of $m = 0$ for constructive interference corresponds to the center line.
- Moving out from the center, the next fringe of any kind occurs when $m = 0$ for destructive interference. Then the next occurs for $m = 1$ for constructive interference, and so on – the bright and dark fringes alternate.
- Not all integer values of m will work, because the absolute value of $\sin\theta$ can never exceed 1. When the absolute value of m gets too high, this relation cannot possibly hold, placing a limit on the number of fringes. This limit is determined by the ratio of the wavelength to the slit separation.

It is sometimes useful to convert this result into measurements of distances from the center line on the screen, rather than the angle θ. To get this, we need the distance L, which was not necessary for the solution above (other than assuming it is much larger than d). Calling the distance from the center line to the m^{th} fringe y_m, we use the fact that the tangent of the angle is the rise over the run ($y_m = L\tan\theta_m$) to get:

$$\begin{aligned} &\text{center of bright fringes:} && y_m = L\tan\left[\sin^{-1} m\frac{\lambda}{d}\right] \\ &\text{totally dark points:} && y_m = L\tan\left[\sin^{-1}\left(m+\tfrac{1}{2}\right)\frac{\lambda}{d}\right] \end{aligned} \qquad m = 0,\ \pm1,\ \pm2, \ldots \tag{3.2.4}$$

So long as we are careful, we can simplify this with a second approximation. If the *angle is small*, then the tangent and sine of that angle are approximately equal. This simplifies the above result to:

$$\text{for small } \theta: \qquad \begin{aligned} &\text{center of bright fringes:} && y_m = m\frac{\lambda L}{d} \\ &\text{totally dark points:} && y_m = \left(m+\tfrac{1}{2}\right)\frac{\lambda L}{d} \end{aligned} \qquad m = 0,\ \pm1,\ \pm2, \ldots \tag{3.2.5}$$

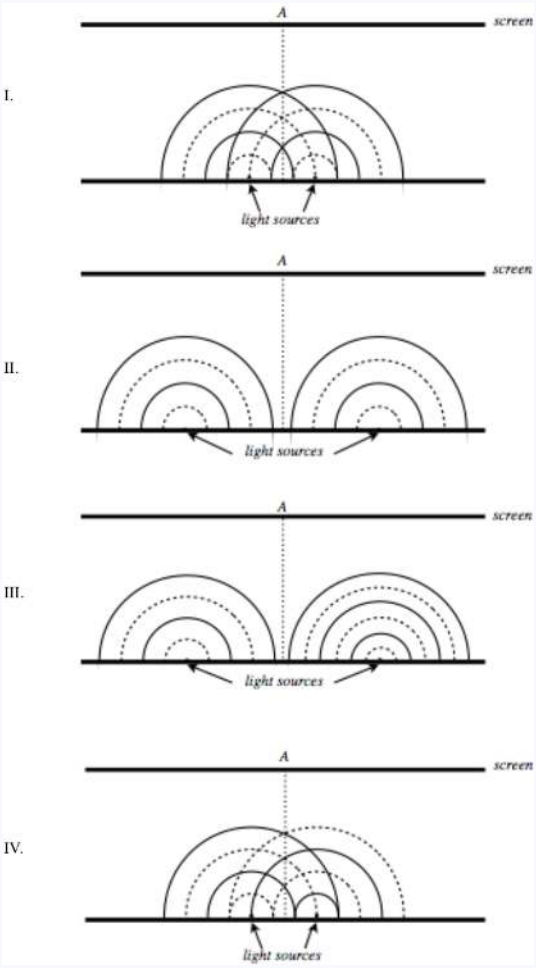
This shows us that for small angles, fringes of the same type are equally-spaced on the screen, with a spacing of:

$$\Delta y = \frac{\lambda L}{d} \tag{3.2.6}$$

Example 3.2.1

Below are four depictions of two point sources of light (not necessarily caused by two slits), using the wave front model. These depictions are "snap shots," meaning they are frozen at an instant in time, but the questions below pertain to what happens in real time. Solid lines represent crests, and the dotted lines troughs. For each case, determine the following, and provide explanations:

a. *Will these sources create a fixed interference pattern on the distant screen?*
b. *If there is an interference pattern, what will appear at the point A on the screen, which is directly across from the midway point between the two sources? That is, will it be a bright fringe, a dark fringe, or something in-between?*
c. *If there is an interference pattern, how many bright fringes will appear on the screen?*



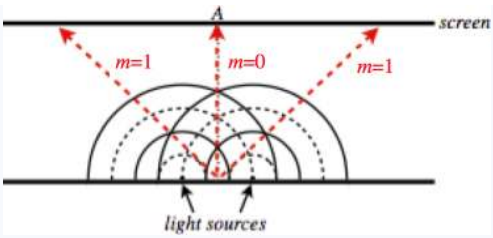
Solution

I.

a. Yes. The sources have the same wavelength (and therefore the same frequency), which means that their interference pattern will not have a time-dependent element to them (i.e. they will not provide the light equivalent of "beats").

b. Bright fringe. The two waves start in phase, and travel equal distances from the sources to get to the center line, so they end up in phase, resulting in constructive interference.

c. One can see by drawing lines through the crossings of crests & troughs that only 3 such lines will strike the screen (parallel to the screen crests match with troughs, so those will not give bright fringes):



We can do this mathematically by noting that these waves start in phase, which means this is equivalent using $d\sin\theta = m\lambda$ for bright fringes, and by noting from the diagram that the two slits are separated by a distance of 1.5λ The fact that $\sin\theta$ can never be greater than 1 puts a limit on m. This is an integer that can't be greater than 1.5, so its maximum value is 1, leaving us with 3 bright fringes.

II.

a. Yes. The same reasons as given above for (I.a) apply.

b. Bright fringe. Same reasoning as II.b

c. Now it is not possible (or at least exceedingly difficult) to draw in the lines that lead to constructive interference, so the mathematical method is the only practical approach. This time the slit separation d is clearly more than 4λ and less than 5λ. This means that the highest integer value of m is 4. With 4 bright fringes on each side of the central bright fringe, the total number is 9.

III.

a. No! These two waves have different wavelengths, and therefore different frequencies, which means that when they interfere, the resulting wave's amplitude (and therefore the brightness) will be time-dependent.

b. N/A

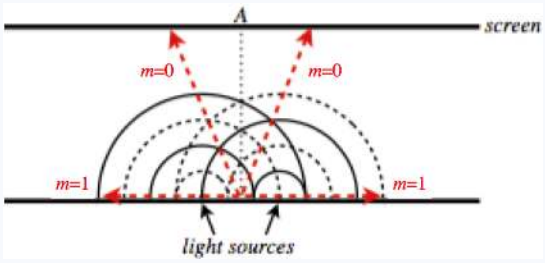
c. N/A

IV.

a. Yes. Back to equal wavelengths.

b. Dark fringe. These waves start out-of-phase by π radians, so when they travel equal distances, they remain out-of-phase.

c. We can once again draw the lines that follow the paths of constructive interference:



The light sources are separated by 1.5λ as they were once before, but now the condition for constructive interference is different, to make up for the starting phase difference. It is now: $d\sin\theta = (m+1/2)\,\lambda$. We see that there are now two bright spots associated with $m = 0$, and although there is a solution for $m = 1$, it gives $\theta = \frac{\pi}{2}$, which means the light never reaches the screen, so the number of bright spots on the screen is 2.

Double Slit Intensity Pattern

The answers above only apply to the specific positions where there is totally destructive or maximally constructive interference. What about the points in between? For this answer, we return to Equation 1.4.10, which relates any phase difference of two waves to the intensity of the wave in comparison to its maximum intensity (when maximal constructive interference occurs). As noted earlier, the only source of phase difference is the distance traveled by the two waves, so:

$$\left.\begin{aligned} I &= I_o\cos^2\left(\frac{\Delta\Phi}{2}\right) \\ \Delta\Phi &= \frac{2\pi}{\lambda}\Delta x \\ \Delta x &= d\sin\theta \end{aligned}\right\} \quad\Rightarrow\quad I(\theta) = I_o\cos^2\left[\frac{\pi d\sin\theta}{\lambda}\right] \tag{3.2.7}$$

It's easy to see that this works correctly for the specific cases of total destructive and maximal constructive interference, as the intensity vanishes for the destructive angles, and equals I_o for the constructive angles. If the angle is small, then we can approximate this answer in terms of the distance from the center line:

$$I(y) = I_o \cos^2\left[\frac{\pi y d}{\lambda L}\right] \tag{3.2.8}$$

Activity

To see all the features of double-slit interference, check out this simulator. To simulate double slit interference for light, take the following steps:

1. Select and click on the "Interference" box.
2. In the control box, click the laser icon:  .
3. In the control box, click the "Screen" toggle box to see the fringes.
4. Click on the green buttons on the lasers to start propagating the light waves.
5. In the control box, you can adjust frequency and slit separation to see the effects on the interference pattern.
6. You can click on the intensity toggle box in the control box to see the graph of the intensity at the screen, as described by Equation 3.2.8.
7. You can even explore this phenomenon *quantitatively* (i.e. check the math derived above) by reading the slit separation in the control box, dragging measuring devices from above the control box, and pausing the simulation.

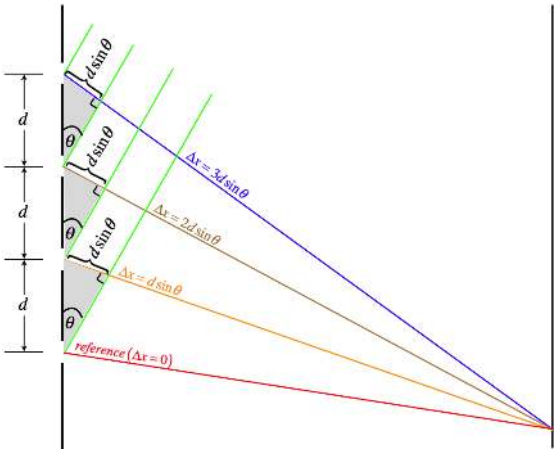
3.3: Diffraction Gratings

Adding More Slits

After having determined the interference pattern associated with two slits, it makes one wonder what would happen if many more (equally-spaced) slits are added. We can recycle our geometrical analysis from the double slit problem to answer this question. Let's look at the example of four slits.

We begin once again with the assumption that the distance to the screen is significantly larger than the separation of adjacent slits: $d \ll L$). Starting with the lowest slit of the four as a "reference" and repeating the double-slit geometry for each slit going up from there, we have a diagram that looks like this:

Figure 3.3.1 - Geometry of Four Slits



The Δx in each case is the difference in distance traveled compared to the reference slit. So the extra distance traveled by the wave following the blue path is three times as great as the extra distance traveled by the wave following the orange path.

Alert

This diagram is blown-up for clarity, but doing so makes the angles quite different from each other. With the proper scale in place the approximations of equal angles (and equal Δx's throughout) would be more apparent.

Okay, so as our first task, we will look for the position where the first bright fringe is located. For this to occur, we need all four waves to be in phase, which means that Δx has to be a full wavelength, giving us the same formula for bright fringes that we found for the double slit:

$$d \sin\theta = m\lambda\ , \qquad m = 0,\ \pm 1,\ \pm 2,\ \ldots \tag{3.3.1}$$

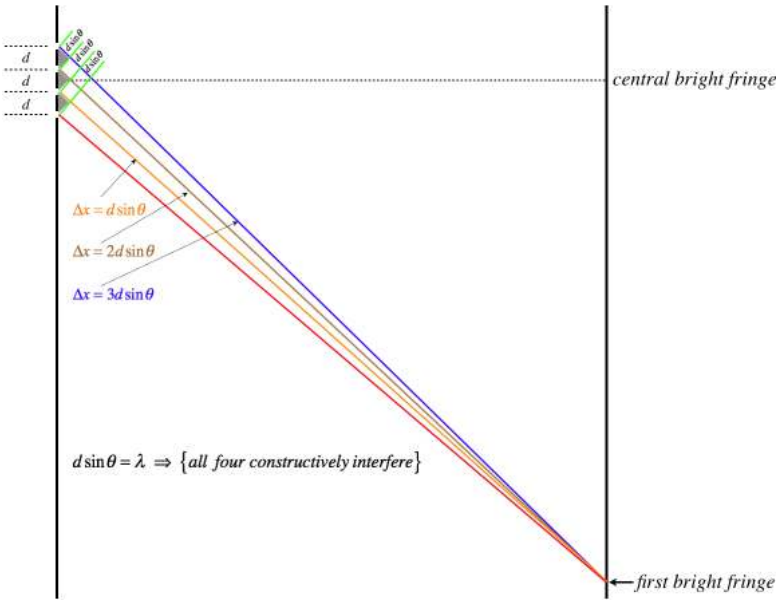
[*It should be noted that the positions of the fringes on the screen are measured from the horizontal line passing through the center of the collection of slits, as we did with the double slit.*]

Does this mean that the result for several slits is identical to that of the double slit? Certainly not! First of all, there are many more sources of light, all interfering constructively, which means that the bright fringes are much brighter. How much brighter? Well, with four slits, as in the example here, the amplitude of a single slit is multiplied by 4, making the intensity (which goes as the square of the amplitude) 16 times greater than a single slit. For the double slit, the intensity was increased by a factor of 4 (the amplitude was doubled). Therefore doubling the number of slits increased the intensity of the bright fringes by a factor of 4. But wait, doubling the number of slits only lets in twice as much energy per second, so how is the intensity increasing so much?

The answer to this puzzle involves how *concentrated* the bright fringes are. All bright fringes have a point of maximum brightness that tapers down to the dark fringes. If the rate at which the brightness tapers down is greater, then the brightness (energy density) near those maximum points can go up, and the energy density near the dark fringes goes down, such that the same total energy hits the screen. But it turns out there is even a little more to it than this, as we will now see.

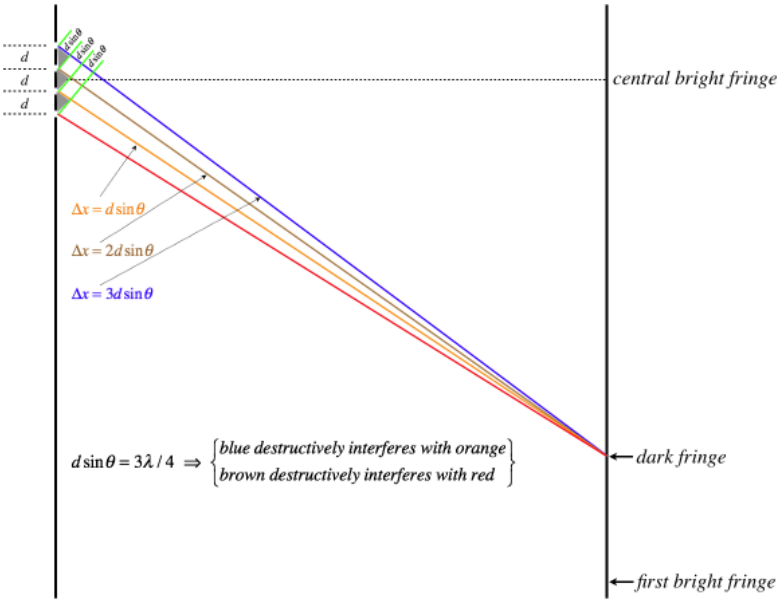
To demonstrate this phenomenon, it becomes necessary to redraw the figure above a little closer to the actual scale. We of course cannot possibly get very close to the actual scale, as slit separations are typically fractions of millimeters, while distances to screens are usually tens or hundreds of centimeters, but we will use what space we can manage. As before, we will use the red line as the reference, and compare the distances traveled by the other three light waves.

Figure 3.3.2a - Finding Dark Fringes



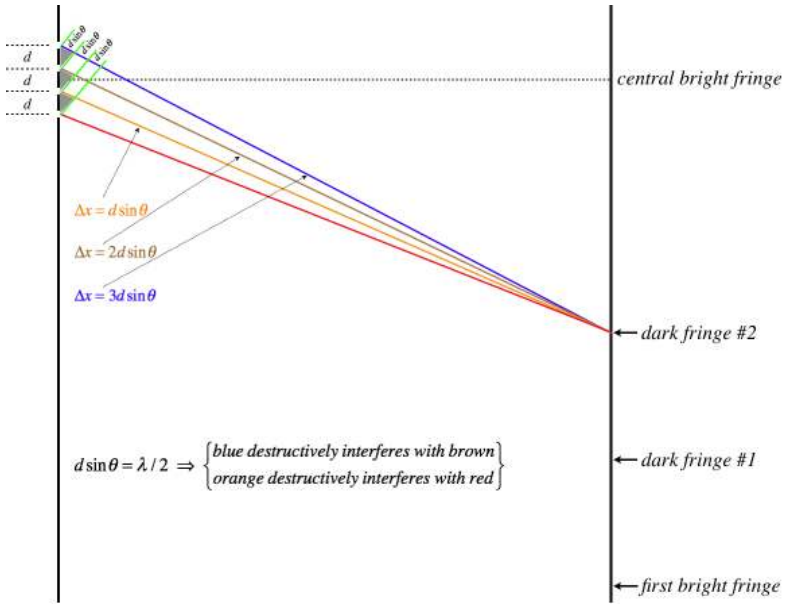
We'll start with the bright fringe, and start working our way closer to the central bright fringe until we hit a dark fringe. Strangely, we find that the first position of total destructive interference we encounter does *not* occur at the halfway point, as it did for the double slit! Note that when the distance $\Delta x = d\sin\theta$ equals *three-quarters of a wavelength*, then the wave that follows the blue path will travel 1.5 wavelengths farther than the wave that follows the the orange path, and as this is an odd number of half wavelengths, these waves will cancel. The same is true for the waves that follow the brown and red paths, which means that position will be completely dark.

Figure 3.3.2b - Finding Dark Fringes



So what happens if we keep going up the screen? We don't find any more maximally-bright fringes (all four waves can't be in phase), but we *do* find another totally dark position. It occurs when the distance $\Delta x = d\sin\theta$ equals *one-half of a wavelength*. In this case, The wave that follows the blue path travels one half-wavelength farther than the wave that follows the brown path, and the waves that follow the orange and red paths also differ in the distance they travel by one half wavelength. So the blue path and red path waves cancel, as do the brown path and yellow path waves, resulting in total darkness.

Figure 3.3.2c - Finding Dark Fringes



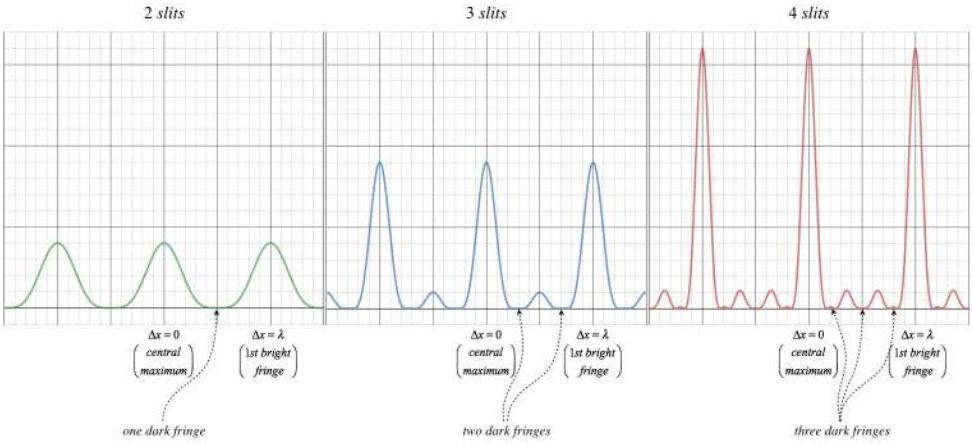
There is one other time when a dark fringe occurs. This happens when the distance $\Delta x = d\sin\theta$ equals *one-quarter of a wavelength*. Once again, alternate slits interfere with each other, as the waves travel distances that differ by a half-wavelength.

We can also show this phenomenon mathematically, by superposing (adding) the wave functions. The waves start in phase at the slits, so all of the phase constants are equal (and we choose them to be zero at $t = 0$), so all that remains of the wave functions is the position dependence. Once again, all that matters are the *differences* in distances traveled with the reference slit (whose difference with itself is zero), so the superposition intensity looks like:

$$\begin{aligned}
I_{2\text{ slits}}(\theta) &= \{f_1(\theta)+f_2(\theta)\}^2 &&= \left\{\cos 0+\cos\left[\left(\tfrac{2\pi}{\lambda}\right)d\sin\theta\right]\right\}^2 \\
I_{3\text{ slits}}(\theta) &= \{f_1(\theta)+f_2(\theta)+f_3(\theta)\}^2 &&= \left\{\cos 0+\cos\left[\left(\tfrac{2\pi}{\lambda}\right)d\sin\theta\right]+\cos\left[\left(\tfrac{2\pi}{\lambda}\right)2d\sin\theta\right]\right\}^2 \\
I_{4\text{ slits}}(\theta) &= \{f_1(\theta)+f_2(\theta)+f_3(\theta)+f_4(\theta)\}^2 &&= \left\{\cos 0+\cos\left[\left(\tfrac{2\pi}{\lambda}\right)d\sin\theta\right]+\cos\left[\left(\tfrac{2\pi}{\lambda}\right)2d\sin\theta\right]+\cos\left[\left(\tfrac{2\pi}{\lambda}\right)3d\sin\theta\right]\right\}^2 \\
I_{n\text{ slits}}(\theta) &= \{f_1(\theta)+f_2(\theta)+\cdots+f_n(\theta)\}^2 &&= \left\{\cos 0+\cos\left[\left(\tfrac{2\pi}{\lambda}\right)d\sin\theta\right]+\cdots+\cos\left[\left(\tfrac{2\pi}{\lambda}\right)(n-1)\,d\sin\theta\right]\right\}^2
\end{aligned} \tag{3.3.2}$$

Putting these functions into a graphing calculator confirms what we found above, as well as what we suspect about n slits – that there are $n-1$ dark fringes between each maximally-bright fringe.

Figure 3.3.3 - Comparison of Interference Patterns by Number of Slits



Notice that the bright fringes for any number of slits occur at the same places as for the double slit (provided they have the same slit separation), and that the number of dark fringes between bright fringes goes up by one every time another slit is added. Also notice that the maximum intensity of the double slit is 4 units, the 3-slit case has a maximum intensity of 9 units, and for 4-slits it is 16 units, as we expect when the amplitude increases by one unit with the addition of each slit. But also notice that the *widths* of the bright fringes get narrower, indicating that the energy becomes more concentrated near the brightness maxima, and less concentrated near the dark fringes.

It turns out that we can mathematically check that the energy is in fact conserved by this mechanism. Recall that the intensity is related to power *density*, which means that if we integrate one of these curves over a full interval of space that the light is landing (say, between adjacent bright maxima), we get a measure of the energy landing in that region per unit time. Once again the graphing calculator comes in handy (unless integrating the intensity functions above is your idea of fun) as areas under these curves between maxima come out to be in relative proportions of 2:3:4 – the total energy landing on the screen every second really is proportional to the number of slits allowing light through!

Adding Many, Many More Slits

We know that the regions where the bright fringes peak get more concentrated light, and that there are more dark fringes between them when the number of slits is increased. One can imagine that in the limit where very many slits are used (a device called a *diffraction grating*), the result is very sharp, very bright lines lines at the points of maximum constructive interference, and darkness everywhere else. As we will see, this will be an extremely useful feature. But there is one assumption we have made here that needs to be emphasized. Because d is so small compared to the distance to the screen, it was easy to ignore the fact that this particular calculation required the assumption that the first bright fringe be farther from the center line than the outermost slit (we assumed that the wavelength was long enough that this had to be true). So creating a sharper interference pattern for a given wavelength of light by adding more slits at the same separation on both sides of the center line has limitations, because when the number of slits gets very large, the added slits go past the bright fringe. However, if more slits are added by *squeezing them closer together* (making d smaller), then for a given wavelength, then not only are there more slits, but the angle to the first bright fringe increases, thanks to the relation $d\sin\theta = m\lambda$.

It is for this reason that diffraction gratings are generally characterized by their *grating density* – the number of slits per unit distance. Of course such a number can be converted into a slit separation: If a diffraction grating has a grating density of 100 slits per cm, then the slits must be separated by $d = \frac{1}{100}cm = 10^{-4}m$. This number can then be used in calculations for the angle at which bright fringes are seen.

It should also be mentioned that like double slits, diffraction gratings do allow for more than one bright fringe (as before, depending upon the ratio of d and λ). For a typical double slit experiment, the goal is usually to show a broad interference pattern – many fringes. If the slit separation is too small, then the angles between the fringes are large, resulting in very few fringes, widely separated, foiling the goal of such an experiment. But use of a diffraction grating has a different goal (very sharp bright fringes), which requires that the slits be separated by much smaller distances. This results in far fewer fringes, separated by large angles. So while the calculation for the angles of bright fringes is the same for both devices, for a given range of wavelengths, their slit separations are usually quite different.

Applications of Diffraction Gratings

It was stated above that sharp bright fringes are very useful in applications. To see why this is so, suppose one wishes to use a diffraction device to measure the wavelength of a monochromatic light. This is straightforward – shine the light through any number of slits with a known slit spacing, and measure the angle at which the first bright fringe is deflected from the central bright fringe, then plug into $d\sin\theta = m\lambda$ (with $m = 1$) and solve for λ. The only real challenge to this procedure is measuring the angle. Of course, if we shine the light onto a screen whose distance we know from the slits, we can measure the distances between the bright fringes, and compute the angle from there. But *still* we have a problem if we want to be precise. If a double-slit is used, then the bright fringe is rather broad, and it might be challenging to get a good measurement of its center. With a diffraction grating, the bright fringe is much better defined. Furthermore, the light we are looking at may not be very intense, and a diffraction crating lets much more of the light in, and the bright fringe is much easier to see than it would be for a double-slit.

But even these two advantages pale in comparison to the third. We have not yet considered what happens if we look at light that is not monochromatic. Suppose the incoming light is a mix of three or four colors. The separate colors don't interfere in a static manner with each other (they can create "beats," but the frequency differences for light are so great that these will not be observable) they only observably interfere with themselves. As such, a beam of light with three colors will exhibit three separate interference patterns when passed though a single device (i.e. they all experience the same slit separation). The wave with color corresponding to the shortest wavelength will have its first bright fringe deflected by the smallest angle. If this light is passed through a double-slit, the interference patterns blend with each other, making it hard to separate the component colors. But a diffraction grating makes three sharp, distinct, first-order bright fringes, making it easy to determine the constituent colors of the incoming light.

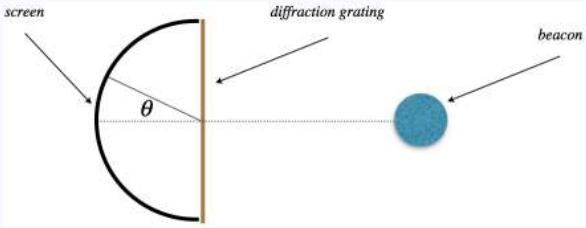
An important part of the fields of chemistry and astronomy is the method of measurement called *spectroscopy*. In Physics 9D, you will learn that matter emits and absorbs light in very peculiar ways. You might think that electrons in atoms can vibrate at any frequency at all and therefore emit or absorb a nice, smooth continuous spectrum of light, but it turns out that they cannot. In fact each atom has a unique “fingerprint” of specific frequencies of light that it emits and absorbs. This means that when light emitted from a certain substance is passed through a diffraction grating, this fingerprint is manifested as a specific set of bright fringes (called *spectral lines*). This means that we can ascertain from a distance (in the case of astronomy, very great distances!) the composition of the matter that is emitting light. These fingerprints are so specific and unique that even if several different substances are emitting light, they can generally be sorted out.

One might worry that since stars are moving relative to the earth, that we might get the elements wrong, since what we will see in the *spectrometer* (a device with a diffraction grating) will measure doppler-shifted wavelengths. But it isn't the exact positions of the spectral lines that tells us the elements emitting the line, but rather their *relative* positions. That is, every spectral line is doppler-shifted, so the "barcode" essentially looks the same for hydrogen regardless of its relative motion, because the whole barcode is just shifted toward longer wavelengths if it is moving away from the spectrometer, and toward the shorter wavelengths if moving toward the spectrometer.

But astronomers can do even more than identify elements in burning stars. We know what the barcode for hydrogen looks like when the source is at rest relative to the spectrometer, so when we see the hydrogen barcode pop up for a star, we can measure how much the barcode in the spectrometer is shifted compared to the stationary case, and we can use the amount of shift to determine how fast the star is moving relative to earth!

Example 3.3.1

A spaceship is fitted with a light beacon before blast-off. The light from this beacon is monochromatic, and when it is shone through the apparatus pictured below, the angle of deflection of the first order bright fringe is measured. The spaceship then blasts off, and after several years of accelerating through outer space, it is moving away from the Earth at a very high rate of speed, and the light from its beacon is shone through the apparatus again (which is still on Earth).



a. *Will the angle of deflection of the first-order bright fringe for the beacon coming from the moving ship be greater or less than the angle measured before blast-off? Explain.*

b. *Suppose deflection angle of the first order bright fringe changes by 10% as a result of the spaceship's motion (so it is either 90% or 110% of what it was before, depending upon your answer above). Find the speed of the spaceship. Assume that the deflection angle is small, so that sine of the angle changes by the same percentage as the angle itself when measured in radians.*

Solution

a. The ship is receding, so the source of the light is moving away from the receiver. This doppler-shifts the light to a lower frequency, which corresponds to a longer wavelength. The relationship between the angle of the first bright fringe and the wavelength is:

$$d\sin\theta = m\lambda \quad \Rightarrow \quad \sin\theta = \frac{\lambda}{d}$$

The separation of the slits doesn't change, so as the wavelength gets longer, the sine of the deflection angle gets bigger, which means the angle itself gets bigger.

b. From our answer above, the deflection angle has grown to 110% of what it was before blastoff. By our small-angle approximation, we can therefore say that the sine of the angle has grown by the same amount, which means that is how much the wavelength has shifted longer. The doppler shift formula (for light) gives a relationship between the sender's frequency and the receiver's frequency when the two are moving away from each other, and we can turn this into a relation between the wavelengths using Equation 2.2.10:

$$f_r = \sqrt{\frac{c-v}{c+v}}f_s \quad \Rightarrow \quad \frac{c}{\lambda_r} = \sqrt{\frac{c-v}{c+v}}\frac{c}{\lambda_s} \quad \Rightarrow \quad \frac{\lambda_r}{\lambda_s} = \sqrt{\frac{c+v}{c-v}}$$

We found that the received wavelength is 10% longer than the sent wavelength, which means that the ratio of these wavelengths is 1.1. Plugging this in allows us to solve for the velocity of the source (i.e. the ship):

$$1.1 = \sqrt{\frac{c+v}{c-v}} \quad \Rightarrow \quad 1.21\,(c-v) = (c+v) \quad \Rightarrow \quad v = \frac{0.21}{2.21}c = 2.9\times10^7\,\frac{m}{s}$$

3.4: Single-Slit Diffraction

Slits Are Not Actually Point Sources

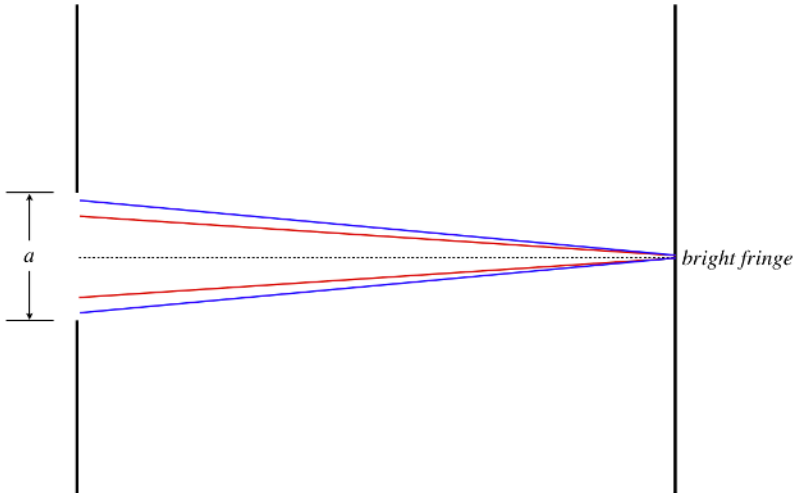
In our discussion of the double slit and diffraction grating, we made the assumption that the gaps that we call slits are so narrow that they can essentially be treated as point sources, making the analysis using Huygens's principle simple to do. But in reality we know that these gaps do not have infinitesimal width, and we need to consider what happens to the light when the approximation of "very thin gaps" breaks down. To do so, we will not consider a grating, or even a double-slit; we'll look at the effect that a single slit of a measurable gap size has on the light that passes through it. Notice that whatever this effect might be, when we extend the result to two or more slits, the effect will occur for every slit, superimposing itself on the multiple-slit interference pattern. But we are getting ahead of ourselves...

We already know that a plane wave passing through a single slit will diffract around the corners, so it will not simply leave a single bar of light on the screen the thickness of the gap – it will spread out. But what else can we say about it? Well, we know that without the aperture, all the Huygens wavelets would continue interfering perfectly to continue the plane wave, but when the portions of the plane wave outside the aperture are excluded, the effects of interference between wavelets is bound to change. We will analyze the effect by essentially following the procedure for many (infinite number) of thin slits that are infinitesimally close together.

Single Slit Interference Pattern

Let's call the gap width of the aperture a, and assume that this is much smaller than the distance to the screen, as in the figure below. We then consider what happens to the wavelets originating from every point within this region. When we look at how the screen opposite a single slit is illuminated, on the screen at the center line we observe a brightness maximum. You can think of such a situation as an infinite number of double-slits that are split by the center line with different slit separations. For every wavelet above the center line, there is a "twin" wavelet on the opposite side of the center line that travels the same distance to the screen (depicted by lines of the same color in the figure below), resulting in constructive interference. Of course, the fact that pairs constructively interfere with each other does not guarantee that the result of two constructively-interfering wavelets will not cancel with two other constructively-interfering wavelets (i.e. one pair creating a doubly-high peak, and the other a doubly-deep trough). In fact this can happen, but if it does, it's only for select wavelets – it can't persist for the entire aperture and leave darkness at the center line. Without going into the math, wavelets find it exceedingly difficult to find canceling partners at the center line, and on balance the interference is highly constructive – the center line is the brightest point in the entire interference pattern.

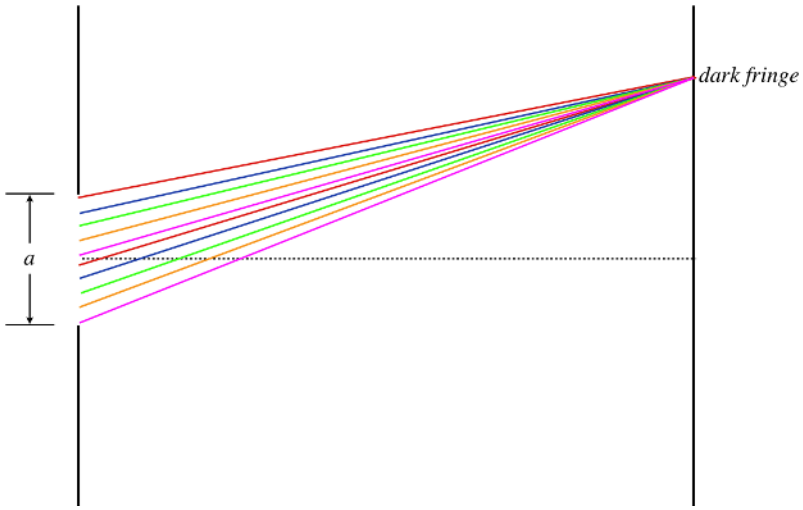
Figure 3.4.1 - Wavelet Pairs Constructively Interfere at the Center Line



Okay, so what about dark fringes – will we see these on the screen? Yes! To see why, we will once again find pairs of wavelets on both sides of the center line, which in this case travel different distances to the screen, differing by one-half wavelength for the first dark fringe. For this case, we pair-off the wavelet originating at the top of the slit with the wavelet originating just below the center line, and continue pairing them as we go down, until the wavelet at the bottom edge pairs with the wavelet originating just above the center line. This is depicted in the figure below with pairs of lines of the same color. The difference in distances for these pairs

will all be the same ($d\sin\theta$, where in this case d is actually $\frac{a}{2}$), and when this difference is one-half wavelength, they all cancel each other pairwise, leaving a dark fringe.

Figure 3.4.2 - Wavelet Pairs Destructively Interfering at the First Dark Fringe



Note that the same geometry holds below the center line as well. Setting the extra distance traveled by the twin wavelets equal to a have wavelength, we get the angle of the first dark fringe:

$$\text{first dark fringe:} \quad \frac{a}{2}\sin\theta = \frac{\lambda}{2} \quad \Rightarrow \quad \sin\theta = \pm\frac{\lambda}{a} \tag{3.4.1}$$

As we move upward on the screen, wavelets will again find their destructive twins and create dark additional dark fringes. It is a bit tricky for us to find the second dark fringe, however. The natural approach is to assume that the next dark fringe occurs when the pairs shown above travel distances that differ by three half-wavelengths, giving the result $\sin\theta = \pm 3\frac{\lambda}{a}$. But in fact this result incorrectly skips the second dark fringe, and goes to the third! To see why, we note that we can pair-off wavelets in a way other than across the center line. Specifically, we can think of this single slit as *two adjacent* single slits, one that has the center line as its lower edge, and one that has the center line as its upper edge. In this case, the wavelets pair-off within the top half, and then again within the bottom half separately. In this case, the only change in the math involves replacing $\frac{a}{2}$ with $\frac{a}{4}$, which means the second dark fringe satisfies:

$$\text{second dark fringe:} \quad \frac{a}{4}\sin\theta = \frac{\lambda}{2} \quad \Rightarrow \quad \sin\theta = \pm 2\frac{\lambda}{a} \tag{3.4.2}$$

We can similarly break the slit into three separate slits, which changes the separation of the starting wavelets to $\frac{a}{6}$, and increments the constant in the formula to 3. For the m^{th} dark fringe, we therefore have:

$$m^{th}\text{ dark fringe:} \quad \sin\theta = \pm m\frac{\lambda}{a} \tag{3.4.3}$$

The bright fringes only approximately follow the same spacing pattern, not exactly located halfway between the dark fringes, but using the pairwise approach doesn't tell us much about the intensity of those bright regions, for the same reason it didn't for the central bright fringe – constructive pairs will not be in phase with other constructive pairs. Significantly more math is required to deal with the intensity of the bright fringes.

Intensity

To compute the intensity of the interference pattern for a single slit, we treat every point in the slit as a source of an individual Huygens wavelet, and sum the contributions of all the waves coming out at an arbitrary angle. One way to think of this is to go back to the diffraction grating case, expressed in Equation 3.3.2. With the slit being completely open, however, the space between the slits (d) goes to zero, and the number of slits (n) goes to infinity. There is of course more to the calculation than this, and either the calculus or the "phasor method" described by many standard physics textbooks will reach the famous result below, and the reader is encouraged to have a look at these derivations. But these derivations do not contribute to the understanding of this

phenomenon, nor are they procedures essential to a wide range of future physics calculations, so we will omit them here, and jump to the end result.

If we define the amplitude of the total wave on the center line to be A_o due to the superposition of all the wavelets, then the amplitude of the wave at an angle θ off the center line is given by:

$$A(\theta) = \frac{\lambda A_o}{\pi a \sin\theta} \sin\left(\frac{\pi a \sin\theta}{\lambda}\right) \tag{3.4.4}$$

Yes, you are reading that right, there is a sine function of θ *within* another sine function. This is often written more succinctly by defining a new variable that is an implicit function of θ:

$$A(\alpha) = A_o \frac{\sin\alpha}{\alpha}, \qquad \alpha(\theta) \equiv \frac{\pi a}{\lambda}\sin\theta \tag{3.4.5}$$

This function comes up frequently enough in math and physics that it has even been given its own name – it is sometimes referred to as a *sinc function*.

Alert

It is important to understand that this expression compares the amplitude at various angles to the amplitude on the center line, equal (or approximately equal) distances from the slit. It does not provide a comparison of the amplitude of the light wave after passing through the slit to the amplitude of the plane wave before it enters the slit.

We know that the intensity of the wave at the center line is proportional to the square of the amplitude there, and that the intensity of the wave at an angle with the center line is proportional to the square of the amplitude there, and that the constants of proportionality are the same in both cases, so we immediately have a comparison of intensities:

$$I(\alpha) = I_o \left[\frac{\sin\alpha}{\alpha}\right]^2 \tag{3.4.6}$$

If the angle θ happens to be small, then α can be written as a function of distance y from the center line on the screen, as we did in Equation 3.2.5 for the double slit, giving:

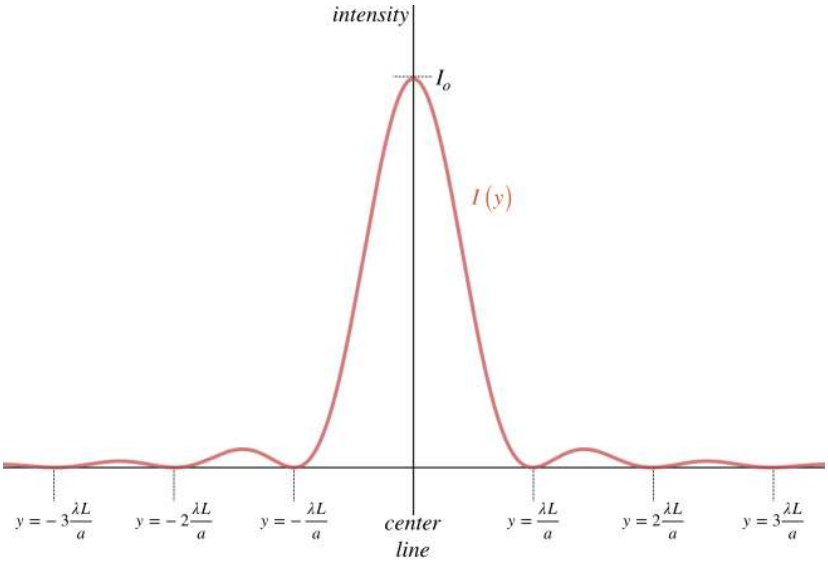
$$\alpha(y) \equiv \frac{\pi a y}{\lambda L}\ , \tag{3.4.7}$$

where, as before, L is the distance from the slit to the screen.

Perhaps you are concerned about the behavior of this function at the center line? After all, the value of the function α there does vanish, and this function appears in the denominator. But the numerator also vanishes at the center line, and L'Hôpital's Rule saves the day, giving the sinc function a value of 1 for $\alpha = 0$, resulting in the intensity equaling I_o, as it should.

A graph of the intensity of the full interference pattern looks like this:

Figure 3.4.3 - Single Slit Diffraction Intensity



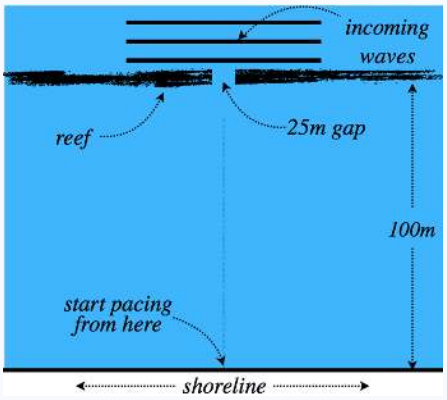
Let's point out a few of the more prominent features of this intensity pattern.

- The dark fringes are regularly spaced, in exactly the manner described by Equation 3.4.3 (note: $\sin\theta \approx \frac{y}{L}$).
- The central bright fringe has an intensity significantly greater than the other bright fringes, more that 20 times greater than the first order peak.
- Using calculus to find the placement of the non-central maxima reveals that they are not quite evenly-spaced – they do not fall halfway between the dark fringes.

We have assumed for simplicity the geometry of a long rectangular slit. If we were instead shining the light through a circular hole, this pattern would occur in every direction of two dimensions, resulting in concentric bright and dark circles, rather than fringes.

Example 3.4.1

You are on a sunny Hawaiian beach, trying to relax after a grueling quarter of Physics 9B. You would like to recline in your beach chair with your feet in the water, but don't want to get crushed by shore break while you snooze. About 100 meters off shore, you see an exposed reef that acts as a breakwater, but there is a gap in it, and waves (whose crests are parallel to the shore) are coming through that gap. While watching the waves, you see a surfer paddle out through the gap, and you use the perspective this event affords you to estimate that the gap is 25 meters wide (the diagram below is not to scale). You time a wave as it comes from the reef, estimating that it takes about 2 minutes for a wave to get to the shore from the gap, and the waves hit the shore roughly every 7 seconds. Starting from the point on the beach directly in line with the center of the gap, roughly how many paces (each pace being 1 meter in length) must you walk along the beach so that you can plant your beach chair and get the minimum wave intensity?



Solution

This is a problem in single-slit diffraction, where we are searching for the first "dark fringe" (place where destructive interference occurs). We can use Equation 3.4.3 for finding the angular deviation from the center line for a single slit, but it requires the wavelength of the wave as well as the slit gap. We have the latter, but we need to calculate the former. We can determine the wave speed and we are given the period, so:

$$\lambda = vT = \left(\frac{100m}{120s}\right)(7s) = 5.83m$$

Now we can plug this wavelength into Equation 3.4.3 to find the angle of the first dark fringe:

$$\sin\theta = \frac{\lambda}{a} \quad \Rightarrow \quad \theta = \sin^{-1}\left(\frac{5.83m}{25m}\right) = 13.5^o$$

The distance from the gap to the shoreline and the angle are known, so we can determine how far along the shore the dark fringe hits:

$$y = x\tan\theta = (100m)\tan 13.5^o = 24m$$

So you need to walk 24 paces.

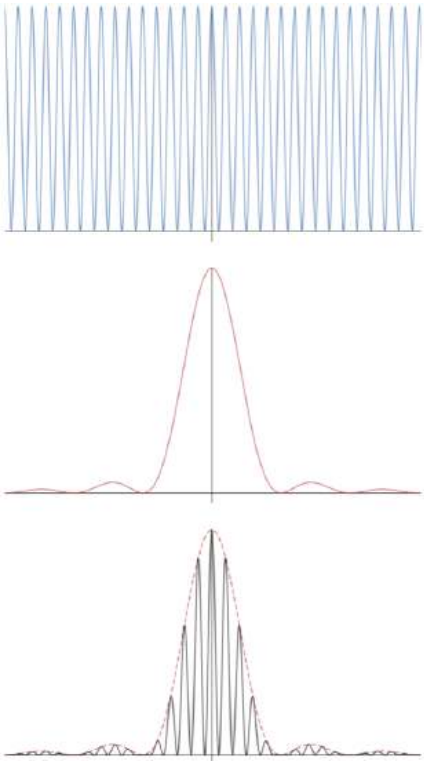
You might be tempted to use the "small angle" equation to solve this more directly, and in fact the angle is quite small. But we have defined our measurement limits in terms of paces, and using the small angle formula we end up with an answer of 23 paces, so while the approximation is very good (it is only off by less than 5%), even our rather coarse measurement scheme notices the difference. [Okay, so "notice" might be too strong a word, as the wave intensity one pace from a minimum and 23 paces from the maximum is not going to be significant.]

Including Gap Size for Double Slits

Our analysis of double slits assumed that the slits were *very* thin, creating point sources. But in the real world, these slits must have finite gap widths (if the widths get too small, too little light gets through to see anything!). So how can we incorporate our result for single slit interference into what we found for double-slit interference? The easiest way to see the answer is to think of the single slit effect as putting limits on the light that comes through. If the light that would reach the screen in the absence of the single slit is a plane wave, then these limits just consist of the single slit intensity pattern (square of the sinc function). If the light destined to reach the screen is instead a double-slit intensity pattern, then the effect of the single slit is to squeeze down the bright peaks (reduce the brightness) so that they conform to the "envelope" of the single slit pattern. That is, the whole intensity pattern of the double-slit becomes the "I_o" for the single slit pattern. Mathematically, this is equivalent to multiplying the intensity functions:

$$\left.\begin{aligned} I_{\text{double slit}} &= I_o\cos^2\left(\frac{\Delta\Phi}{2}\right) & \Delta\Phi &= \frac{2\pi}{\lambda}d\sin\theta \\ I_{\text{single slit}} &= I_o\left[\frac{\sin\alpha}{\alpha}\right]^2 & \alpha &= \frac{\pi a}{\lambda}\sin\theta \end{aligned}\right\} \quad \Rightarrow \quad I_{\text{both}} = I_o\cos^2\left(\frac{\Delta\Phi}{2}\right)\left[\frac{\sin\alpha}{\alpha}\right]^2 \qquad (3.4.8)$$

Figure 3.4.4 - Intensity Pattern for Double Slit with Finite Gap Widths



How does this actually appear to someone viewing it on the screen? The usual double-slit pattern is there, but the fringes are not all equally-bright. The center fringes are very bright, and it quickly tapers off. The single slit pattern is apparent in the brightness of the double-slit fringes. Notice, by the way, that we have assumed here that the slit separation is larger than the gap widths. This is apparent from the fact that the distance between dark fringes for the double slit is much smaller than it is for the single slit, and the separations are inversely-proportional to the slit separation d for the double slit, and inversely-proportional to the gap width a for the single slit.

Example 3.4.2

Light is shone through a double slit apparatus whose slit gaps are wide enough to also exhibit single slit interference. The slit spacing for this apparatus is 4 times as great as the gap sizes. How many bright fringes from the double slit pattern appear within the central maximum of the single slit pattern (i.e. between the first order dark fringes)?

Solution

We are given that $d = 4a$, so comparing the bright fringe equation for the double-slit (Equation 3.2.3) with the dark fringe equation of the single-slit (Equation 3.4.3), we see that the 4^{th}-order bright fringe of the former coincides with the 1^{st} dark fringe of the latter:

$$\left.\begin{array}{ll} \text{double-slit bright fringes:} & m_{\text{double-slit}}\,\lambda = d\sin\theta \\ \text{single-slit dark fringes:} & m_{\text{single-slit}}\,\lambda = a\sin\theta \\ \text{given:} & d = 4a \end{array}\right\} \quad \Rightarrow \quad m_{\text{double-slit}} = 4m_{\text{single-slit}}$$

This means that the 4^{th}-order double-slit bright fringe won't appear, as the destructive interference of the single slit will wipe it away. The central bright fringe and the three fringes (on each side) lie between these two endpoints. This makes a total 7 double-slit bright fringes within the central single-slit maximum.

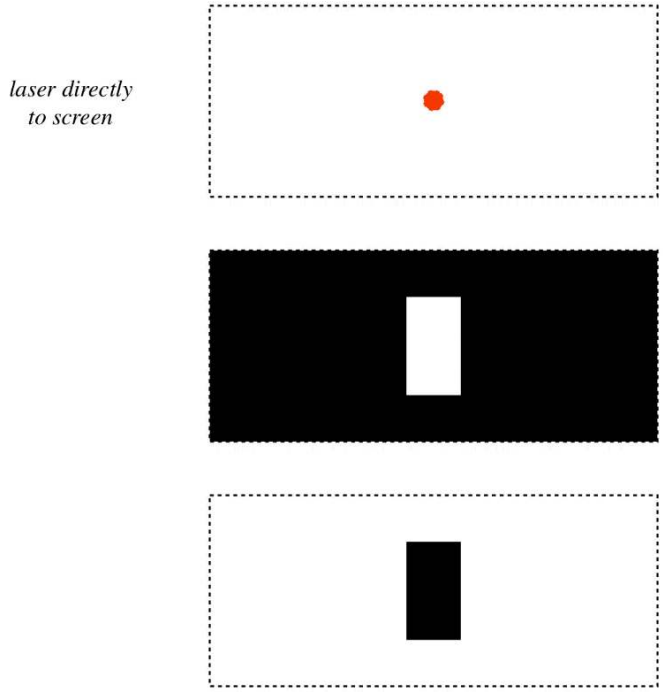
Diffraction for an "Inverse Slit"

We can use our knowledge of waves to determine the light pattern we will see when the incoming plane wave diffracts around a thin barrier. Imagine starting with a plane barrier, out of which we cut a tiny sliver. Described above is what we see if coherent light

is shone through the opening we have created in the barrier, but what if we shine the same light on *just the sliver*? That is, instead of only allowing light to pass through a thin space, we let the light pass everywhere *except* the thin space.

Imagine a tight laser beam in three different situations: First, it goes straight to the screen unimpeded. As with all laser beams, it spreads very little during its journey. Second, it encounters a thin slit that is a little bit smaller than the width of the beam. Naturally a single-slit diffraction pattern appears on the screen. And third, the beam encounters only a sliver that has the same dimensions as the single slit, so that the outer edges of the beam go past the edges of the sliver. Our question is what happens in this third case.

Figure 3.4.5 - Three Laser Beam Results



If the light from the second case was allowed to superpose with the light from the third case, it should be pretty clear that the result will be the first case. But for the second case, some light lands outside the beam's confines (thanks to diffraction), which means that for the superposition to occur, the third case must also send light to those outer regions with exactly the same amplitudes as the slit, though the light from the sliver must be π radians out of phase with the light from the slit. But if the sliver is by itself, the light it sends outside the beam region doesn't cancel with anything, which means it shows up on the screen. The end result is that the interference pattern outside the beam region must be the same for the sliver as it was for the slit.

What about the central bright fringe? For a single slit, the central maximum is not as bright as the unimpeded beam (because some of the light energy is diverted by diffraction). For the superposition to apply, this means that the region directly behind the sliver must also be illuminated. The relative brightness of the central maximum with the outer fringes may be different for the slit and the sliver, but the fringe spacings are the same in both cases, giving essentially the same diffraction patterns for both cases. This phenomenon is known as *Babinet's principle*.

3.5: Thin Film Interference

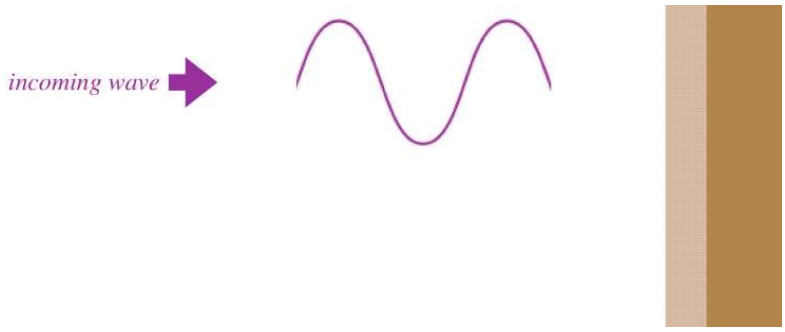
The Basic Idea

We have already seen three physical systems that result in interference patterns. While they all result in different patterns, they all function in pretty much the same way: A single wave is broken into multiple in-phase wavelets à la Huygens's principle, and these wavelets interfere with each other after traveling different distances to a position on a screen. Here we will see another interference phenomenon, and this one is also based on two waves traveling different distances, but this comes about due to reflection rather than diffraction.

An important element to this is that waves that strike a surface of a new medium partially reflect and partially transmit. This allows for the possibility that a single incoming wave can result in *two* waves being reflected off a thin, transparent film. Part of the wave reflects off the front surface of the film, and the other part off the rear surface of the film. These two reflected waves both come away from the film in the same direction, but they travel different distances in the process, because one of them traverses the thickness of the film twice, while the other does not. This difference can lead to destructive interference, meaning that *no* light is reflected!

This only scratches the surface of this phenomenon, however, because there are two other very important things going on that we have to take into account. The first of these is that waves which reflect off new media can possibly experience a phase shift of π if they reflect off a medium in which the wave moves more slowly (see Section 1.5 to review this phenomenon). The wave may make a phase shift at the front surface, the rear surface, both, or neither. As the phase difference between the two waves is the only factor that determines whether or not there is destructive interference, knowing whether or not each reflected wave has changed its phase by π is critical.

Figure 3.5.1 – Thin Film Destructive Interference



In the figure above, part of the incoming wave reflects off the front surface of the transparent film (the red wave), and the rest of it is transmitted into the film, after which it reflects off the rear surface (the blue wave). [*Note that the different colors of these waves are used to distinguish them from each other, not to represent red and blue light – these waves have the same frequencies when they interfere.*] These two waves come out together and interfere, but the red wave has a "head start" both in displacement (the thickness of the film), and in time – it is already propagating left while the incoming wave is still moving right, on its way to the rear surface. Of course, the film thickness can be adjusted to whatever we like, and in this particular case, it is one quarter of a wavelength of the light, which results in the destructive interference, as we can see in multiple ways.

To determine the interference of these two waves, we have to compute their total phase difference $\Delta\Phi$ at the point when they superpose. So let's consider the position and time when they first encounter each other – at the front surface after the blue wave has reflected back to that point...

- That position is the starting point of the red wave, so $x_{red} = 0$. For the blue wave it is a position one quarter wavelength from its origin, so $x_{blue} = \frac{\lambda}{4}$, giving:

$$\Delta x = x_{blue} - x_{red} = \frac{\lambda}{4}.$$

- At the moment when they meet, the red wave has been propagating for one-half period (it propagates for the period of time that the incoming wave travels a quarter wavelength to the right, plus the time that the blue wave propagates a quarter wavelength

back to the left), so $t_{red} = \frac{T}{2}$. The blue wave only travels a quarter wavelength by the time the two waves superpose, so it has been propagating for a quarter of a period: $t_{blue} = \frac{T}{4}$. This gives us a difference in time-of-propagation of:

$$\Delta t = t_{blue} - t_{red} = -\frac{T}{4}.$$

- Both waves experience a phase shift upon reflection, and come from a common incoming wave where they were in phase, so there is no difference in their phase constants:

$$\Delta\phi = \phi_{blue} - \phi_{red} = 0.$$

Putting all this together gives us the phase difference of the two waves when they rejoin (note that by choosing values of x to be positive measured to the left, the wave is moving in the positive direction, which means that the position and time parts of the phase must have opposite signs):

$$\Delta\Phi = \frac{2\pi}{\lambda}\Delta x - \frac{2\pi}{T}\Delta t + \Delta\phi = \frac{2\pi}{\lambda}\left(\frac{\lambda}{4}\right) - \frac{2\pi}{T}\left(-\frac{T}{4}\right) + 0 = \pi \tag{3.5.1}$$

With the two waves out of phase by π, they interfere destructively:

$$I = I_o \cos^2\left(\frac{\Delta\Phi}{2}\right) = I_o \cos^2\left(\frac{\pi}{2}\right) = 0 \tag{3.5.2}$$

We can just as easily ignore the time element by choosing the zero time to be when the incoming wave first strikes the front surface. In this case, both waves start at the same moment (making $\Delta t = 0$), but the wave that strikes the rear surface travels an extra *half* wavelength, since it has to make a round-trip across the film. Of course, the same answer results, and this is a somewhat simpler way to view the phase difference.

Note that a film thickness is not the only way that destructive interference can occur. If the thickness was instead three-quarters of a wavelength, then the round-trip distance for the transmitted wave is 1.5 wavelengths, and it once again emerges from the film out of phase with the reflected wave by π radians. We will summarize the effect of film thickness shortly, but there are a couple other loose ends we need to tie up first.

The Effect of Phase Shifts

In the example above, we mentioned the phase shifts that occurred at the boundaries of the media, but they didn't seem to take part in the calculation. This is because the same phase shift occurred at both boundaries. Suppose there was no third material involved (depicted in dark brown in the figure above), and that the film was by itself, surrounded on both sides with the exterior medium (presumably air, but no necessarily). In this case, there would be a phase shift at the reflection with the front surface, but no phase shift at the rear surface. Then the blue wave in the figure would not emerge upside-down, and it would come out *in phase* with the red wave. Mathematically this is not hard to see, as the space (and time) parts of the phase difference calculation are unchanged, but now we have the blue wave starting with a different phase than the red wave: $\Delta\phi = \pi$. When we put this into the calculation of the phase difference, we find that the two waves emerge *in phase*. For the waves to emerge out of phase, the film thickness would need to be different. Specifically, a thickness of one-half wavelength (or some integer number of half wavelengths) will make the distance traveled a full wavelength, which would normally make them in phase, but one of the waves is phase-shifted by π, putting them out of phase by that amount.

Example 3.5.1

A thin film of glass is in flush contact with a thin film of transparent plastic. Light travels faster through the air than through the glass, and faster through the glass than through the plastic. Monochromatic light is shone on both sides of this combination (the same frequency of light on both sides), and there is a negligible amount light reflected from either side. If the two films are now separated slightly to allow for a small air gap between them, and we repeat the process with the same light, what will we see in the way of light reflections from the two sides?

Solution

The only change that occurs with the separation is that the second reflection in both thin films is now off a surface in contact with air. This means there will be no phase shift at that reflection, since air has a lower index of refraction than either film. Before the separation, the second reflection within the plastic was off a faster medium (glass), so there is no change to the

phase shift for the plastic, and the same interference as before (destructive) will result. But the second reflection for the glass film was previously off a slower medium (plastic), so changing that reflection so that it is off air will make it go from a π radian phase shift to no phase shift. With the film thickness the same as before, this means that light that previously emerged from the glass π out of phase with the first reflection is now in phase with it, so light will be seen reflected off the glass film.

Light in a Medium

There is one element of this phenomenon that we have not yet accounted for. Clearly the film thickness has to be just right for the timing of the two reflected waves to come out exactly π out of phase. But when it comes to timing, there is another consideration – the wave moving through the film is moving through a different medium than the wave that reflects off the first surface, which means the two waves are *moving at a different speeds*. Clearly this will have to play a role in the timing that leads to the interference effect. Let's have a look at the effects of media on the speed of light.

As we stated in Section 3.1, light will travel through a vacuum, and its maximum speed occurs through that (non)medium. When light propagates through other medium that is transparent to it, it slows down. Going into what physical attributes of the media go into slowing down the light and by how much is beyond the scope of this course, but generally the effect is boiled down to a single constant called the *index of refraction*. This constant (n) is dimensionless, and is a number greater than 1 which provides the speed of light through a medium in terms of the speed through the vacuum (c) according to:

$$v = \frac{c}{n} \tag{3.5.3}$$

When we discussed what happens when a wave passes from one medium to another, we concluded that the frequency remains the same, and the wavelength changes along with the velocity. This means that if a light wave is traveling from a medium with index of refraction n_1 to a new medium with index of refraction n_2, then the unchanging frequency gives the following relationship between the two wavelengths:

$$f_1 = f_2 \quad \Rightarrow \quad \frac{v_1}{\lambda_1} = \frac{v_2}{\lambda_2} \quad \Rightarrow \quad \frac{c}{n_1\lambda_1} = \frac{c}{n_2\lambda_2} \quad \Rightarrow \quad n_1\lambda_1 = n_2\lambda_2 \tag{3.5.4}$$

Since the index of refraction is larger where the light travels slower, then light passing into a slower/faster medium will exhibit a reduction/increase in wavelength.

All we need to do to complete the analysis above is use the proper wavelength for the light within the film. That is, with two phase shifts, the film still needs to have a thickness of one quarter (or three quarters, or five quarters, etc.) of the wavelength of the light, but that wavelength must be *as measured within the film*. Putting it all together, destructive interference occurs when light is reflected off a thin film when the light reflected off the front and rear faces emerge π radians out of phase, which occurs under the following circumstances (the Greek letter τ is used for film thickness, to avoid confusion with the time variable):

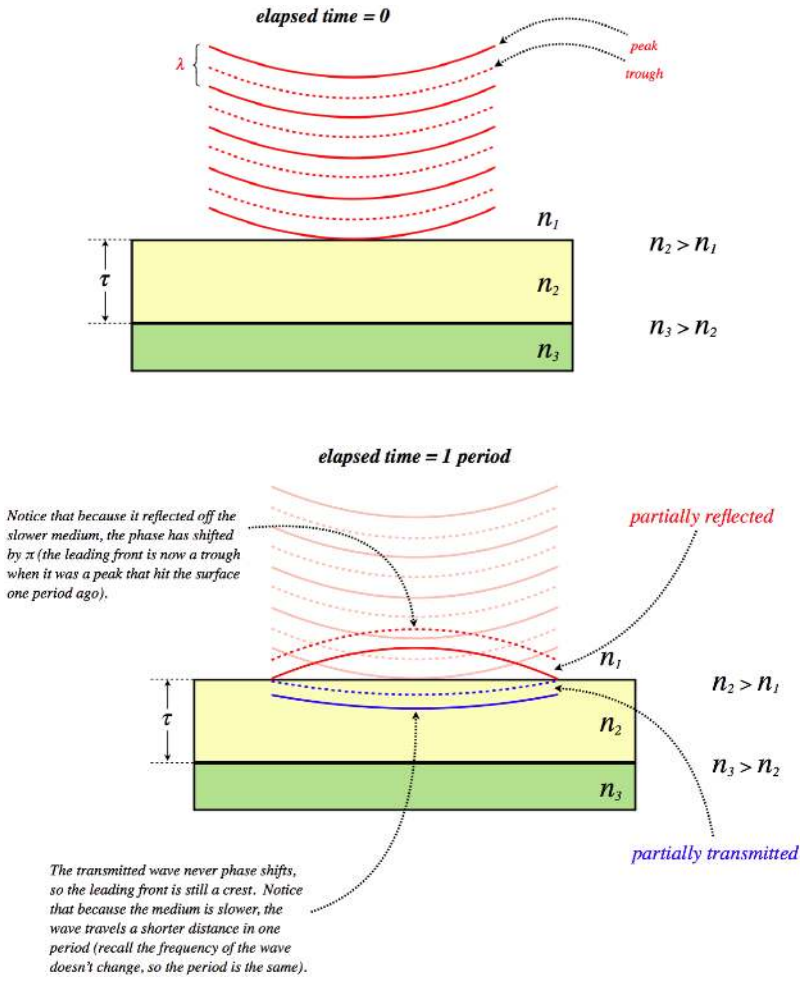
$$\begin{aligned} &\text{phase shifts at only one surface:} && 2\tau = m\lambda_{in\ film} \\ &\text{phase shifts at both or neither surface:} && 2\tau = \left(m - \tfrac{1}{2}\right)\lambda_{in\ film} \end{aligned} \qquad m = 1,\ 2,\ \ldots \tag{3.5.5}$$

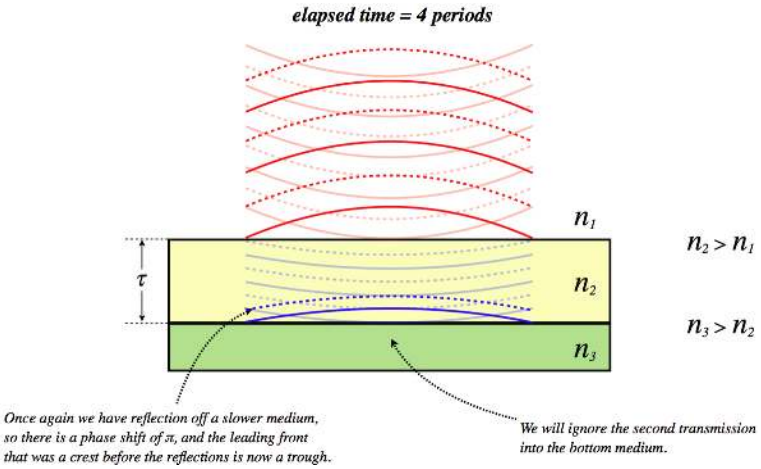
You are encouraged to convince yourself that these formulas do result in a phase difference of π.

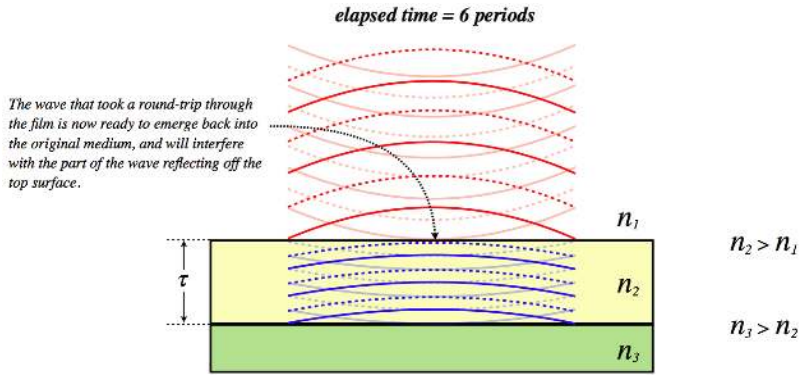
Putting It All Together

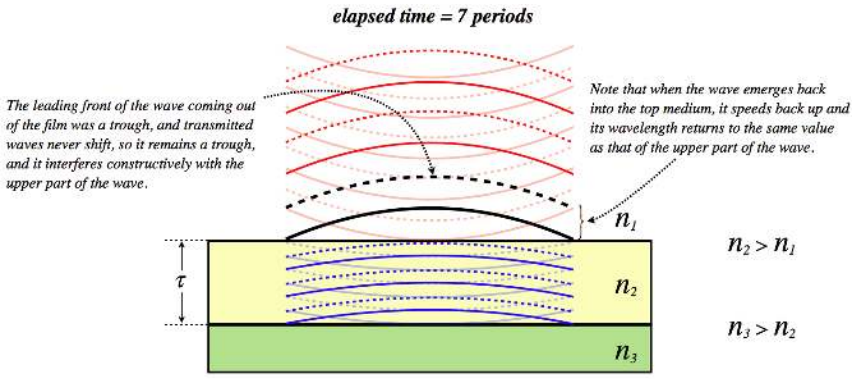
Let's put together a series of diagrams that reveal step-by-step what happens in thin film interference. There are several circumstances possible, but we will choose a film that has a thickness of three wavelengths of the light (as measured within that film), and assume the light is coming from a medium with a lower index of refraction, while behind the film is a medium with a higher index of refraction.

Figure 3.5.2 – Step-By-Step Thin Film Interference


elapsed time = 0
λ
peak
trough
n_1
$n_2 > n_1$
τ
n_2
$n_3 > n_2$
n_3
elapsed time = 1 period
Notice that because it reflected off the slower medium, the phase has shifted by π (the leading front is now a trough when it was a peak that hit the surface one period ago).
partially reflected
n_1
$n_2 > n_1$
τ
n_2
$n_3 > n_2$
n_3
partially transmitted
The transmitted wave never phase shifts, so the leading front is still a crest. Notice that because the medium is slower, the wave travels a shorter distance in one period (recall the frequency of the wave doesn't change, so the period is the same).
elapsed time = 3 periods
Clearly if we keep waiting an integer number of periods, the picture up here will not change.
n_1
$n_2 > n_1$
τ
n_2
$n_3 > n_2$
n_3
A second reflection occurs as the transmitted wave reaches the bottom medium.







$$\Phi_1 = \Phi_{incoming} + \underbrace{\pi}_{\substack{\text{reflection off} \\ \text{slower medium} \\ [n_2 > n_1]}}$$

$$\Phi_2 = \Phi_{incoming} + 2\pi\left(\frac{\Delta x}{\lambda}\right) + \underbrace{\pi}_{\substack{\text{reflection off} \\ \text{slower medium} \\ [n_3 > n_2]}} = \Phi_{incoming} + 2\pi\left(\frac{2\tau}{\lambda}\right) + \pi = \Phi_{incoming} + 2\pi\left(\frac{6\lambda}{\lambda}\right) + \pi = \Phi_{incoming} + 13\pi$$

$$\Delta\Phi = 13\pi - \pi = 12\pi \Rightarrow \text{constructive interference}$$

Applications

This phenomenon can be observed and exploited in several ways. Let's start with observations...

Sunlight that reflects off a thin film of oil floating on a puddle of water will exhibit thin film interference in an interesting manner that comes about because the light is comprised of many wavelengths, and the film of oil is not uniform in thickness. The light will strike one part of the film, where the thickness happens to cause destructive interference for light of a certain wavelength. While it is not totally destructive for nearby wavelengths, they are out of phase by a number very close to π, which means that they are

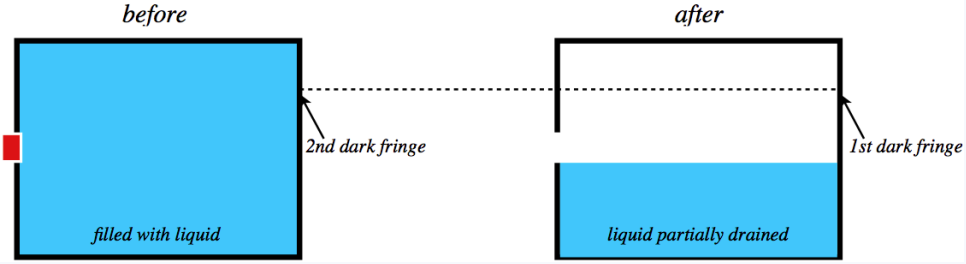
barely seen. So if the thickness happens to cause destructive interference for light near the red end of the spectrum, the reflected light will look more blue. At other points in the film, the thickness may cause the blue light to be canceled. The result is a reflection of a rainbow of colors. What is more, the variation in color tracks the thickness, so the observed rainbow swirls look like the grade lines on a topographical map, with each line of constant color indicating a different film thickness. These thickness variation rainbows can also be seen other thin films, such as soap bubbles.

A practical application of this effect are anti-glare films. When light strikes a thin film with air on both sides, if the film allows for no light to be reflected (i.e. all the reflected light destructively interferes), then conservation of energy requires that *all* of the light passes through the film to the other side. One place where one might want as much light as possible to pass through is a camera lens. Of course, cameras generally take photographs of objects illuminated by the entire spectrum of visible light, and it is impossible for thin films to create destructive interference for reflected light of all wavelengths at once. So generally the film chosen works for the middle of the spectrum (green light), which means that it doesn't work well for the ends of the spectrum (red and violet). With essentially only red and violet light able to be reflected by the film on a camera lens, the lens takes on a dark purple appearance.

The following example, while not a thin film problem, nevertheless incorporates the effect of a medium on interference patterns.

Example 3.5.2

A chamber is filled with an unknown liquid, held within by a small rectangular hole in its side, as shown in the diagram below. The plug does not fit perfectly, and it allows light to pass through its top and bottom edges. Coherent monochromatic light is shone through these tiny gaps, and an interference pattern appears on the opposite wall of the chamber. The plug is then removed from the rectangular hole, and some of the liquid drains out. After it does so, a new interference pattern emerges above the level of the liquid (assume this occurs with negligible effect from the light reflected off the surface of the liquid). It is noted that the position of the first dark fringe of this new pattern exactly coincides with the position of the second dark fringe that appeared when the plug was still in place. Find the index of refraction of the unknown liquid. The index of refraction of air is approximately the same as a vacuum: $n_{air} = 1.0$.



Solution

When the plug is in, the small slits at the outer edges that allow light to come in function as a double slit. The separation of these slits is the width of the hole, which we will call d. *The angle at which the dark fringes occur for the double slit is given by the usual double slit relation, Equation 3.2.3. We see the second fringe, which corresponds to* $m = 1$, *and the light is passing through the liquid, so the wavelength in this equation is the wavelength within the liquid:*

$$d \sin\theta = \left(m + \frac{1}{2}\right) \lambda_{liquid} = \frac{3}{2}\lambda_{liquid}$$

The source of light remains unchanged, so whether the light is passing through the liquid or (after the chamber drains) the air, the frequency is the same. This means that the wavelength of the light through the liquid is related to the wavelength through the air according to:

$$n_{air}\lambda_{air} = n_{liquid}\lambda_{liquid} \quad \Rightarrow \quad \lambda_{liquid} = \frac{\lambda_{air}}{n},$$

where n *is the index of refraction we are looking for.*

When the plug comes out, the hole now becomes a single slit with the light traveling through air, and the gap size of this single slit (which we typically denote as a*) is exactly equal to the double slit separation we used above:* d. *The first dark*

fringe for a single slit pattern corresponds to $m = 1$ in the formula, so putting this (and $a = d$) into the formula gives us a relation between the wavelength of the light in air, the gap width, and the deflection angle:

$$a \sin\theta = m\lambda_{air} \quad \Rightarrow \quad d \sin\theta = \lambda_{air}$$

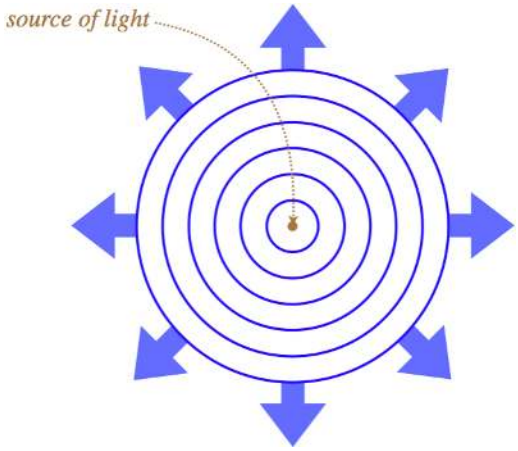
Putting these three equations together, we find that all of the unknowns except for the index of refraction of the liquid cancel out, leaving simply $n = 1.5$.

3.6: Reflection, Refraction, and Dispersion

Rays

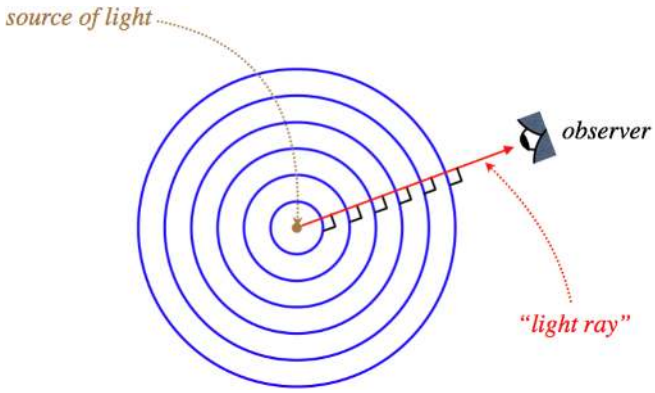
As we consider more phenomena associated with light, one of our primary concerns will be the direction that light is traveling. We already know that light, like any wave, travels in a direction perpendicular to its planes of constant phase:

Figure 3.6.1 – Light Waves Travel in Several Directions at Once



So in our wave view of light, we say that the light wave is traveling in many directions at once, but now we are going to change our perspective to that of an observer and a source. When we do that, we narrow down all the possible directions of the light wave motion to a single line, which we call a *light ray*. This is a directed line that originates at the source of light, and ends at the observer of the light:

Figure 3.6.2 – Source and Observer Define a Ray



Alert

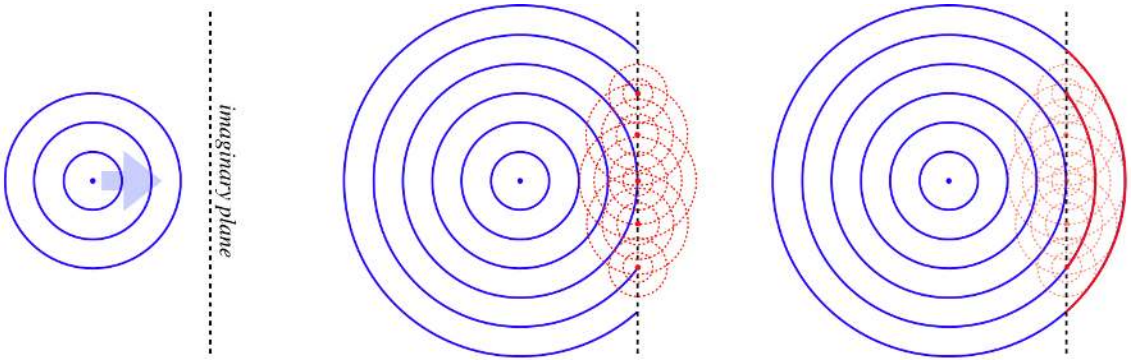
When most people encounter the idea of a light ray for the first time, what they think of is a thinly-confined laser beam. This is ***not*** *what is meant here! The ray has no physical meaning in terms of the confinement of light – we just use it as a simple geometrical device to link a source to an observer. Always keep in mind that the actual physical manifestation of the light is a wave that is usually traveling in many directions at once! Our use of rays will become so ubiquitous that this will be easy to forget.*

Reflection

Consider a point source of light that sends out a spherical wave toward an imaginary flat plane, as in the left diagram below. When the wave reaches this plane, then according to Huygens's principle, we can look at every point on the plane and treat it as a point source for an individual wavelet (center diagram below). These wavelets are not in phase, because they are all travel different

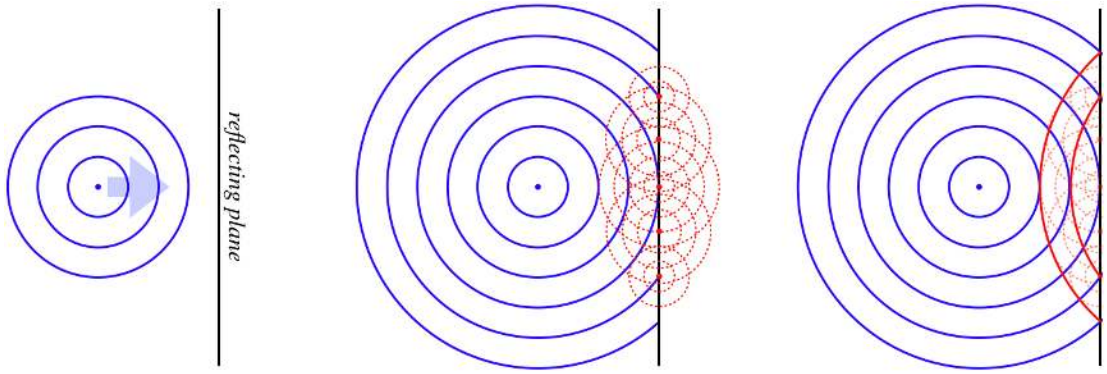
distances from the source to the plane, and when they are superposed, we know the result is what we see, which is a continued spherical wave (right diagram below).

Figure 3.6.3 – Spherical Wave Passes Through Imaginary Plane



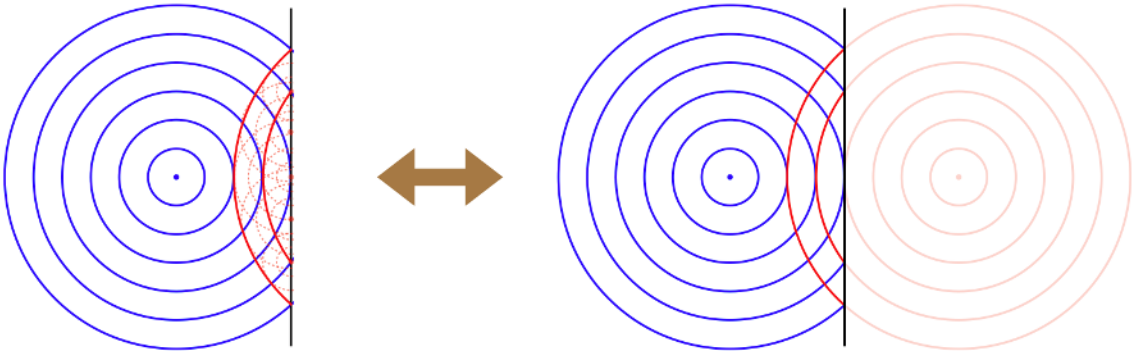
Now suppose the plane is not imaginary, but instead reflects the wave. Every point on this plane becomes a source of a wavelet, but this time, the wave created by these wavelets is going in the opposite direction. The wavelets have the same relative phases as in the previous case, and they are completely symmetric, so they superpose to give the same total wave as before, with the exception that it is a mirror image of the case of the imaginary plane:

Figure 3.6.4 – Spherical Wave Reflects Off Plane



Thanks to the symmetry of the situation, it's not difficult to see that the reflected wave is identical to a spherical wave that has originated from a point on the opposite side of the reflecting plane, exactly the same distance from the plane as the source, and along the line that runs through the source perpendicular to the surface:

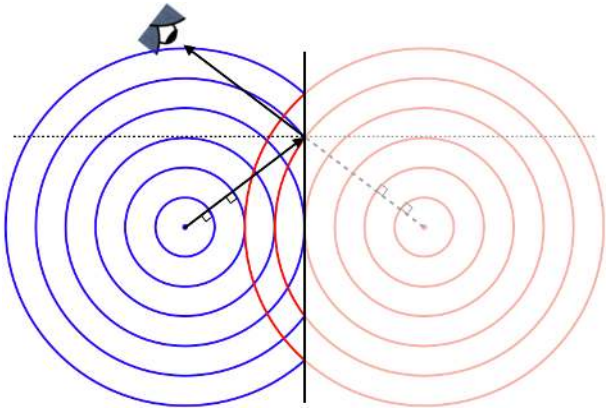
Figure 3.6.5 – Image of a Reflected Wave



Of course, there isn't *actually* a point light source on the other side of the reflecting plane, it's just that someone looking at the reflected light – no matter where they look from – will see the wave originating from the direction of that point. We call such a point an *image* of the original source of the light.

Now let's put this result in terms of light rays. To do this, we need a source and an observer, and this case, we will require also that a reflection has taken place. Once again drawing the rays perpendicular to the wave fronts, we get:

Figure 3.6.6 – Law of Reflection



It's clear from the symmetry of the situation that the angle the ray makes with the perpendicular (the horizontal dotted line) to the reflecting plane as it approaches, is the same as the angle it makes after it is reflected. This gives us the *law of reflection*, which states that the incoming angle (*angle of incidence*) equals the outgoing angle (*angle of reflection*):

$$\theta_i = \theta_r \tag{3.6.1}$$

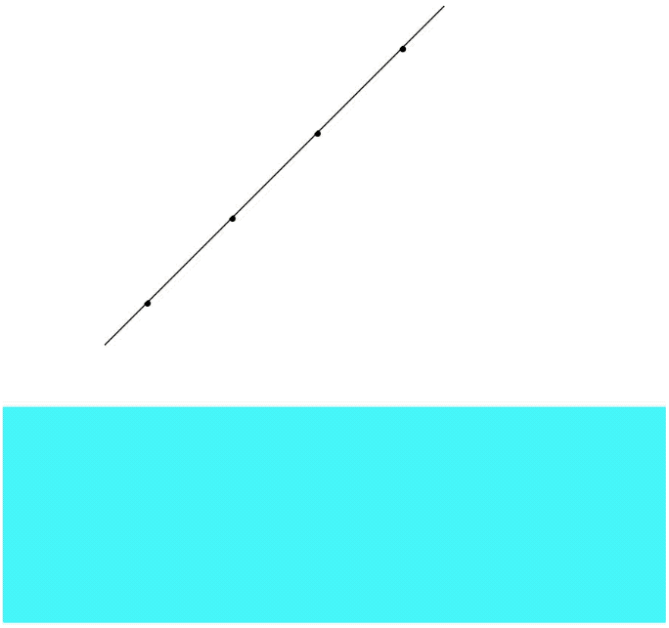
The beauty of introducing rays is that from this point on, we can discuss sources and observers without a complicated reference to the spherical waves and Huygens's principle – we can just use the law of reflection and pure geometry.

Refraction

We saw that light waves have the capability of changing the direction of the rays associated with it through diffraction. We now consider another way that such a direction change can occur. This process, called *refraction*, comes about when a wave moves into a new medium. To get to the essence of this phenomenon from Huygens's principle, we don't have a symmetry trick like we did for reflection, so rather than use a point source of the light, we can look at the effect that changing the medium has on a plane wave.

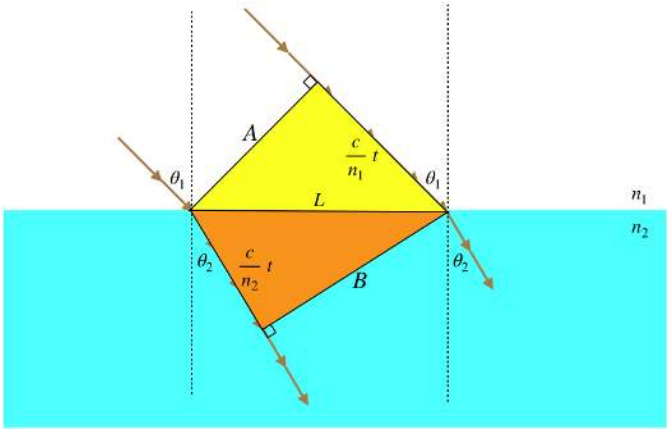
We saw in Figure 3.1.2 how a plane wave propagates according to Huygens's Principle. We can't sketch every one wavelets emerging from the infinite number of points on the wavefront, but we can sketch a few representative wavelets, and if those wavelets have propagated for equal periods of time, then a line tangent to all the wavelets will represent the next wavefront. It's clear that following this procedure for a plane wave will continue the plane wave in the same direction. But now let's imagine that such a plane wave approaches a new medium from an angle, as shown in the figure below. As each point on the wave front comes in contact with the new medium, it becomes a source for a new Huygens wavelet *within the medium*. These wavelets will travel at a different rate than they traveled in the previous medium (in the figure, the light wave is slowing down in the new medium). This means that the distance the wave in medium #1 travels is farther than it travels in medium #2 during the same time. The effect is a bending of the direction of the plane wave in medium #2 relative to medium #1.

Figure 3.6.7 – Huygens's Principle Refracts a Plane Wave



The amount that the direction of the light ray changes when the wave enters a new medium depends upon how much the wave slows down or speeds up upon changing media. In other words, it depends upon the indices of refraction of the two media. We can actually calculate this effect by freezing the figure above and looking at some triangles:

Figure 3.6.8 – The Geometry of Refraction



We are looking at what happens to a wavefront when it passes from position A to position B. The left side of the wave front is traveling within medium #2, during the same time period that the right side is traveling through medium #1. The rays are by definition perpendicular to the wavefronts, and we have defined the angles the rays make with the perpendicular in each medium as θ_1 and θ_2. Before we do any of the math at all, we immediately note:

Light passing from a faster medium into a slower medium bends toward the perpendicular, and light passing from a slower medium to a faster medium bends away from the perpendicular.

While the second of these conclusions is not expressed in our figure, it's not hard to see that it must be true, if we just imagine the wavefronts in the figure moving up to the left from medium #2 to medium #1.

Now for the math. We have two right triangles (yellow and orange) with a common hypotenuse of length we have called L. The distance between wavefronts in the upper medium is the speed of the wave there ($\frac{c}{n_1}$) multiplied by the time spent propagating, while the distance measured within the lower medium is calculated the same way, with a different speed ($\frac{c}{n_2}$). The angle θ_1 (shown on the right side of the diagram) is clearly the complement of the acute angle on the right-hand-side of the yellow triangle, which makes it equal to the acute angle on the left-hand-side of the yellow triangle. We therefore have:

$$\sin\theta_1 = \frac{\left(\frac{c}{n_1}\right)t}{L} \tag{3.6.2}$$

Similarly we find for θ_2:

$$\sin\theta_2 = \frac{\left(\frac{c}{n_2}\right)t}{L} \tag{3.6.3}$$

Dividing these two equations results in c and L dropping out, leaving:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{3.6.4}$$

This relationship between the rays of a light wave which changes media is called *the law of refraction*, or *Snell's law*. While this works in either direction of light propagation, for reasons that will be clear next, it is generally accepted that the "1" subscript applies to the medium where the light is coming from, and the "2" subscript the medium that the light is going into.

Total Internal Reflection

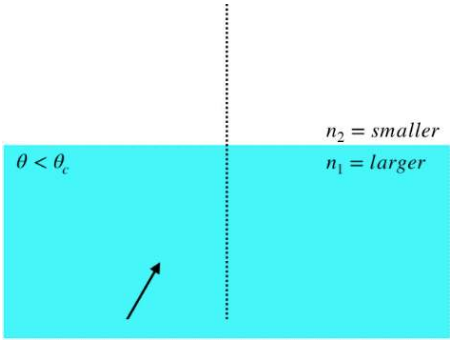
It was noted above that light which passes from a slower medium to a faster one bends away from the perpendicular. What happens then if the incoming angle is made larger and larger (obviously it can't be more than 90^o)? For example, suppose we have $n_1 = 2.0$, $\theta_1 = 45^o$, and $n_2 = 1.0$. Plugging these values into Snell's law gives:

$$\sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1 = 2.0 \cdot \sin 45^o = 1.4 \tag{3.6.5}$$

The sine function can never exceed 1, so there is no solution to this. This means that the *light incident at this angle cannot be transmitted into the new medium*. Every time light strikes a new medium some can be transmitted, and some reflected, so this result tells us that all of it must be reflected back into the medium in which it started. This phenomenon is called *total internal reflection*. The angle at which all of this first blows up is the one where the outgoing angle equals 90^o (the outgoing light refracts parallel to the surface between the two media). This angle is called the *critical angle*, and is computed by choosing the outgoing angle to be 90^o:

$$n_1 \sin\theta_c = n_2 \sin 90^o \quad \Rightarrow \quad \theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) \tag{3.6.6}$$

Figure 3.6.9 – Partial and Total Internal Reflections By Incident Angle



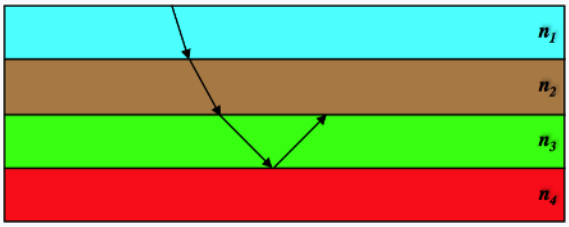
Note that there is at least partial reflection (obeying the law of reflection) every time the light hits the surface, but all of the light along that ray is only reflected when the ray's angle exceeds the critical angle.

Alert

Note that when light is coming from one medium to another, unless that light is a plane wave, it will be moving in many directions at once. Only the portions of the light wave with rays that equal or exceed the critical angle are not transmitted into the new medium. So the word "total" in "total internal reflection" to express the fraction of light at a specific angle that is reflected back, not necessarily the fraction of all the light that is reflected back.

Example 3.6.1

The diagram to the right shows the path of a ray of monochromatic light as it hits the surfaces between four different media (only the primary ray is considered – partial reflections are ignored). Order the four media according to the magnitudes of their indices of refraction.



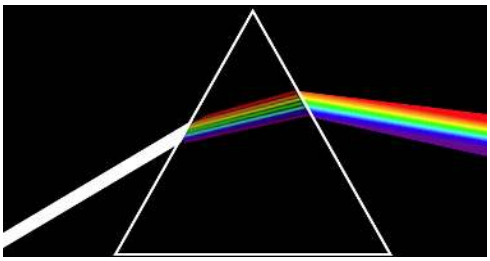
Solution

We know from Snell's Law that when light passes from a higher index to a lower one, it bends away from the perpendicular, so we immediately have $n_1 > n_2 > n_3$. For the ray to reflect back from the fourth medium, it has to be a total internal reflection (we are only considering primary rays, so this is not a partial reflection), which can only occur when light is going from a higher index of refraction to a lower one, so $n_3 > n_4$.

Dispersion

What determines the index of refraction for a medium is a very complicated problem in E&M, but there is one easily-observable fact: The amount that a ray bends as it enters a new medium is dependent upon the light's frequency. Specifically, the higher the frequency of the light, the more it bends – it essentially experiences a higher index of refraction when its frequency is higher. This phenomenon is most evident when white light is shone through a refracting object. The most iconic example of this is white light through a prism.

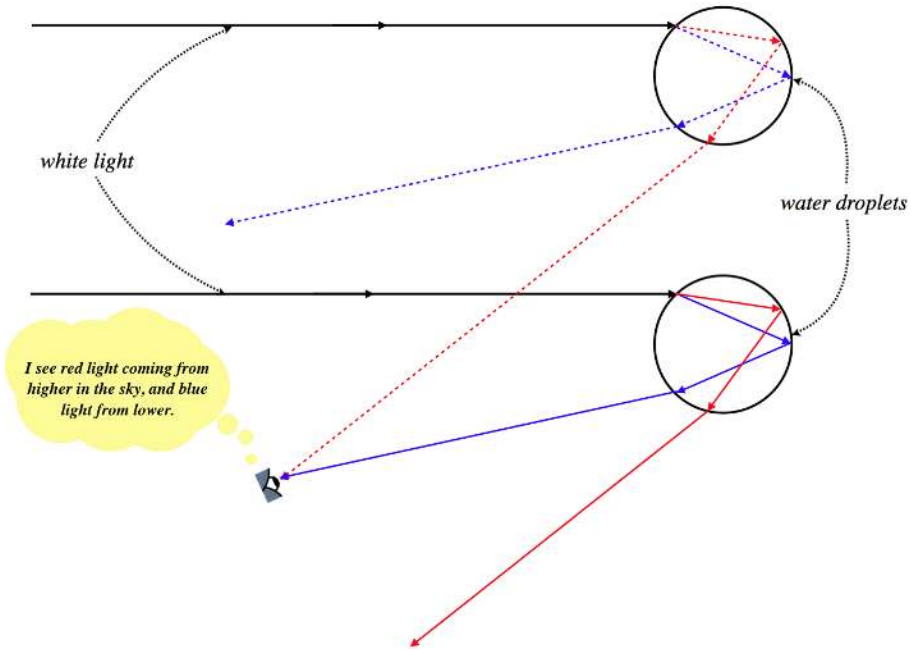
Figure 3.6.10 – Dispersion Through a Prism



The emergence of the fully-separated spectrum of colors from a prism is reminiscent of a rainbow, and in fact rainbows are also a result of dispersion. Unlike the prism depicted above, however, internal reflection is an integral part of the rainbow effect (and in fact prisms can also feature internal reflection).

A droplet of water suspended in the atmosphere is a refracting sphere. White light that enters near the top of the droplet gets dispersed inside the droplet, reflects, and then gets dispersed as it exits the droplet, sending rays of different-colored light in different directions. The diagram below shows this effect for rays of red and blue light for two droplets.

Figure 3.6.11 – Rainbows



A few things to note here:

- Notice that the sun always needs to be behind the observer in order to witness a rainbow. That's why it seems to move as you move, and why reaching the "end of the rainbow" is impossible (unless you can catch a leprechaun).
- The reason it is shaped like a bow is that the sun is nearly a point source, so the geometry is symmetric around the line joining the sun and the observer. If you create a "human-made rainbow" with a light and some mist, you can get close to an entire circle (minus whatever light your body blocks out).
- The secondary rainbow above the primary one comes from the light that enters the *bottom* of the droplets, and has *two* internal reflections. This reversed direction of the light bouncing around inside the droplets results in the colors being reversed (the violet is at the top and the red at the bottom).

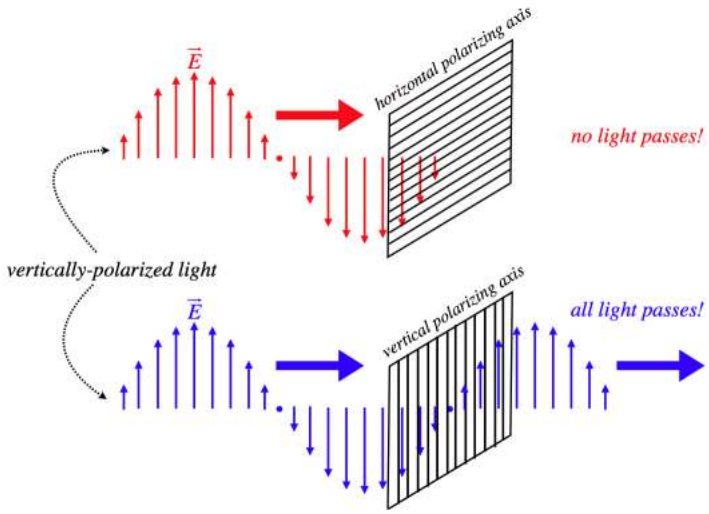
3.7: Polarization

Polarization Filters (Polaroids)

As stated previously when discussing the speed of light waves through transparent media, the mechanisms that govern light propagation through media are complicated. There is little we can say about it in this class, except to say that because the light wave is electromagnetic in nature, it interacts with electric charge, which is present in all matter. It so happens that it is possible to construct a solid substance which greatly restricts oscillatory motion of electric charges along a single dimension. The upshot of this is that the charges react to electric fields along one direction (or rather, components of electric fields along one direction), while they don't react along a perpendicular direction.

This material can have a dramatic effect on light passing through it. If the light is plane-polarized (see Figure 3.1.1), then its propagation through a medium will be affected by the preferential orientation of charge oscillations. When the light polarization is aligned with what we define as the *polarizing axis* of the substance, then little of the light is absorbed by the substance (i.e. the substance is transparent to this light), while if the light is polarized perpendicular to the polarizing axis, then virtually all of the light is absorbed. Such a filter is called a *polaroid* or *polarizer*.

Figure 3.7.1 – Light Through a Polaroid



One interesting application of this phenomenon is 3-D movies. Long ago someone came up with a brilliant idea for making movies projected onto a 2-D screen appear in 3-D. The idea is based on the fact that a large component (but not the only one) of seeing in 3-D is stereo vision. Your right eye sees objects from one perspective, while your left eye sees it from a slightly different perspective. You can see this is true by holding up your finger in a fixed position and alternately opening-and-closing each eye. Your finger's position appears to change relative to the background. This inventor's idea was to project not one but *two* images on the same screen. One image is recorded from the perspective of the right eye, and the other from the perspective of the left eye. Then the trick is to make the right-perspective image invisible to the left eye and the left-perspective image invisible to the right eye, so that each eye sees only its own perspective. The original inventor did this with colors – red lenses obscure red images, and yellow lenses obscure yellow light, so films were recorded from two perspectives, and each perspective was projected in a different color – one red and one yellow. But today we like our movies to be in realistic colors, so someone came up with the idea of projecting the two images with differently-polarized light, and then give viewers glasses that only admit the properly-polarized light into the respective eyes.

We have overly-simplified things here, in a couple of ways. First of all, a light wave does not have to arrive at the polarizer in either a parallel or perpendicular orientation – it could be aligned at any angle with the polarizing axis. What happens then? Well, electric fields are *vector fields*, which means they can be broken into components, so the component of the electric field that is parallel to the polarizing axis gets through, and the other component is absorbed.

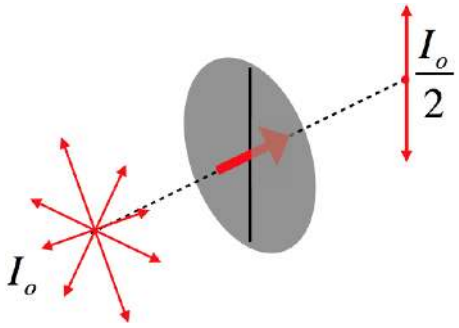
The second oversimplification is that not all of the individual light waves that come from a source are necessarily polarized in the same direction. In fact "natural" light from light bulbs and the sun is "unpolarized," which comes about because each of the individual light sources (atoms) are aligned in random orientations, and all send out random, unaligned light waves. When such

light is passed through a polaroid, half the light gets through. To see why this should be so, break every electric field vector of every wave into components parallel and perpendicular to the polarizing axis. Because the wave polarization directions are randomly-oriented, there is no reason to expect there to be a greater sum of components along one axis than another. By “half the light gets through,” what do we mean? We mean that the intensity drops by one half. We look at the more general case of intensity next.

Intensity

We can express the fact that half of natural light gets through a polaroid in a diagram as follows:

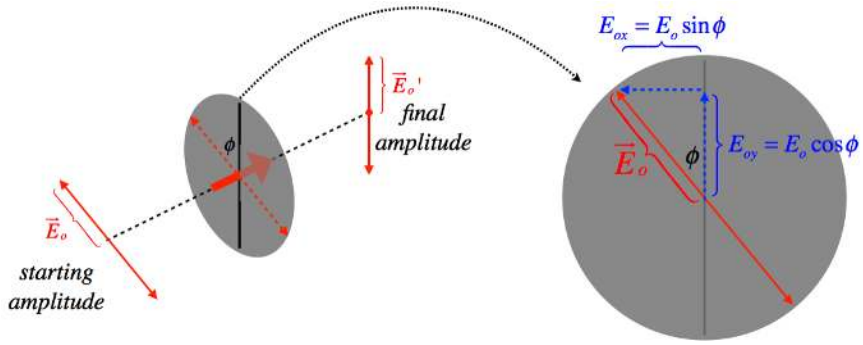
Figure 3.7.2 – Natural Light Intensity Halves Through a Polaroid



Now let's consider what happens if we send the natural light through *two* polaroids in succession. Clearly when the light reaches the second polaroid it will be plane-polarized from the first one. If the second polaroid is oriented the same as the first, then all the light gets through, and the intensity is unchanged, and if its polarizing axis is at right angles to the first polaroid, then no light will get through it. But now we seek to determine the intensity of the light that passes through the second polaroid if the angle between their polarizing axes is somewhere between 0^o and 90^o.

This process all comes down to what happens to the electric field vectors. After passing through the first polaroid, all the electric field vectors are aligned with that polaroid's polarizing axis. When those vectors come upon the second polaroid, just the component of the field vector that is aligned with the new axis gets through, resulting in a new vector shorter than the original.

Figure 3.7.3 – Polarized Light Through a Polaroid



Resolving the original electric field vector into components parallel and perpendicular to the polarizing axis, and keeping only the parallel part means that the new electric field vector magnitude is:

$$E_o' = E_o \cos\phi \tag{3.7.1}$$

The electric field vector is the *amplitude* of the light wave, and we are interested in the *intensity*. As with any other wave, the intensity is proportional to the square of the amplitude, so the relationship between the outgoing intensity I and incoming intensity I_o is:

$$I = I_o \cos^2\phi \tag{3.7.2}$$

This is known as *Malus's law*. Notice that it works exactly as we expect for the cases where the angle happens to be 0^o and 90^o.

Example 3.7.1

Unpolarized light enters a series of four polaroids with axes of polarization that are each rotated 30^o clockwise from the previous polaroid, making angles of 0^o, 30^o, 60^o, and 90^o with some common reference point. What fraction of the intensity of the incoming light is the intensity of the outgoing light?

Solution

When the unpolarized light passes through the first filter, the intensity is cut in half and comes out polarized at 0^o. Then it passes through three successive filters, and applying Malus's law for each 30^o change of polarization angle brings in a factor of 0.75 for each polaroid. The result is that the final intensity is:

$$I = I_o \left(\frac{1}{2}\right)\left(\cos^2 30^o\right)^3 = I_o \left(\frac{1}{2}\right)\left(\frac{3}{4}\right)^3 = \frac{27}{128} I_o$$

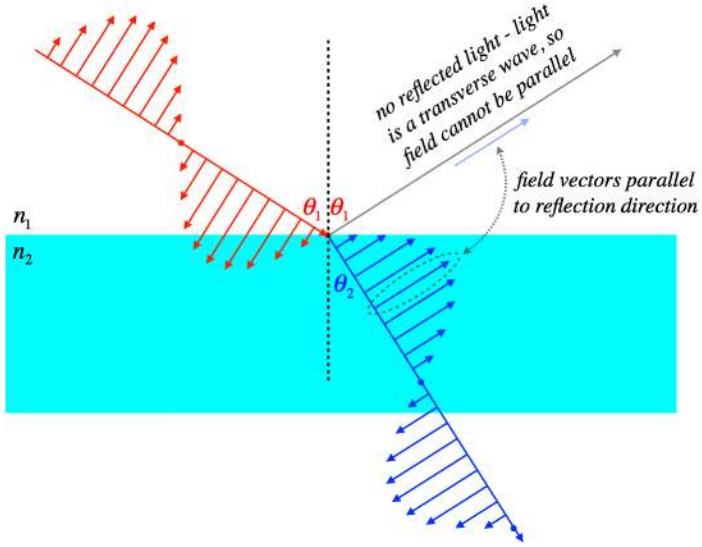
*One might expect that since the first and last polaroids are at right angles to each other, no light at all should emerge from the last polaroid. But when the light passes through a polaroid, it gains a new polarization aligned with that polaroid's polarization axis, and has no "memory" of its previous plane of polarization. Unless two **consecutive** polaroids are at right angles, some light will always get through each polaroid.*

Polarization By Reflection

While most natural light is unpolarized and we can polarize it with a polaroid, it turns out that is not the only way it can be polarized. A more "natural" way to create polarized light exists thanks to reflection. As we have said many times, when light (or any wave) strikes an interface between two media, it is partially transmitted and partially reflected.

Consider the following scenario: Light polarized in the vertical direction strikes an interface between media such that the reflected ray aligns with the electric field vectors of the transmitted ray. There is an important principle in physics that states that the conditions at the boundary have to work out properly. This means that the electric field vector of the incoming light must add up properly to the electric field after striking the interface. The electric field vector can of course be written in components with the "x-direction" being the electric field direction of the transmitted wave, and the "y-direction" being the direction of the reflected ray (which is perpendicular to the transmitted ray). But the outgoing light cannot have an electric field vector pointing along its direction of motion (light is a transverse wave), so no light reflects!

Figure 3.7.4 – Reflection of Polarized Light



Of course this result is only for vertically-polarized incoming light, so unpolarized light that reflects at this angle will have its vertical component removed, which means that the reflected light is horizontally-polarized. More generally, light that is reflected off a surface at just the right angle will be polarized *parallel to that surface*. It also happens that if the angle is not just right, then while the light is not entirely polarized, it is partially so (depending upon how close to the correct angle the reflection is). By "partially polarized," we mean that the amplitude of light waves measured (using a polaroid) along one direction is not the same as

the amplitude measured along the orthogonal direction. In practice this means that a polaroid aligned parallel to a surface from which the light is reflected will admit more light than a polaroid aligned perpendicular to that surface.

We can easily write down an expression for the "special angle" at which total polarization occurs (this is known as the *Brewster angle*), by noting that for this angle the reflected ray makes a right angle with the transmitted ray (because the field vector of the transmitted wave is perpendicular to the transmitted ray and is parallel to the reflected ray). Combining this fact with Snell's law gives the Brewster angle, θ_B:

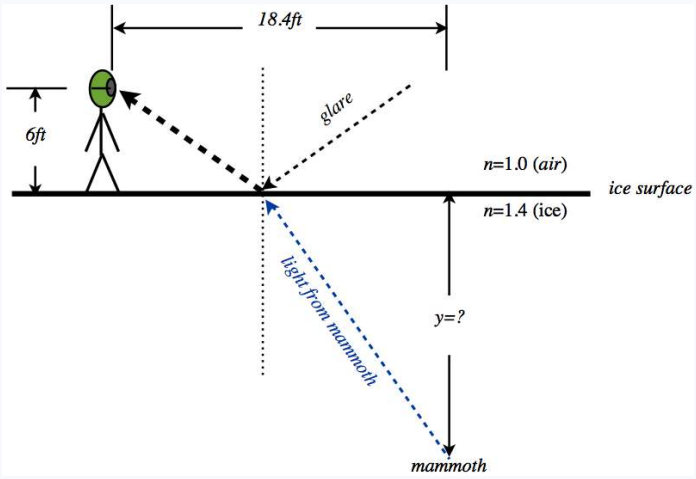
$$n_1 \sin\theta_B = n_2 \sin\theta_2 = n_2 \sin(90^o - \theta_B) = n_2 \cos\theta_B \quad \Rightarrow \quad \tan\theta_B = \frac{n_2}{n_1}\,, \tag{3.7.3}$$

where n_1 is the index of refraction of the medium within which the reflection is occurring, and n_2 is the index of refraction of the medium off which the reflection is occurring.

A nice application of this effect involves polaroid sunglasses. Most glare from sunlight comes off surfaces that are horizontal (roads, lake surfaces, etc.), which means that the light that reflects off such surfaces has a relatively small fraction of its polarization in the vertical direction. This means that if we place polaroids in front of our eyes that are allow only vertically-polarized light to pass, then very little of the horizontally-polarized glare gets through. Of course, only half of the non-glare light gets through as well, but at least one's vision of light of important objects (on coming cars or boats, etc.) does not have to compete with the incoming light from glare.

Example 3.7.2

A paleontologist is looking for the remains of a wooly mammoth in an unusually clear section of a glacier. The glare off the ice from the sun makes it hard for her to see, so she puts on her polarized sun glasses and is immediately rewarded when, along the line where the glare is cut to zero, she finds what she is looking for. Now she just needs to figure out how deep the carcass is. Fortunately she has a physicist (you) on staff. You measure the height of her eyes above the ice surface to be $6\ ft$, and you measure the distance from the position where she first saw the beast through the glare, to the point where you can look straight down at it. This distance is $18.4\ ft$. You estimate the index of refraction of the ice to be 1.4. Find the depth of the wooly mammoth.



Solution

For the polarized sunglasses to remove all the glare, the angle the light makes with the perpendicular to the ice must be Brewster's angle, so:

$$\tan\theta = \frac{n_2}{n_1} = 1.4$$

From the right triangle on the left, we can derive the distance from the paleontologist to the point of reflection:

$$\tan\theta = \frac{x_1}{6\ ft} \quad \Rightarrow \quad x_1 = (1.4)\,(6\ ft) = 8.4\ ft$$

We can use this distance to derive the horizontal distance from the point of reflection to the point on the ice directly above the mammoth:

$$x_2 = 18.4\ ft - 8.4\ ft = 10.0\ ft$$

The Brewster angle occurs when the reflected light makes a right angle with the transmitted light, and from symmetry (just reverse the direction of the light to see this), that is also true of the incoming glare and the light from the mammoth. Therefore we can use x_2 and the tangent of the angle to get the depth:

$$\tan\theta = \frac{y}{x_2} \quad \Rightarrow \quad y = (1.4)\,(10.0\ ft) = 14.0 ft$$

CHAPTER OVERVIEW

4: Geometrical Optics

4.1: Images

Using Rays to Locate Light Sources

In our discussion leading to the law of reflection, we found (Figure 3.6.6) that if we view the light from a point source after it is reflected off a plane, then the spherical wave we see appears to originate not from the point source, but from another point *behind* the reflecting plane. We will now look more generally at locating this apparent origin of a source of light (which we have dubbed "image"). We are through with doing the hard work of analyzing optical systems with Huygens's principle and wave propagation, however. Instead, we will now embrace the tool of rays, keeping in mind as we do the fact that rays do not represent what is actually happening with light – they just make it easier to do the geometry of linking actual light sources with the images observed. This approach to the study of certain aspects of light is therefore generally referred to as *geometrical optics*.

Alert

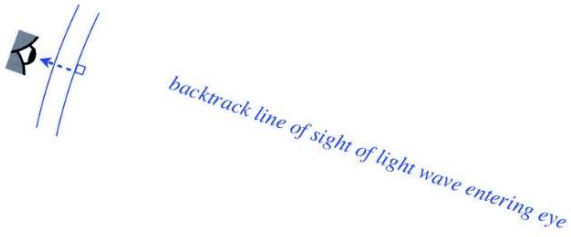
This point cannot be emphasized enough: We will sketch rays so that we can do geometry, but these are just imaginary lines that we use for our own purposes to simplify our work – they do not represent actual light, and in many cases we will see, there isn't any light at all where we have drawn the ray!

Up to now, we have discussed only plane waves and point sources of light, but when we look at something, we receive light from every part of it that is visible to us. That is, there are actually several point sources of light. For the purpose of simplifying our analysis, we will reduce the objects we look at to their bare essence, but they must be more nuanced than merely a single point. At the very least, we would like to have some measure of the size and orientation of what we are looking at. As we will see, the path that light takes from the object to our eye can alter not only the location of the apparent source of the light, but also what we perceive to be its size and orientation.

The simplest way to exhibit the two attributes of size and orientation is with an arrow. The directionality of the arrow gives us a sense of orientation, and the length of the arrow a measure of its size. If the *object* arrow points upward, then depending upon what happens to the light in transit from the object to the observer, the *image* arrow will either point upward as well (it is *upright*, or *erect*), or it will reverse direction and point downward (*inverted*). When we compare the lengths of the object and image arrows, the image arrow is *laterally magnified* if it is longer than the object arrow, and *laterally diminished* if it is shorter. Doing this reduces our analysis to using rays and geometry to determine where the base and point of the image arrow are located.

Rays radiate outward from a point source (such as the point of the image arrow). If we draw a single ray that enters our eye, we only know the direction from which the light has come, not how far away it is. The way that we locate the source of the light is to view *another* ray. This allows us to backtrack the rays to see where they intersect.

Figure 4.1.1 – Finding an Image



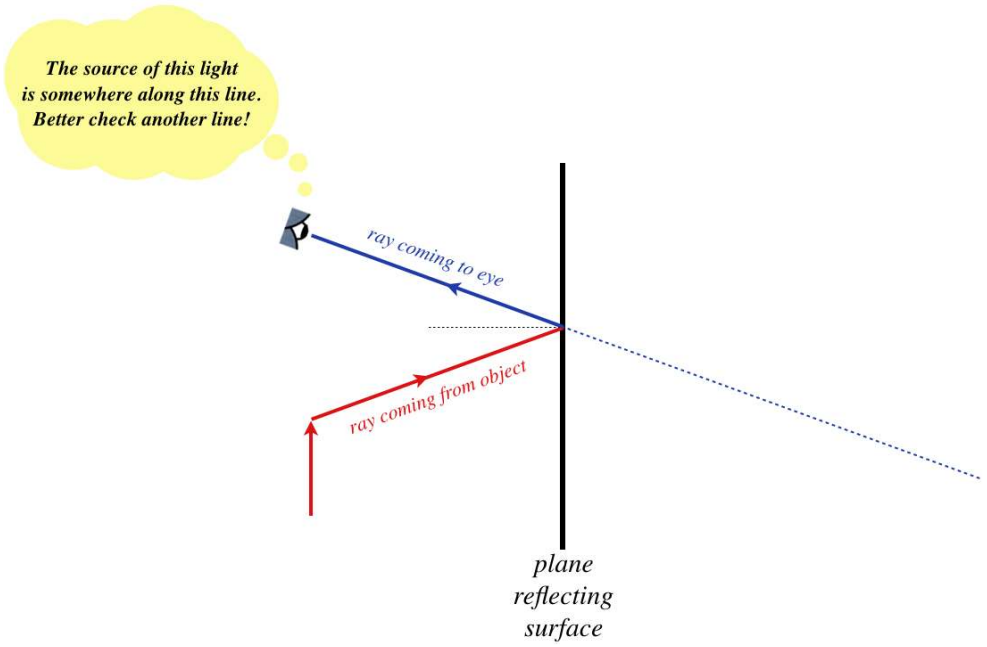
To get the full arrow image, this process would need to be followed for both the pointed end and the base. When looking directly at an object, following this process locates the image exactly where the object is located. The image location (where we see the light coming from) is the same as the object location (where the light is actually coming from), the size of the image is the same as the size of the object, and the image is the same orientation as the object. This is all so boring! What interests us is when the image data

differs from the object data, and this can only be accomplished when the light reaches our eyes in a less direct manner. Then when we backtrack the rays they will converge at a different place than the source from which they actually originated.

Plane Reflector

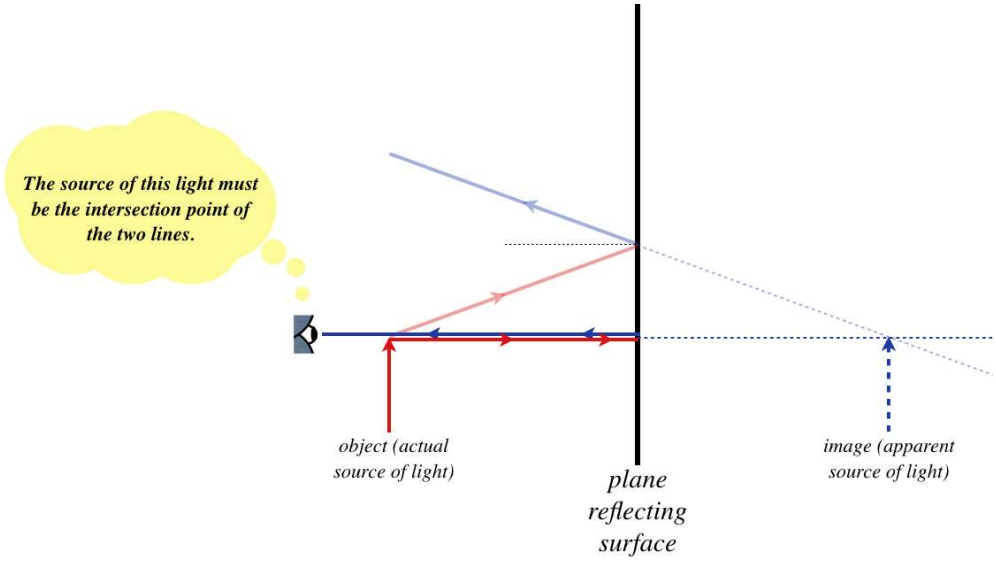
We used a clever symmetry argument with Huygens's principle in Section 3.6 to find the image of a point source of light in the presence of a plane reflector. We'll now repeat this conclusion with the ray method, using the law of reflection and basic geometry. We start by drawing the object, and selecting a ray that comes out of it, reflects, and enters an eye.

Figure 4.1.2 – Plane Reflector: First Ray



The law of reflection assures us that the angle that the ray coming from the object makes with the perpendicular to the plane is the same as the angle the ray coming into the eye makes. We'll use this fact in a moment, but first we need to pick another ray. To get the answer we can pick any ray we like, but as using these rays is merely our trick for finding the image, we might as well make the geometry we have to do as easy on ourselves as possible. Let's choose as our second ray the one that strikes the reflector perpendicularly, and comes straight back (we can place the eye directly behind the tip of the arrow).

Figure 4.1.3 – Plane Reflector: Two Intersecting Rays



Okay, now for the geometry. Thanks to the law of reflection, we can see that the angle the first (blue) image ray makes with the horizontal is the same as the angle that the first (red) object ray makes with the horizontal. The two right triangles formed by rays on the opposite sides of the reflector are therefore similar, and since they share a side, they are congruent. The image is therefore the same distance to the right of the reflector as the object is to the left of it. If we repeat this process for the base of the arrow, we get the same result, and with the object and image bases and tips directly across from each other, the image is clearly the same size as the object (neither magnified nor diminished), and the image is upright (the same orientation as the object).

As simple as this example is, it shows clearly a case where the *apparent* source of light (i.e. the image) is disconnected from the *actual* source of light. No matter how complicated the geometry might get, the method of finding the image is the same – determine the position that the observer backtracks two (or more) rays to.

Alert

One last warning, then you are on your own... It is easy to get caught up in the process of finding the image (backtrack rays, do geometry, etc.), but this process is not tracking any actual ***physical*** *process – the light is not traveling in rays, and in fact there is* ***not even any light*** *where we have drawn image rays to the right of the reflector! Using rays is just a trick that makes our lives easier. For this plane refractor, what is actually physically happening to the light is better depicted by Figure 3.6.5.*

Sign Conventions

Not all examples in geometrical optics are as easy to do geometrically as the plane reflector, so it is helpful to develop some mathematical definitions, formulas, and conventions that can be used for all cases. We'll begin the process of making such definitions here, and will continue to add to them as the number of situations we can cover grows. In particular, in order for mathematical formulas to work out properly, we will need to define when a variable is positive or negative. These are referred to as *sign conventions*, and at times these can be quite confusing. This is because – at least at first – the conventions will seem overly complicated or cryptic. This is necessary, because making definitions that are too simple forces us to make new definitions later for more complicated situations, and we want to avoid this.

For the plane refractor (and in fact for every optical device we will examine), there is one critical region along the horizontal axis. For the case above, it is the plane at which the reflection occurs. We choose this important boundary to be an "origin" of sorts, from which distances are measured. We start with the following variable definitions and their associated sign conventions:

name	symbol	positive when	negative when
$object\ distance$	s	$object\ on\ incoming\ side$	$object\ on\ other\ side$
$image\ distance$	s'	$image\ on\ outgoing\ side$	$image\ on\ other\ side$

(4.1.1)

Okay a clarification is in order here. "Incoming" and "outgoing" refer to the direction of the rays relative to the origin – rays pointing toward the reflector (or more generally, the "origin") are on the "incoming" side, and rays pointing away from it are on the outgoing side. While "incoming" and "outgoing" sound like opposites, they are not. This is very clear for the case of the reflector, where the incoming and outgoing sides are the same (the rays come in from the left, and go back out to the left). But in other cases we will examine, the rays will pass through the "origin," making the incoming and outgoing sides different.

For the plane reflector, we found that the magnitudes of the object and image distances are equal, but what about the signs? Well, the light is coming into the reflector from the left, and the object is on the left, so it is on the incoming side, which means that $s > 0$. The light is going away from the mirror on the left side, so it is the outgoing side, and the image is located on the *other* side, so $s' < 0$. We therefore conclude, for the plane reflector:

$$\text{plane reflector:} \quad s = -s' \tag{4.1.2}$$

We will see many more sign conventions, and we'll add them to the list as they come up.

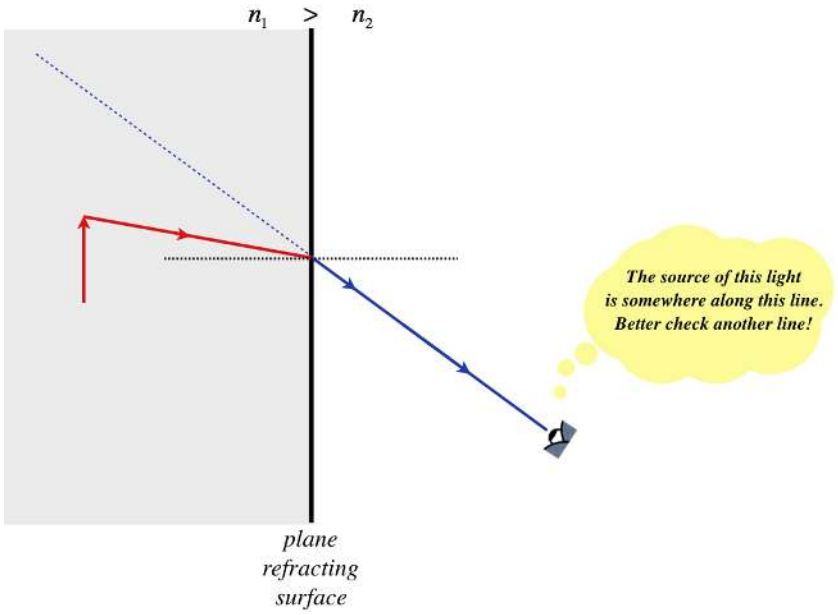
Plane Refractor

Obviously any way that we can get the rays representing the light to change direction will lead to a case where the image position does not coincide with the object position. Reflection is one way to make this happen, and another way is refraction. We will therefore examine the simplest case of a refraction-induced image – the plane refractor. This consists of two regions with different indices of refraction, delineated by a plane boundary, such as looking at an object below the surface of a pool of still water.

The light coming from the object (which we will say is in a medium with index of refraction n_1) will bend as it crosses the boundary into a new medium, where the observer resides, with index of refraction n_2, according to Snell's law (Equation 3.6.4). In this case, the plane boundary between the two media is the "origin" from which we measure everything.

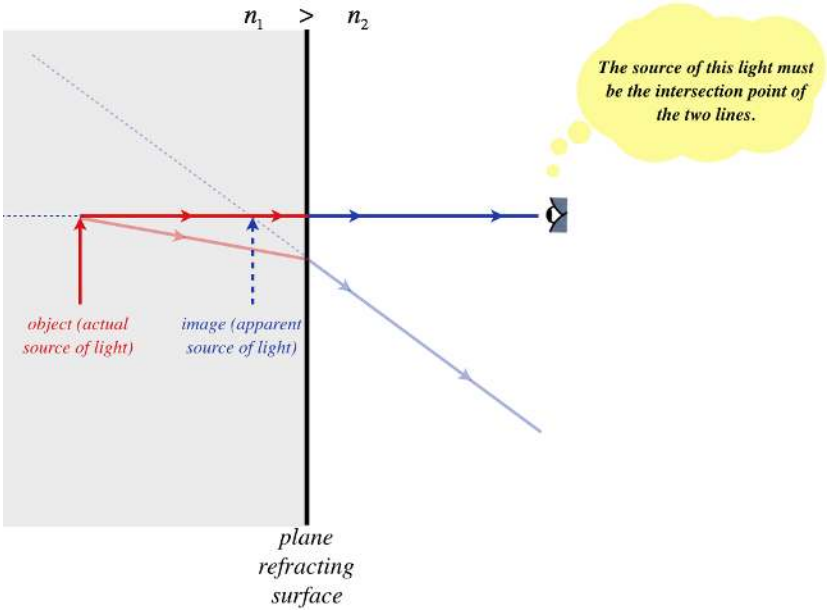
Now proceed as we did for the reflector: Start with an arbitrary ray that comes from the object, passes through the boundary, and enters the eye of an observer. For the sake of drawing a figure, we will look at the case that the index of refraction of the medium where the object resides is greater than the index of refraction of the medium of the observer.

Figure 4.1.4 – Plane Refractor: First Ray



We need another ray to find the intersection point that defines the position of the image, and once again we will make things easy on ourselves by choosing the ray that strikes the refracting plane perpendicularly. This ray passes straight through without bending, making the location of the intersection point quite simple.

Figure 4.1.5 – Plane Refractor: Two Intersecting Rays



Before we go on to do the math associated with the geometry, there is one important thing to point out here. Unlike the plane reflector case, where we could show using Huygens's principle that all of the rays that emerge backtrack to the same point, it is far

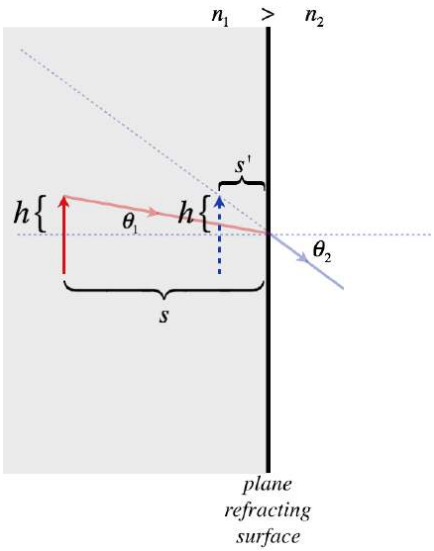
from clear that this must be true in this case. In fact it *isn't true*! That is, if we carefully keep drawing rays for other angles, the angled rays will intersect the direct (un-deflected) ray at different points. This is not difficult to prove: There is some point on the surface for which a ray from the object is totally internally reflected. Light rays coming out of the object at angles greater than this never escape the medium, so they cannot be viewed by the observer, so they cannot be backtracked to the same image as any other rays.

Despite this "problem," we can nevertheless get a useful *approximate* result. If we assume that the object is viewed from angles not too far off a direct viewing (i.e. small refracted angles), then the image's position can be determined to a close approximation. Under this approximation, we'll first assert the following, which will be useful shortly:

$$\text{for a small angle } \theta: \quad \sin\theta \approx \tan\theta \tag{4.1.3}$$

Now let's label some values from the figure above. Namely, we'll define the incoming angle θ_1, the refracted angle θ_2, the height of the above the axis h, the object distance s, and the image distance s' (note we have already defined the two indices of refraction):

Figure 4.1.6 – Geometry of the Plane Refractor



Start by constructing the tangents of the two angles:

$$\tan\theta_1 = \frac{h}{s} \qquad \tan\theta_2 = \frac{h}{s'} \tag{4.1.4}$$

Now apply the small angle approximation to turn these tangents into sines, and apply Snell's law:

$$\left.\begin{aligned} \frac{h}{s} &\approx \sin\theta_1 \\ \frac{h}{s'} &\approx \sin\theta_2 \\ n_1\sin\theta_1 &= n_2\sin\theta_2 \end{aligned}\right\} \quad s' = \frac{n_2}{n_1}s \tag{4.1.5}$$

So just as for the plane reflector where we found the relationship between the object and image distance (Equation 4.1.2), we have done the same for the plane reflector. Let's check to see if the sign conventions work out properly for this equation. The object is to the left of the refracting plane, and the light reaches the refracting plane from the left, so the object is on the incoming side, which means that $s > 0$. The light is leaving the refracting plane going to the right (to the observer), so the right hand side is the outgoing side. The image is on the left, so it is on the side "other than" the outgoing side. Therefore the image distance is actually negative. Indices of refraction are always positive, so we need a minus sign to make the formula correct with our sign conventions:

$$\text{plane refractor:} \qquad s' = -\frac{n_2}{n_1}s \tag{4.1.6}$$

We should also note that the magnitudes work out properly – the fact that $n_1 > n_2$ assures that the image is closer to the refracting plane than the object. If the object had instead been in the region with the lower index of refraction, then the image would have

been farther from the plane of refraction than the object.

4.2: Magnification

Lateral Magnification

You may have noticed that in the diagrams for the plane reflector and refractor, we oriented the object arrow so that it was parallel to the reflecting/refracting surface. We should take a moment to say a few words about this, as well as define a few terms.

The horizontal line that is perpendicular to and bisects the surface that causes reflection or refraction is called the *optical axis*. We will almost exclusively work with object arrows that are oriented perpendicular to this axis, as we did in the previous section. The reason for this is that the "object and image distances" are only well-defined when every point is the same distance from the surface. Shortly we will say a few words about objects that are not oriented so conveniently, but unless explicitly stated, one should assume that this perpendicular-to-optical-axis is in effect.

In the cases of both the plane reflector and plane refractor, we found that the object and image arrows were exactly the same size. This result is special to these cases – in the more general cases we will examine soon, the object and image sizes will be different. We define the *lateral magnification* of the image as the ratio of the image size y' to the object size y:

$$M \equiv \frac{y'}{y} \tag{4.2.1}$$

For both the plane reflector and the plane refractor that we examined in the previous section, this ratio is 1. We know this because when we compute the positions of the head and base of the image arrow, we find that the separations of these points y' is the same as the distance separating the head and base of the object arrow y.

As with every quantity we will define in geometrical optics, there is a sign convention defined for lateral magnification. We do this by insisting that the values of y and y' are positive when the arrow is pointing in one direction, and negative when pointing in the other. Let's say we define upward as the positive direction for the two cases we have studied so far. In both cases, an up-oriented object arrow resulted in a up-oriented image arrow, so for both cases y and y' were both positive. Therefore the lateral magnifications for both cases are positive. Clearly when the orientations of the object and image are the same (either both positive or both negative) – which we have previously defined as "upright") then the magnification is positive, and when the image is inverted relative to the object, the magnification is negative. Adding this sign convention to our list gives:

name	symbol	positive when	negative when
$object\ distance$	s	$object\ on\ incoming\ side$	$object\ on\ other\ side$
$image\ distance$	s'	$image\ on\ outgoing\ side$	$image\ on\ other\ side$
$lateral\ magnification$	M	$image\ upright$	$image\ inverted$

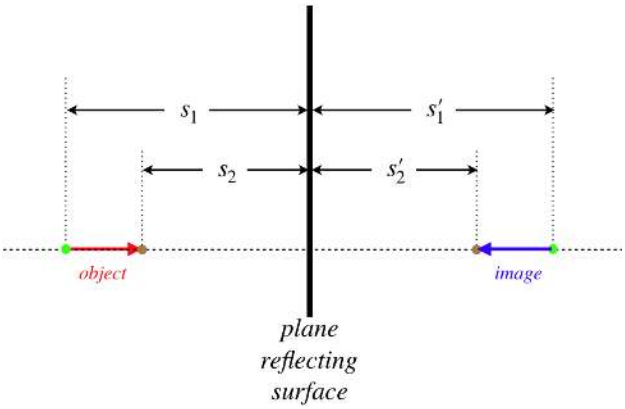
(4.2.2)

Longitudinal Magnification

The word "lateral" is appended above because it only applies to the dimensions of the objects perpendicular to the optical axis. As stated above, this is almost exclusively what we will be working with, but before we forge ahead, let's take a moment to look at magnification in the cases of the plane reflector when the object arrow is parallel to the optical axis.

Calling the object length y, we seek to compute the image length y'. We do this by computing the position of each end of the image arrow separately, and then using those positions to compute the length. The image of each point is exactly the same distance from the reflecting plane on the opposite side:

Figure 4.2.1 – Longitudinal Reflection



The extension of the object is $y = s_2 - s_1$ and the extension of the image is $y' = s_2' - s_1'$, and since, for the plane reflector $s = -s'$, we get that these are the same length. We can use the same definition for the magnitude of longitudinal magnification as we did for lateral magnification (keeping in mind here that the measurement of y is along the optical axis rather than perpendicular to it), giving:

$$|M| = \frac{|y'|}{|y|} = 1 \tag{4.2.3}$$

The magnitude of 1 indicates the object and image are the same length. We see also that the arrows are pointing in opposite directions, which means the image is inverted. So both the lateral and longitudinal cases of the plane reflector result in images that are neither magnified nor diminished, but in the lateral case the image is upright, while it is inverted in the longitudinal case.

Example 4.2.1

An object in a medium with index of refraction n_1 is viewed through a plane refractor from a medium with an index of refraction n_2. Find the magnitude of longitudinal magnification of the image, and indicate whether the image is upright or inverted.

Solution

We follow the same procedure as above, computing distances to the refracting plane for the front and rear of the image arrow. Defining, as we did above, s_1 and s_2 as the object distances of the rear and point respectively of the object arrow, we compute the image distances using Equation 4.1.6 to be:

$$s_1' = -\frac{n_2}{n_1} s_1 \ , \qquad s_2' = -\frac{n_2}{n_1} s_2$$

As above, we write the lengths of the objects and images as the differences of these quantities:

$$y = s_2 - s_1 \ , \qquad y' = s_2' - s_1'$$

Putting these together and computing the magnitude gives:

$$|M| = \frac{|y'|}{|y|} = \frac{\left|\frac{n_2}{n_1} s_2 - \frac{n_2}{n_1} s_1\right|}{|s_2 - s_1|} = \frac{n_2}{n_1}$$

So the image is magnified if $n_2 > n_1$, and diminished if the opposite is true. Clearly the object and image arrows point in the same direction (whichever end of the object is closer to the plane, that same end is closer to the plane for the image) so this image is upright.

Interesting that this result is the exact opposite of the plane reflector – upright instead of inverted, and either magnified or diminished rather than unchanged.

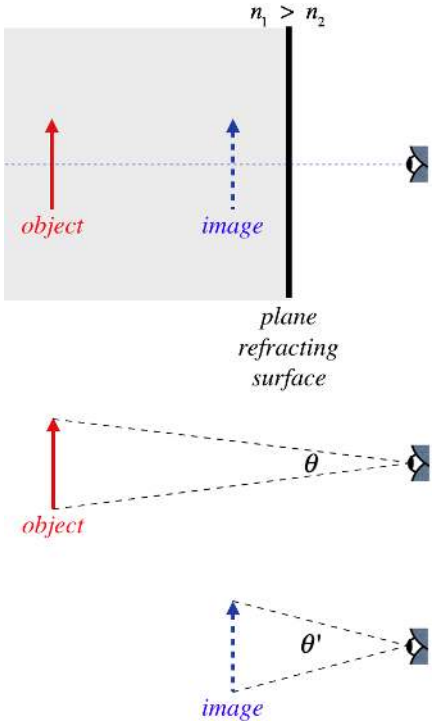
Angular Magnification

As logical as these two definitions of magnification are, neither of them actually captures what we think of when we hear the word "magnification." When an object we see is magnified, we generally mean that we see it *better* – its details are more evident. An

item does not increase in size (its dimensions remain the same), but we can see it better when it is closer. When the light from an object is refracted by a plane such that the image is closer to the viewer than the object, the image and object are the same size, *but the image is closer, so we can see it better*. This is a condition that most people would refer to as "magnified." Indeed, if you look at a penny at the bottom of a swimming pool, it certainly *looks* larger than another penny in the same position with the pool empty. So how do we account for this commonsense definition of magnification?

What makes something look large or small regardless of its actual size is the percentage of our *field of view* that it captures. This is determined by the *angle* the image subtends in our view. Let's see what this means for the refracting plane. When the object is in a region with a higher index of refraction, we know that the image is the same size as the object, but it is closer to the observer, causing it to subtend a larger angle.

Figure 4.2.2 – Angular Magnification By Refracting Plane



Clearly $\theta' > \theta$, which means the image takes up a bigger portion of our field of view than the object would, if it was viewed directly (without the refracting plane). We can measure the amount of this *angular magnification* with a simple ratio of the angles:

$$M_\theta \equiv \frac{\theta'}{\theta} \tag{4.2.4}$$

Alert

While we define a sign convention for lateral magnification to indicate whether the image is upright or inverted, for angular magnification, it is standard to only take into account the magnitude, so the directions that the angles are swept-out are not important, and we essentially just use absolute values. It is understood that the lateral magnification takes into account everything about the image, while the angular magnification only handles the size of the field of view that the image occupies.

Example 4.2.2

An object in a medium with index of refraction n_1 is viewed through a plane refractor from a medium with an index of refraction n_2. If the viewer is positioned right next to the plane refractor, find the angular magnification of the image.

Solution

We can use the figure above, except imagine that the eye is right at the plane. The optical axis divides both the object and image in half, so the ratio of these half-angles will also equal the angular magnification. These angles are part of right

triangles formed by the object and image, with one side (the height of half the object/image, which we will call y) being the same for both. The tangents of the angles are therefore:

$$\tan\theta = \frac{y}{s}, \qquad \tan\theta' = \frac{y}{s'}$$

The angles, as aleways, are assume to be small, so the tangents are approximately equal to the angles themselves (measured in radians), which means we have:

$$M_\theta = \frac{\theta'}{\theta} \approx \frac{\tan\theta'}{\tan\theta} = \frac{\frac{y}{s'}}{\frac{y}{s}} = \frac{s}{s'}$$

Now just plug in Equation 4.1.6:

$$M_\theta = \frac{n_1}{n_2}$$

Sure enough, an observer looking at an object that is within a medium of higher index of refraction than the surrounding region sees an image that is magnified angularly (i.e. takes up a wider field of view), even though the image is no larger than the object, because the image is just closer.

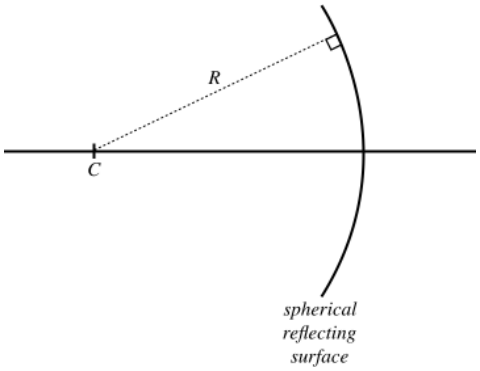
4.3: Spherical Reflectors

Finding the location of an image viewed in a plane reflector is a rather simple matter, but now we will look at more interesting cases where the reflector (which we will often refer to as "mirrors," though this does not need to be specifically the case) is curved. We will not examine *general* curves (the analytical geometry would be quite daunting!), but will stick to spherical surfaces. Such a surface comes in two varieties – *concave* and *convex*. The former refers to a shape where the light enters the hollow region (like a cave!) before reflecting, and the latter refers to the opposite shape – one that bulges out into the region where the light is.

Concave Mirrors – Object and Image Distances

We begin by defining the geometry we are working with. This mirror's curve is a section from a sphere, which means that if we draw lines perpendicular from its surface, the lines all intersect at a single point, which we call "C." The distance to this point from the mirror is the radius of curvature of the sphere, R. The optical axis intersects the mirror perpendicularly at its center point, called its *vertex*.

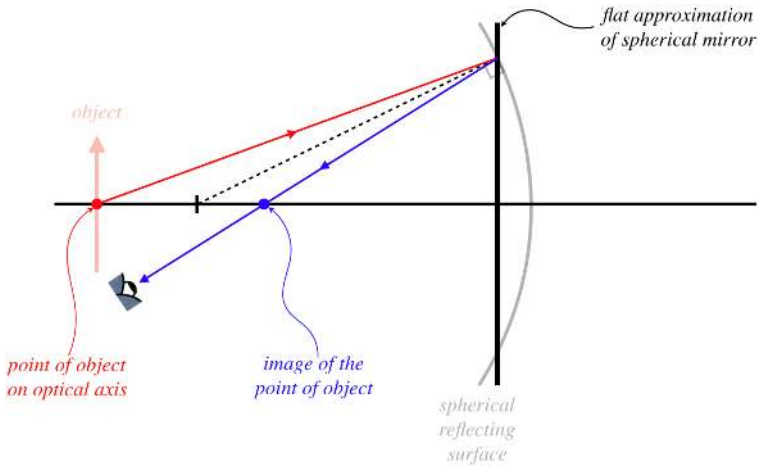
Figure 4.3.1 – Geometry of Concave Mirror



From here we use our special trick of considering rays, rather than waves. We do this by employing the law of reflection and doing the necessary geometry. Our goal here is to use this method to derive the position of an image s' in terms of the position of the object s, and the only other variable we have here, which is R. The trouble is, how do we define s and s' in this case? In the case of the flat plane, every point on our object arrow (which is perpendicular to the optical axis) is the same distance from the the reflector, but that is clearly not the case for this curved surface. We therefore make the assumption that the radius of curvature of the mirror is large enough (i.e. it is flat enough) that the mirror's position can be treated like the position of a plane, and this is the position from which the quantities s and s' are measured. The result is that an object that is totally lateral (perpendicular to the optical axis) will result in an image that is also lateral.

Using this "flat approximation," we will do the geometry to find the image of a point on the object that lies on the optical axis. The diagram will be clearer for this approximation if we draw the mirror as a straight vertical line, and for no particular reason, we will choose the object to be outside of the center point of the mirror. The critical principle here is the law of reflection. A ray starting at the object that strikes the mirror will reflect off it such that the incident angle equals the reflected angle, where these angles are measured relative to the perpendicular – the radial line shown in the figure above.

Figure 4.3.2 – Image of a Point on the Optical Axis – Picture



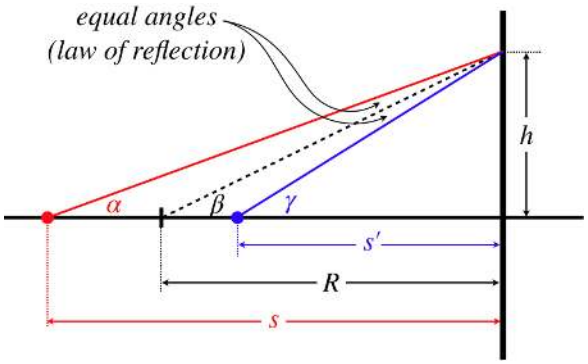
We are only looking for how the object and image distances are related here, and since we are assuming that the image remains lateral, we only need to do the calculation for a single point (we will have more work to do later to determine the nature of the full image). But how do we know that the image of this point also lies on the optical axis? Remember, just viewing a single ray is not enough to locate the image! We need a second ray. Rather than clutter our diagram, we simply note that the ray that leaves the same point heading *along* the optical axis strikes the mirror at a right angle, and therefore reflects straight back. That gives us a ray intersection that lands right at the point indicated – indeed the image of a point on the optical axis must also lie on the optical axis.

Digression

As with the case of the image due to a flat refractor, if one carefully sketches the rays properly for a spherical mirror, the rays actually do not all pass through a common point. For our purposes, this deviation from an exact result will not be a problem, but for optical systems involving spherical surfaces that require great precision, this unwanted blurriness of the image is a consequence of the inexactitude of the geometry known as spherical aberration.

Okay, so let's clear away the clutter of the above figure and label all the important features so that we can do some geometry and apply our usual small angle approximation...

Figure 4.3.3 – Image of a Point on the Optical Axis – Geometry



It's obvious that the angle β has a value that falls between the values of the angles α (which is smaller) and γ (which is larger). But when the other two angles indicated in the triangle happen to be equal as they are in this case thanks to the law of reflection, it happens that the angle β is precisely the *arithmetic mean* of α and γ (geometric proof of this feature of interior and exterior angles of triangles is left as an exercise for the reader). In other words:

$$\beta = \frac{\alpha+\gamma}{2} \quad \Rightarrow \quad \alpha+\gamma = 2\beta \tag{4.3.1}$$

We can read off the tangents of these angles, and apply the small-angle approximation to get:

$$\begin{aligned} \alpha &\approx \tan\alpha = \frac{h}{s} \\ \beta &\approx \tan\beta = \frac{h}{R} \\ \gamma &\approx \tan\gamma = \frac{h}{s'} \end{aligned} \tag{4.3.2}$$

Combining these equations, we get the following simple relationship between the object and image distances and the radius of curvature of the mirror:

$$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} \tag{4.3.3}$$

We have dealt with these distances as absolute values, so we need to double-check this result and make sure it satisfies our sign conventions. The light is coming into the mirror from the left, and the object is on the left, so according to our sign convention, the object distance is positive. The light leaves the mirror on the left side, and the image is on the left side, so the image is on the outgoing side, which means that it too is a positive value. With both of these values positive, the value of the R must also be positive, and this equation holds as it stands. We don't want to make any assumptions about the sign of R for a convex mirror, so for now we will just note that R is positive if the mirror is concave. Or put into the language of our conventions, R is positive if the center of the sphere is on the outgoing side of the light.

name	symbol	positive when	negative when
object distance	s	*object on incoming side*	*object on other side*
image distance	s'	*image on outgoing side*	*image on other side*
lateral magnification	M	*image upright*	*image inverted*
radius of curvature	R	*center of sphere on outgoing side*	???

(4.3.4)

Real and Virtual Images

We arbitrarily chose a position for the object, and found that the image distance was also positive. The value of $\frac{2}{R}$ is a fixed value, but we can move the object as close to the mirror as we like. Then the value of s can be made arbitrarily small (but still positive), which means that $\frac{1}{s}$ can be made arbitrarily large. If this latter value exceeds the value of $\frac{2}{R}$, then for Equation 4.3.3 to hold, the value of s' must be *negative*. This means that the image of an object close to the mirror lands on the side other than the outgoing side of the light, which in this case is to the *right of the mirror*.

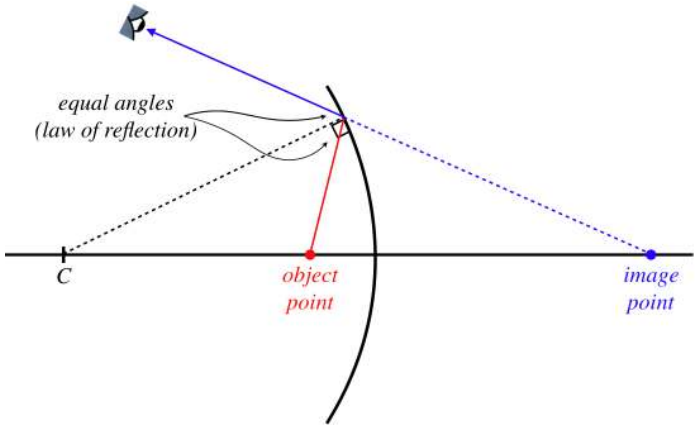
At first this seems impossible – the light never gets behind the mirror, so how can the right rays possibly intersect there to make an image? This calls for an important reminder...

Alert

Images are not formed from light rays running into other light rays! First of all, it is dangerous to use the word "formed" here, because it implies that a new object is created, as though optics works like a 3-d printer. The idea of an image is simply something we invented to express the fact that the apparent source of light is different from the actual source (different in its location, size, and possibly orientation). Second, we locate the image by looking at light and tracing imaginary "rays" (which we also invented) back to their apparent origin. There is nothing in this scheme that requires actual light to intersect with itself. If the backtrack of the rays lands in a region where there is no light (as was the case for the flat reflector), then that is fine.

Let's see how this happens with a new figure. As before, a ray that follows the optical axis bounces straight back from the vertex of the mirror, so the image must lie on the optical axis, but now when the object point is close we have:

Figure 4.3.4 – Image Point Behind the Mirror



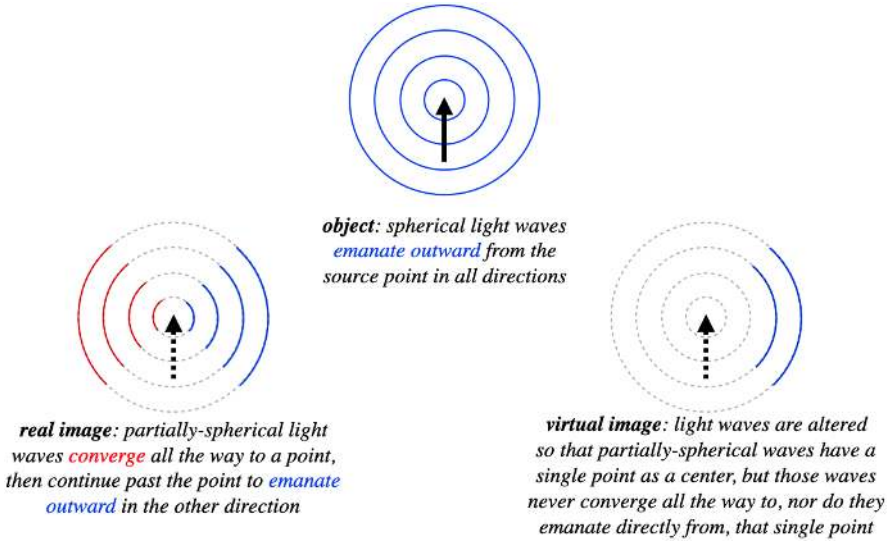
This image is found the same way as the previous image (because that is the *definition* of an image) – by backtracking reflected rays to the point where they intersect. Nevertheless, there seems to be something distinctly different about this image that exists where there is no light, and the previous image that exists where the actual light is. These images don't *look* any different – it's not like one is darker than the other or anything, but there is one small difference about them that we will discuss later. For now, we will simply give them different names. Images that exist at a point where the light actually passes are called *real*, and images that exist where there is not actually any light (at least no light from the object that reaches the eye) are called *virtual*.

Alert

*These monikers of "real" and "virtual" are quite dangerous, as they often lead to great confusion. People encountering these terms tend to think that real images are "actually there," while virtual images are not. The truth is that **neither** is "actually there," whatever that means! Both types of images are simply where we perceive the light to be coming from – there is not actually any light emitted from that spot in either case. Don't let their names fool you into thinking there is really anything fundamentally different about these two types of images.*

To understand the differences between an object, a real image, and a virtual image, it might help to think about the light once again as waves. For a point object (like the tip of our usual arrow), light travels outward in all directions in a full spherical wave. For a real image, a partial spherical wave propagates *inward* toward the point image, and then outward from it on the other side. For a virtual image, a partial spherical wave is moving outward only, with the point image at its center, but the wave does not exist in the immediate vicinity of the point image.

Figure 4.3.5 – Objects and Images in Terms of Light Waves



In every case, an eye viewing the outward-moving spherical light waves will trace the source back to the center of the sphere, which is the "apparent source of the light."

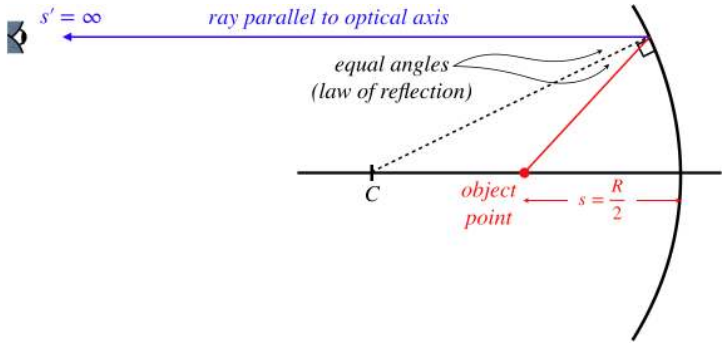
Concave Mirrors – Principal Rays

We can now find the image given the location of the object, either by drawing a diagram and doing the geometry for a single point, or by plugging into the equation. But this doesn't tell us everything about the image. We will now develop the geometry a little further in order to determine whether the image is larger or smaller than the object, and whether it inverts or remains upright.

We can use what we have done above to locate the one point on the image that intersects the optical axis, and to simplify matters we will place the base of the object arrow on this axis. So all we need it to find the location of the tip of the image arrow. The tip sends out spherical waves that are reflected by the mirror, which means we have an infinite number of rays to choose from to perform our geometry tricks. We only need two such rays to get an intersection point, so let's see if we can pick rays that make our job as easy as possible. It turns out that there are four such rays, called the *principal rays*. They are not physically any more important than any of the infinitude of other rays – they just make our task of geometry simpler.

Before we describe how the principal rays are defined, we need to take a quick look back at Equation 4.3.3. We saw that the image distance can be either positive or negative, depending upon the placement of the object. What happens if we start the object far away and gradually move it closer to the mirror. At some point the image must transition from real to virtual. What is the borderline position of the object that defines this transition? Clearly this occurs when $s = \frac{R}{2}$, which results in the image position going to infinity. What does the ray look like that passes through the image point at infinity? In this limit, the outgoing ray must be parallel to the optical axis.

Figure 4.3.6 – Image Point at Infinity Means Ray Parallel to Optical Axis



If we swap the object and image (the law of reflection doesn't care which ray is coming in or going out), so that the object is at $s = \infty$, then the image of this point will be on the optical axis a distance of $\frac{R}{2}$ from the mirror. This is true of *every* ray that strikes the mirror from a direction parallel to the optical axis. That is, every incoming light ray *converges* to this point, and this point is called the *focal point* of the mirror.

The existence of a focal point gives us two principal rays – one that comes in parallel to the optical axis (note it doesn't have to come from infinity – it still follows the same path), which reflects through the focal point, and one that passes through the focal point or comes from the direction of the focal point, which reflects parallel to the optical axis. A third principal ray is one that passes through or comes from the direction of the center point. When such a ray strikes the mirror, it hits it at a right angle, so it comes straight back. And the fourth principal ray is the one that strikes the vertex of the mirror, which reflects at the same angle with the optical axis from which it arrived. Again, these rays are "principal" only because they are easy to sketch and work with geometrically, and only two of them are needed to find the image location, though all four intersect at the same point. To summarize:

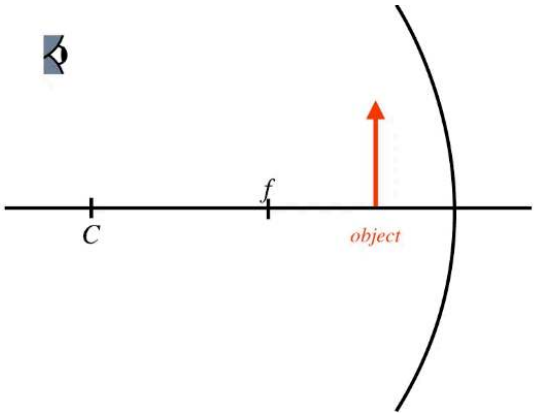
Figure 4.3.7 – Principal Rays of a Concave Spherical Mirror

What we notice immediately is that we have our first inverted image thus far in our exploration of geometrical optics. It is also diminished in size. Both of these results are specific to the placement of the object, so one must be careful about drawing a general conclusion from this. This process of sketching the principal rays is called a *ray trace*, and while it is not particularly useful for quantitative work (unless, for example, you work carefully on graph paper), it does quickly give an idea of the rough position of the image, a rough measure of the lateral magnification, and the image's orientation. Shortly we will examine how to get this information through direct mathematics.

We can see the importance of the placement of the object in the final result by performing the ray trace with the object closer to the mirror than the focal point. The principal ray that comes in parallel to the axis and the ray that strikes the vertex are as easy to do as the case shown above, but what about the ray that passes through the focal point and the ray that passes through the center point? If the light from the object has to go backward to go through these points, then it will not strike the mirror at all. Do we simply have to concede that we only have two principal rays to work with? No!

The rays only represent the direction that a single point on the light wavefront is moving, so their value as a tool comes from their *directions*, not from there actual paths. A ray that comes to the mirror *from the direction* of the center point or focal point will behave after reflection exactly like a ray that does pass through those points. So all we need to do to trace these rays is start them at C or f (depending upon which ray you want), and send it through the tip of the object toward the mirror. [*Note: The part of the ray where there is not actually any light is often represented with a dotted line.*]

Figure 4.3.8 – Principal Rays from an Object Close to a Concave Spherical Mirror



Notice that the outgoing rays don't cross each other as they did in the previous example. Again, it is only the *backtracking* of rays that has any meaning in our use of rays. We simply have to backtrack these rays a bit farther than before, into a region where there is no light passing through. Put another way, the previous example resulted in a real image, while this example results in a virtual image. Although the other two principal rays were omitted in the above figure, if they had been included, then they would also be backtracked to the same point in space.

Alert

The curvature of the mirror has been exaggerated for effect in the figure, and a careful ray trace should actually treat it as flat. If you try to confirm that the other principal rays converge to the same point and don't correct this flaw, you will find that they will not quite converge.

Concave Mirrors – Mathematics

The focal point is a distance of $\frac{R}{2}$ from the mirror, a distance known as the *focal length* of that mirror, designated the variable f (we stopped discussing waves just in time to be able to reuse this symbol without confusing it with frequency). In terms of the focal length, Equation 4.3.3 becomes:

$$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f} \tag{4.3.5}$$

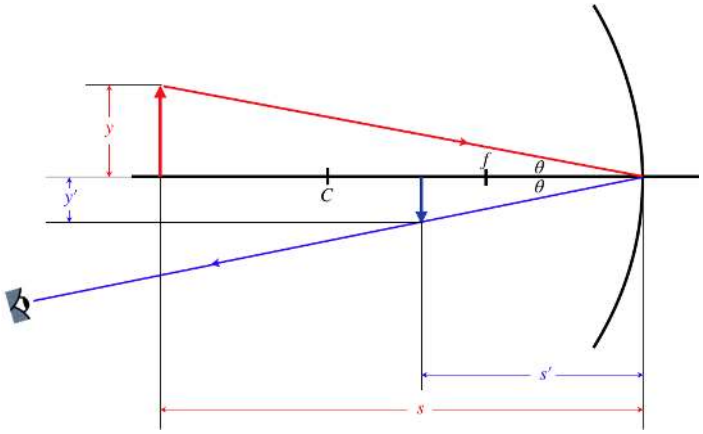
So far we have only discussed concave mirrors, and we know in this context that our sign convention requires $R>0$, which means that the same must be true for the focal length, which we now add to our list:

name	symbol	positive when	negative when
$object\ distance$	s	$object\ on\ incoming\ side$	$object\ on\ other\ side$
$image\ distance$	s'	$image\ on\ outgoing\ side$	$image\ on\ other\ side$
$lateral\ magnification$	M	$image\ upright$	$image\ inverted$
$radius\ of\ curvature$	R	$center\ of\ sphere\ on\ outgoing\ side$	???
$focal\ length$	f	$focal\ point\ on\ outgoing\ side$	???

(4.3.6)

We already know how to compute the image distance s' from the object distance and radius of curvature. Let's look at how we can calculate the lateral magnification from these inputs. We know the lateral magnification in terms of the heights of the object and image, and with just one of the principal rays (the 4^{th} one, that strikes the vertex), we can accomplish our goal.

Figure 4.3.9 – Geometry of Lateral Magnification



The triangles formed by the object and image are similar (the reflection angles are equal, and they are both right triangles), which means the ratios of the lengths of their sides are equal. Using this and noting that y is positive, y' is negative, and both s and s' are positive (according to our sign conventions), we have:

$$M = \frac{y'}{y} = -\frac{s'}{s} \tag{4.3.7}$$

This can then be combined with Equation 4.3.3 to determine the lateral magnification in terms of the placement of the object and the radius of curvature of the sphere. While this derivation of the lateral magnification was performed for the case of the real image, the same result comes out for the virtual image represented by Figure 4.3.8, and the method is exactly the same – similar triangles are formed using the ray that strikes the vertex, and the ratios are the same as here. Interestingly, because the image is not on the outgoing side of the mirror in this second case, then according to our sign conventions $s' < 0$, which makes the lateral magnification positive, and indeed the image is upright, as it should be for a positive lateral magnification!

Example 4.3.1

An object is moved away from a concave mirror to a position that is twice as far from the reflecting surface. In the process, the size of the image seen in the reflection goes down by a factor of 3. Find distance that separated the object and the surface before it was moved, measured in terms of the radius of the spherical reflector.

Solution

Using Equation 4.3.3, we can solve for the magnification in terms of the object distance and radius of curvature of the sphere:

$$\left.\begin{aligned} M &= -\frac{s'}{s} \\ \frac{1}{s'} &= \frac{2}{R} - \frac{1}{s} \end{aligned}\right\} \quad M = \frac{R}{R-2s}$$

For the object at two different positions, there are two different magnifications. The object size never changes, so for the image to get smaller by a factor of 3, the magnification must get smaller by a factor of 3. Calling M_1 the magnification at the initial object distance s and M_2 the magnification at the new object distance $2s$, we have:

$$\frac{1}{3} = \frac{M_2}{M_1} = \frac{\frac{R}{R-4s}}{\frac{R}{R-2s}} = \frac{R-2s}{R-4s} \quad \Rightarrow \quad s = R$$

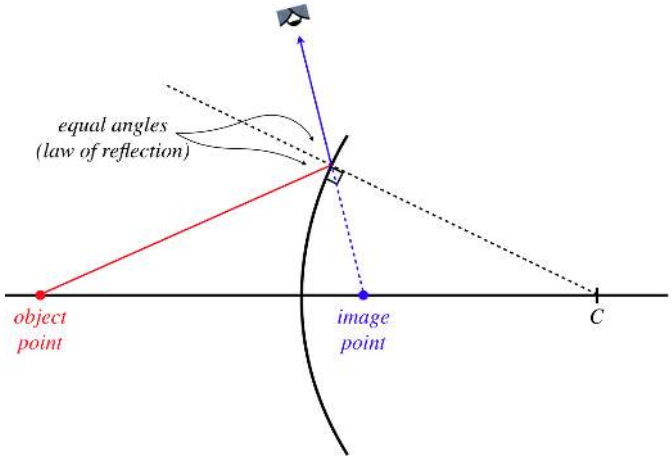
Convex Mirrors – Object and Image Distances

As much work as we have done above, it only covers one of the two varieties of spherical reflector. Fortunately, we will not need to repeat every single step above for convex mirrors, because the principles behind the geometrical optics involved are the same. The only difference we have to address is the way that light behaves when it is reflected off a convex surface. After that, we will see that the ray traces are fairly straightforward (assuming we fully understand the ones we have done already), and the mathematics is even easier to extend to the convex case.

The single physical principle that must be satisfied is the law of reflection. When a ray strikes a convex spherical reflector, the incoming angle and outgoing angle are measured relative to the line that emanates outward from the center of the sphere. As before, if we place our

object point on the optical axis, the image point must also reside on that axis, so we can draw a basic picture of what the relative positions of the object and image should look like, as we did for the concave case in Figure 4.3.2.

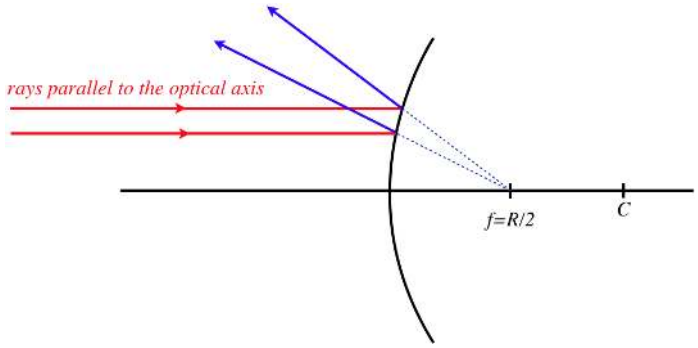
Figure 4.3.10 – Image of a Point on the Optical Axis for Convex Reflector – Picture



The first thing we notice for this case upon closer examination is that the image comes out to be virtual *no matter where the object is located*. Using our sign conventions, we have a positive object distance (the object is on the incoming side of the mirror), and a negative image distance (the image is not on the outgoing side of the mirror). If we want to use an equation that looks anything like Equation 4.3.3, then consider when the object is very far (read "infinitely far") away... The radius of curvature has to be *negative*. We will incorporate this into our sign conventions shortly.

As we determined previously, an object infinitely far away produces rays that are parallel to the optical axis. In the concave case, such rays converged to the focal point of the mirror. In this case, the rays all *diverge* from the focal point, which in this case resides on the opposite side of the mirror. The math works out the same as before regarding the focal length's relation to the radius of curvature, namely: $f = \frac{R}{2}$.

Figure 4.3.11 – Rays Parallel to Optical Axis Diverge from Focal Point After Reflection



While we will not repeat the geometry here, it conveniently works out that all the math we developed before, including Equation 4.3.3 can be used unchanged for the convex mirror, provided we now treat the radius and focal length as *negative* numbers. That is, we can now fill-in the unknown parts of the sign conventions:

name	symbol	positive when	negative when
$object\ distance$	s	$object\ on\ incoming\ side$	$object\ on\ other\ side$
$image\ distance$	s'	$image\ on\ outgoing\ side$	$image\ on\ other\ side$
$lateral\ magnification$	M	$image\ upright$	$image\ inverted$
$radius\ of\ curvature$	R	$center\ of\ sphere\ on\ outgoing\ side$	$center\ of\ sphere\ on\ other\ side$
$focal\ length$	f	$focal\ point\ on\ outgoing\ side$	$focal\ point\ on\ other\ side$

(4.3.8)

Convex Mirrors – Principal Rays

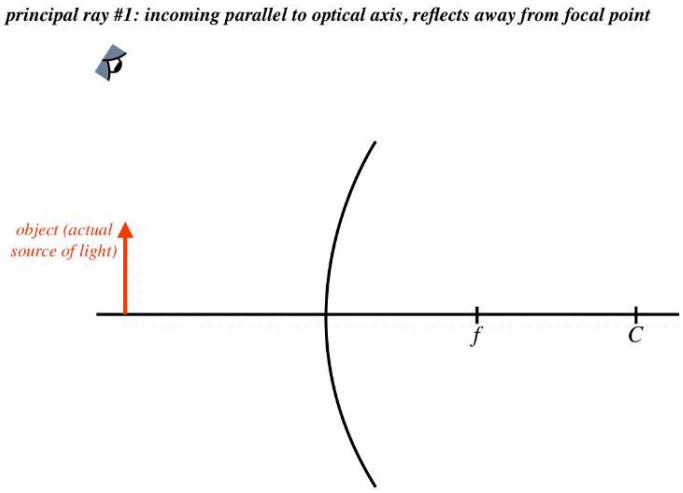
When it comes to ray traces for convex spherical mirrors, we follow the same four methods as for the concave mirror, though we have to keep in mind that all images are going to turn out to be virtual, so we can never expect outgoing rays to intersect – we always have to backtrack behind the mirror to find the intersection point.

The ray that strikes the vertex of the mirror (which we called "principal ray #4" in Figure 4.3.6 above) is sketched exactly as before, but there are subtle differences for the other three principal rays.

- Principal ray #1 comes in parallel to the optical axis, and therefore reflects *away* from the focal point, rather that toward it as it did in the concave case.
- Principal ray #2 can't pass through the focal point as it does for the concave case, but it heads *toward* it. It then reflects back parallel to the optical axis.
- Principal ray #3 can't pass through the center point as it does for the concave case, but it heads *toward* it. This causes it to strike the surface at a right angle, so it bounces straight back.

As before, all four of these rays (to a very good approximation, if the mirror can be treated as fairly flat) appear to an observer to emanate from a common point behind the mirror, which is the position of the virtual image.

Figure 4.3.12 – Principal Rays of a Convex Mirror



We can double-check this result with the math of Equation 4.3.3, to see that everything remains consistent. We already said that a negative value of R (and f) and a positive value of s requires that the value of s' be negative, confirming that the image must be on the dark side of the mirror. In addition, for the sum of $\frac{1}{s}$ and $\frac{1}{s'}$ for come out negative, the *absolute value* of s' must be smaller than that of s. This means that the image is always closer to the mirror than the object. We can see that this is true for the ray trace, and that in fact it must be true no matter where the object is placed.

Another check we can do is the lateral magnification. The ray trace shows that the image is diminished and upright. Given that the math shows that s' is always negative and always has a smaller absolute value than s, this lateral magnification result is confirmed using Equation 4.3.7.

The effect of a diminished image is that a greater field of view is possible in the space occupied by the mirror. This explains why car side mirrors are convex – the greater field of view allows the driver to see blind spots. The drawback is that when the overall field of view is increased, the percentage of that field occupied by a single object (such as another car) goes down. We are accustomed to interpreting objects occupying small fractions of the field of view as being far away, which is why these mirrors often have the warning, "Objects in mirror are closer than they appear."

Example 4.3.2

A spherical shell is reflective on both sides. When the reflection of an object is viewed in the convex side, the image is 40% of the size of the object. If the shell is now turned around so that the reflection is viewed in the concave side, determine the size of the image (compared to the object), and whether the image is upright or inverted. Assume that the distance between the shell and object are unchanged after the shell is rotated.

Solution

In terms of the focal length of the reflector, the magnification is:

$$\left.\begin{aligned} M &= -\frac{s'}{s} \\ \frac{1}{s'} &= \frac{1}{f} - \frac{1}{s} \end{aligned}\right\} \quad M = \frac{f}{f-s}$$

We are given that the magnification initially is $\frac{2}{5}$, so solving for s in terms of f gives:

$$\frac{2}{5} = \frac{f}{f-s} \quad \Rightarrow \quad s = -\frac{3}{2}f$$

The negative sign comes in because the focal length for a convex reflector is negative, while the object distance is positive. When the reflector is turned around, the focal length becomes positive, and since the distance between the object and mirror is unchanged, we now have that the object distance is: $s = +\frac{3}{2}f$. Plugging this into the magnification gives:

$$M = \frac{f}{f-\frac{3}{2}f} = -2$$

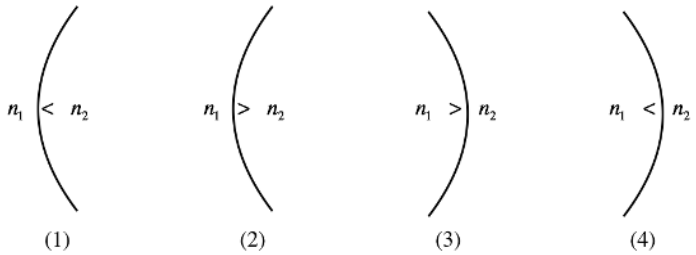
So the new image is twice as large as the object, and because the value of the magnification is negative, the image is inverted.

4.4: Spherical Refractors

Spherical Surface Refractions

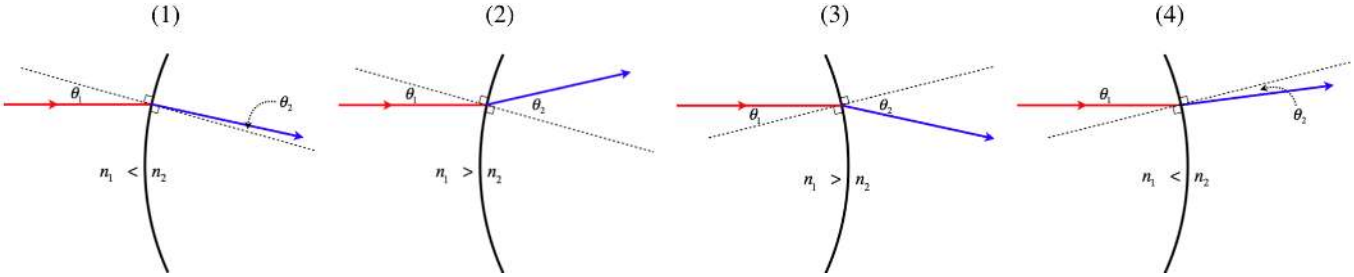
In Section 4.1, when we completed plane reflectors, we moved on to plane refractors, and having done spherical reflections, we now follow the same progression. There were two possibilities for reflectors – concave and convex. While the same two shapes are available for refractors, there are actually *four* distinct possibilities, since for each of the two shapes, the transition of media can be either from slower medium to faster, or vice-versa. If we assume light is moving left to right in each of the cases, the diagram below lists all four possibilities.

Figure 4.4.1 – Four Varieties of Spherical Refraction



Before diving into finding images for these cases, let's take a moment to get some sense of what happens to a ray that encounters these boundaries in each case. In particular, we will be looking to see if the ray bends toward or away from the optical axis as it crosses the boundary between the two media. Central to this discussion is the law of refraction, the conceptual basis of which is that a light ray entering a slower medium bends toward the perpendicular to the surface, while a ray that enters a faster medium bends away from this perpendicular.

Figure 4.4.2 – Ray Convergence and Divergence for the Four Cases of Spherical Refraction



It is important to note that neither the shape of the border, nor change in the speed of the light, determines by itself the convergence or divergence of rays. Only the combination of both of these factors makes this determination.

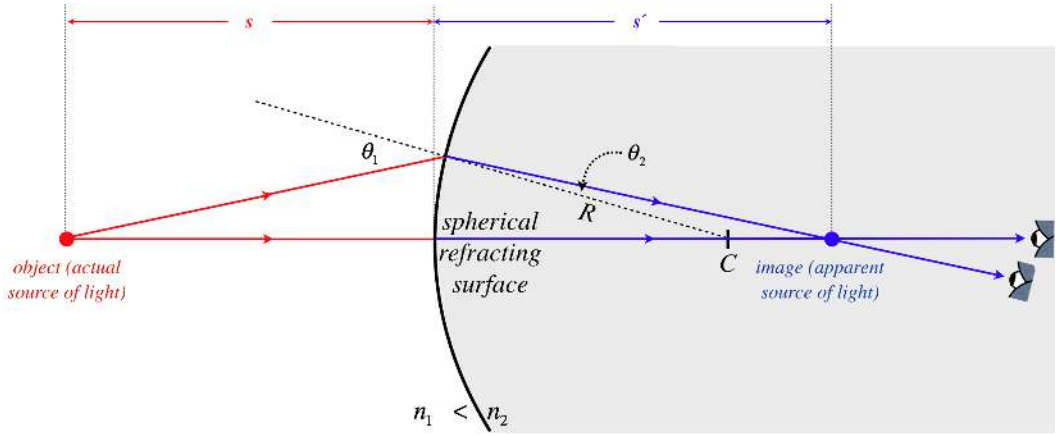
Object and Image Distances

As with the case of reflectors, it is clear that the image of an object point on the optical axis is also on the optical axis, because a ray along the optical axis will strike the boundary at a right angle and will pass straight through without bending. Our task is therefore to determine how far the image is from the vertex, given the distance of the object from the vertex. In the diagrams that follow, we will maintain the following conventions:

- The object is always to the left of the boundary, and the observer to the right, viewing the light after it passes through.
- The shaded region is the one with the *higher* index of refraction. This means the light travels slower there, and the angle on that side of the boundary is always the smaller one.

In each case, we will look at two rays – one that is along the optical axis, and the other that strikes the boundary off-axis. Backtracking the second ray to the optical axis will locate the image point. We will diagram each of the four cases shown in the figure above...

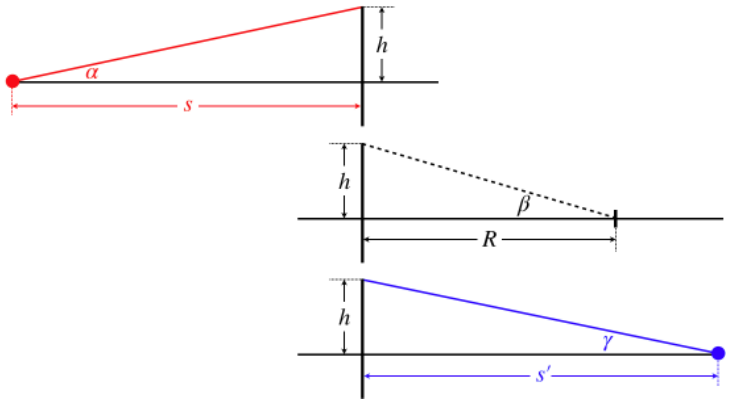
Figure 4.4.3 – Image of a Point on the Optical Axis – Picture (Case 1)



Before we go on, it is important to note that for this case, we have assumed that the object is not too close to the refracting surface. Imagine what happens if we move it closer: The angle θ_1 grows, which causes angle θ_2 to grow, according to Snell's law. But if θ_2 grows too much, then the outgoing ray may not converge to the axis on its way to the eye. If we move the object close enough to the refracting surface, then the image ends up on the left side of the surface. The light rays don't actually cross the axis there, but the eye sees the light coming from that point.

As with the case of the spherical reflector, we need to make the approximation that the spherical surface is flat. The geometry here is significantly more involved than it was for the reflector, because in the reflection case, the two angles at the surface were equal, while here they are related by Snell's law (Equation 3.6.4). The simplest approach to this geometry is to first reduce the clutter by separating the three right triangles involved and naming all the relevant angles...

Figure 4.4.4a – Image of a Point on the Optical Axis – Geometry (Case 1)

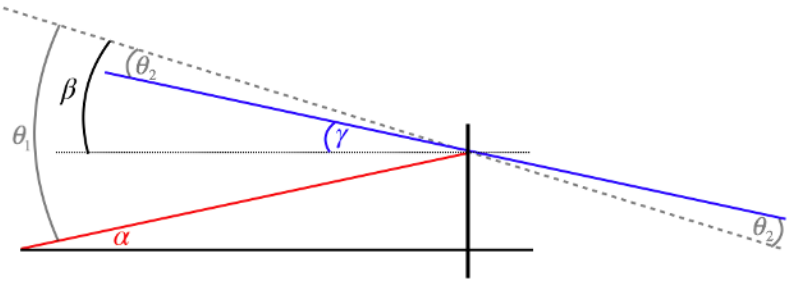


We can now relate these angles to the sides of the triangles, using our usual small-angle approximation:

$$\begin{aligned} \alpha \approx \tan\alpha &= \frac{h}{s} \\ \beta \approx \tan\beta &= \frac{h}{R} \\ \gamma \approx \tan\gamma &= \frac{h}{s'} \end{aligned} \tag{4.4.1}$$

Though the pictures are different, so far this is actually no different from what we obtained in Equations 4.3.2. The difference comes in how we relate these angles to each other. In the case of the spherical reflector, the law of reflection gave us the simple relation of Equation 4.3.1. Here we have to employ Snell's law to obtain a link between these angles. To see how these angles relate to each other, one more diagram is called for, this one focusing on the point at the boundary where the ray passes through.

Figure 4.4.4b – Image of a Point on the Optical Axis – Geometry (Case 1)



This diagram includes the angles from the three triangles, as well as the two refraction angles, θ_1 and θ_2. We can now do the geometry necessary to relate them to each other, and combine them with the triangle equations above:

$$\theta_1 = \alpha + \beta = \frac{h}{s} + \frac{h}{R}\,, \qquad \theta_2 = \beta - \gamma = \frac{h}{R} - \frac{h}{s'} \tag{4.4.2}$$

We also have Snell's law relating θ_1 and θ_2, along with the small angle approximation $\sin\theta \approx \theta$:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \quad \Rightarrow \quad n_1\theta_1 \approx n_2\theta_2 \tag{4.4.3}$$

Plugging this in above gives a result reminiscent of Equation 4.3.3, with the difference coming from the two indices of refraction:

$$\left.\begin{aligned} n_1\theta_1 &= n_1\left(\frac{h}{s} + \frac{h}{R}\right) \\ n_2\theta_2 &= n_2\left(\frac{h}{R} - \frac{h}{s'}\right) \\ n_1\theta_1 &= n_2\theta_2 \end{aligned}\right\} \quad \frac{n_1}{s} + \frac{n_2}{s'} = \frac{n_2 - n_1}{R} \tag{4.4.4}$$

Although this is different from the result of the reflector, we can still define a focal length. As before, we do this by solving for the image distance of an object at an infinite distance:

$$\frac{n_1}{\infty} + \frac{n_2}{f} = \frac{n_2 - n_1}{R} \quad \Rightarrow \quad f = \frac{n_2}{n_2 - n_1}R \tag{4.4.5}$$

Notice that while the focal length of a reflector is shorter than the radius of the mirror, for a refractor it is longer. So Equation 4.4.4 can be written in terms of the focal length as:

$$\frac{n_1}{s} + \frac{n_2}{s'} = \frac{n_2}{f} \tag{4.4.6}$$

The ray traces of the other three cases look different from case 1:

Figure 4.4.5 – Image of a Point on the Optical Axis – Picture (Case 2)

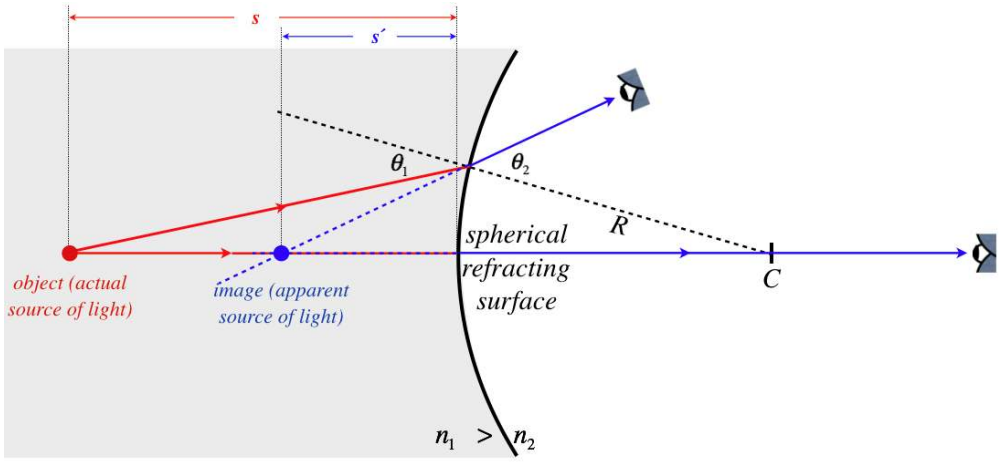


Figure 4.4.6 – Image of a Point on the Optical Axis – Picture (Case 3)

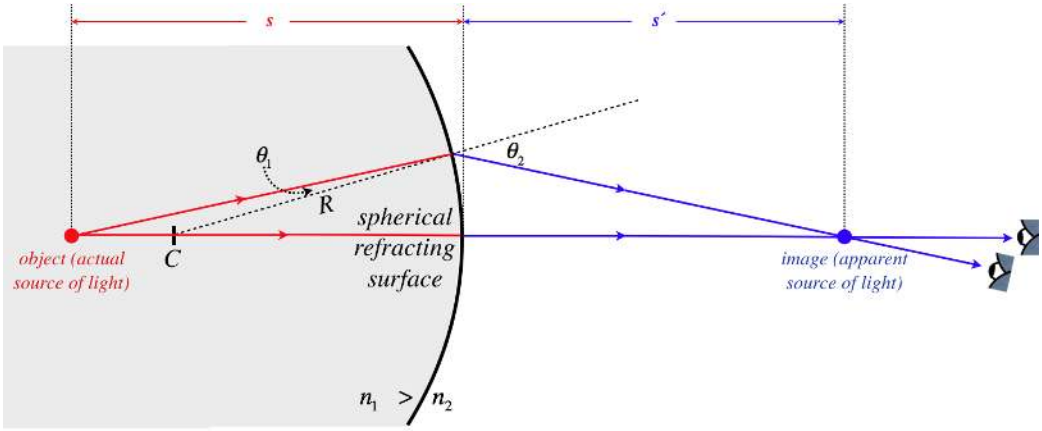
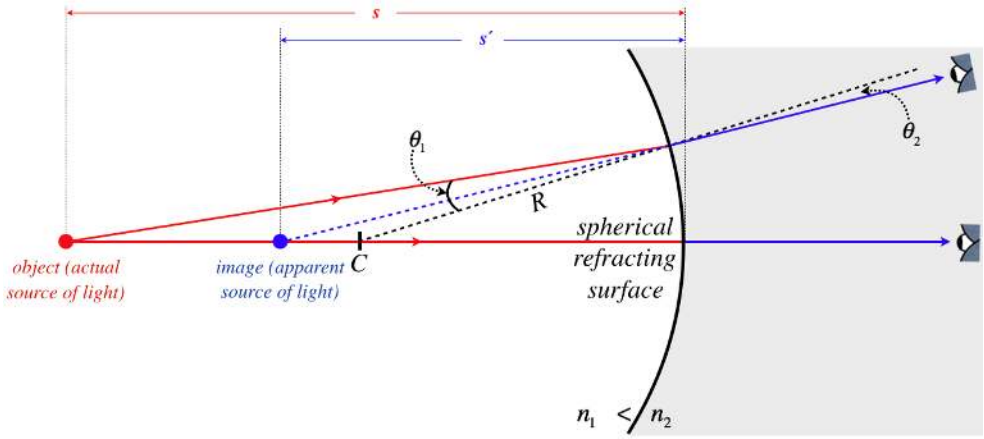


Figure 4.4.7 – Image of a Point on the Optical Axis – Picture (Case 4)



We will not re-work the geometry for cases 2 though 4, but rather will state without proof that the same relationship between the object distance, image distance, and radius of curvature holds, *provided the sign conventions we have established are maintained*.

Checking Sign Conventions

Let's do a quick check to see if the diagrams above match what we expect from the formula and sign conventions.

Case 1:

As we saw with the reflector, whenever a surface causes rays to converge, the image can either be real or virtual, depending upon the object distance. If the object is closer to the surface than the focal point ($s < f$), the image is virtual ($s' < 0$), and if it is farther from the surface than the focal point ($s > f$), the image is real ($s' > 0$). In Figure 4.4.3, the object distance is greater than the focal length. We know this because the ray shown converges down to the optical axis. If it was at the focal point, the ray would converge to a line parallel to the optical axis, and if it was inside the focal point, it would not even converge as far as the parallel.

- The object is on the incoming side of the surface and is farther from the surface than the focal point: $s > 0$, $s > f$
- The center of the sphere is on the outgoing side: $R > 0$
- The light is going from faster medium to slower medium: $n_2 > n_1$, $f = \dfrac{n_2}{n_2 - n_1} R > 0$
- Rearranging Equation 4.4.6, we can show that the image distance must be positive, which agrees with the diagram that shows the image to be on the outgoing side of the surface:

$$\frac{n_2}{s'} = \frac{n_2}{f} - \frac{n_1}{s} \tag{4.4.7}$$

With $s > f > 0$ and $n_2 > n_1$, the first term on the right hand side of this equation must be greater than the second term, making the difference (and s') positive.

Example 4.4.1

Figure 4.4.3 shows the image to be farther from the surface than the center of the sphere. For the conditions given for that diagram, confirm mathematically that this must be true.

Solution

Rearranging Equation 4.4.4, we have:

$$\frac{n_1}{s}+\frac{n_1}{R}=\frac{n_2}{R}-\frac{n_2}{s'} \tag{4.4.8}$$

The left hand side of this equation is positive, so on the right hand side the second term must be smaller than the first term, which means the second term's denominator must be larger than the first's: $s' > R$.

Case 2:

In this case, the surface causes the rays to diverge, so the nature of the image is not affected by the magnitude of the object distance.

- The object is on the incoming side of surface: $s > 0$
- The center of the sphere is on the outgoing side: $R > 0$
- The light is going from slower medium to faster medium, which means that the focal length and radius have opposite signs (i.e. the focal point is on the opposite side of the surface as the center point): $n_2 < n_1$, $f=\frac{n_2}{n_2-n_1}R<0$
- Rearranging Equation 4.4.6, we can show that the image distance must be negative, which agrees with the diagram that shows the image to not be on the outgoing side of the surface:

$$\frac{n_1}{s}=\frac{n_2}{f}-\frac{n_2}{s'} \tag{4.4.9}$$

The left hand side of this equation is positive, and the first term on the right hand side is negative. This means that s' must be negative.

Example 4.4.2

An object is viewed through a spherical refracting surface like that in case 2, with $n_1 = 1.2$ and $n_2 = 1.0$. Can the image be seen at a distance from the surface equal to the radius of the sphere? If so, determine the distance that the object must be from the surface (in terms of the radius of the sphere) for this to happen.

Solution

For this case, the $s' < 0$ and $R > 0$, so the condition we are asked to check is: $s' = -R$. Plugging this into Equation 4.4.4, we have:

$$\frac{n_1}{s}+\frac{n_2}{-R}=\frac{n_2-n_1}{R} \quad \Rightarrow \quad \frac{n_1}{s}=\frac{2n_2-n_1}{R}$$

So we see that this is only a possibility when $2n_2 > n_1$, which happens to be the case here. Plugging in the values for n_1 and n_2, we find that the object must be a distance 1.25 times the radius of the sphere from the surface.

Cases 3 and 4:

The remaining cases follow similarly with the first two. Case 3 involves converging rays ($f > 0$) like Case 1, which comes about because $R < 0$ (center is not on the outgoing side), and $n_2 - n_1 < 0$ (light going from slower medium to faster one). It should be noted that Case 3, like Case 1, includes a possibility that is not depicted in its associated figure. Figure 4.4.6 assumes that the object point is sufficiently far from the refracting surface (making θ_1 sufficiently large) that Snell's law will result in the ray being bent back down to the axis. The fact that a second figure was not drawn for either Case 1 or Case 3 should not confuse the reader into thinking that having the rays converge back to the axis is the only possibility, or that the mathematical result is any different in such situations (the sign conventions make it all work out!).

Case 4 involves diverging rays ($f < 0$) like case 2 because $R < 0$ and $n_2 - n_1 > 0$.

Lateral Magnification

We do not need to go through all the principal rays to derive the lateral magnification relation for the spherical refractor. We can simply note that a ray coming from the tip of an object arrow that approaches the refractor parallel to the optical axis is a distance y from the optical axis. Then after hitting the surface, it either converges toward or away from the focal point. The tip of the image arrow then lies on this angled part of the ray, a distance of $\pm y'$ from the axis. These two heights form similar right triangles. The right triangle associated with the object has legs of length y and f, while the right triangle associated with the image has legs of length y' and $s' - f$. Setting the ratios of the lengths of their legs equal, we have an expression for the lateral magnification:

$$M = \frac{y'}{y} = -\frac{s' - f}{f} \tag{4.4.10}$$

From Equation 4.4.6, we have:

$$\frac{n_1}{s} = n_2\left(\frac{1}{f} - \frac{1}{s'}\right) \quad \Rightarrow \quad \frac{n_1}{n_2 s} = \frac{s' - f}{f s'} \quad \Rightarrow \quad M = -\frac{n_1}{n_2}\frac{s'}{s} \tag{4.4.11}$$

Comparing this lateral magnification with that of the spherical reflector, we see that a factor of $\frac{n_1}{n_2}$ has been introduced. It seems clear that a greater difference in the two indices of refraction will cause the light to refract more, which should lead to more magnification. Comparing with the result from the plane refractor (Equation 4.1.6), we see that the lateral magnification is confirmed to be +1.

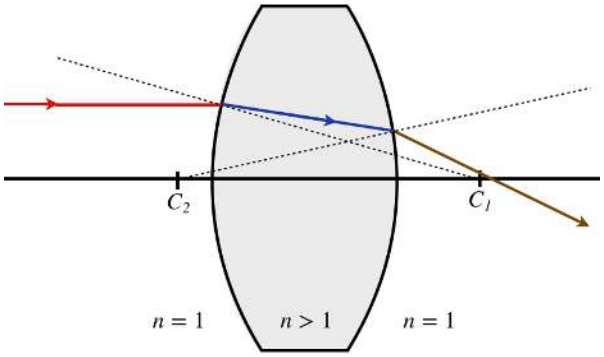
4.5: Thin Lenses

Building a Lens

While it is interesting to see how images are formed as light passes from one index of refraction to another through a spherical surface, this only rarely happens in the real world. It is far more common for light to come from an object that is in the air ($n = 1.00$), pass through a transparent region with a higher index of refraction, then return again to the air before being observed. The light will then be refracted at both surfaces of the region during its journey from the object to the point of observation. If both surfaces of the region are spherical, then the light will either converge or diverge at each surface (depending upon which of the four cases of Section 4.4 applies). Such an optical device is called a *lens*.

Let's take a look at a glass lens that is convex on both sides, which is in air. When light enters one side, it is going from a lower index of refraction (air) to a higher one (glass) that is convex, which means we have case #1, which is a converging refraction. Now the light is inside the glass, and its next encounter is with a *concave* surface (from the perspective of inside the glass), passing from a higher index of refraction to a lower one. This is case #3, which is also a converging refraction. [*See Figure 4.4.2 for references to the four possible cases.*] A double-convex lens is therefore a converging lens.

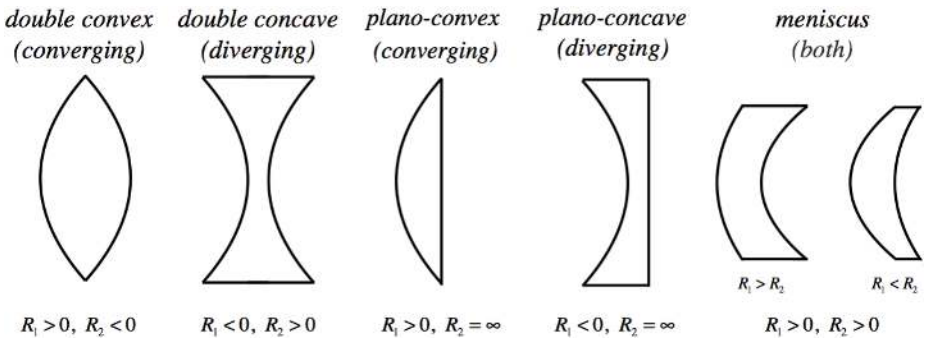
Figure 4.5.1 – Double Convex Lens



Note that the two surfaces do not necessarily have the same radius of curvature. The surface with the shorter radius (the one that is more sharply-curved) will bend the light more than the other surface.

We can in fact put together any combination of surfaces we like in the construction of a lens. A double concave lens will result in the light experiencing cases 2 and 4, both of which are diverging, causing the light rays that pass through the lens to diverge. The various combinations of surfaces is shown in the diagram below. Note that the signs of the radii given are found using the sign convention and assuming that the light is passing from left to right, and that a flat plane can be described as a spherical surface with an infinite radius.

Figure 4.5.2 – Varieties of Lenses



Tracing a ray through a lens is a daunting bit of geometry, and as always, when a calculation is daunting, we make a simplifying approximation. In this case, the simplifying principle is called the *thin lens approximation*. This asserts, not surprisingly, that the lens is very thin, which means that the bending of the light that occurs at each surface essentially happens at the same position. The two refractions are still successive (a refraction occurs at one surface then the other), but the light travels no distance getting from the point where it changes direction the first time to where it changes direction a second time. This simplifies things greatly,

because measurements of quantities like object and image distances are all referenced to the same place, no matter which surface is in play.

Defining a Single Focal Length for a Lens

We already know that each surface of a lens has its own focal length, but when light that comes upon the lens parallel to the optical axis, it must go somewhere after it passes through, which means that the lens as a whole must have a unique focal length. We now seek what this will be for a lens of index of refraction n in air with radii of curvature of its two sides equal to R_1 and R_2.

The key trick to use here is that after light passes through the first surface, it appears to be coming from somewhere else (i.e. from the position of its image). The second surface then receives the light, and as far as it "knows," the light is coming from this new position. That means that the object position for the second surface is the image position of the first surface. All that remains is to do two successive image locations, while getting all the sign conventions right, and the final image position will be what is seen as the apparent origin of light that has passed through both surfaces of the lens.

Okay, so if we are looking for the focal length of the lens, we naturally start with a ray that comes into the lens parallel, and look for the distance from the lens that it crosses the axis. A parallel ray is equivalent to an object distance of infinity ($s_1 = \infty$), the first medium is air ($n_1 = 1$), and the second medium is that of the lens ($n_2 = n$), so the image for the first surface (with radius R_1) is easy enough to find using Equation 4.4.4:

$$\frac{1}{\infty} + \frac{n}{s_1'} = \frac{n-1}{R_1} \quad \Rightarrow \quad \frac{n}{s_1'} = \frac{n-1}{R_1} \tag{4.5.1}$$

The image position for this first surface now becomes the object position for the second surface. Because of our thin lens approximation, the magnitude of the image distance for the first surface equals the magnitude of the object distance of the second surface, but now the issue of the signs must be addressed.

Suppose the first surface is concave. Then this surface is case 4, and the ray is made to diverge at this surface, resulting in an image that is to the left of the lens (we are still assuming the light is moving left-to-right). With this image to the left of the second surface and the light moving left-to-right, the object distance for the second surface is positive (on the incoming side). But the image distance calculated was negative (the divergence caused a virtual image), so while the image distance s_1' for the first surface and the object distance for the second surface s_2 are equal in magnitude, they have opposite signs. We can therefore make the substitution $s_2 = -s_1'$ for the equation at the second surface. Note that this time the light is passing from a region with index of refraction n to air:

$$\frac{1-n}{R_2} = \frac{n}{s_2} + \frac{1}{s_2'} = \frac{n}{-s_1'} + \frac{1}{s_2'} \tag{4.5.2}$$

Plugging in the result of Equation 4.5.1 and noting that the final image position is the focal point of the whole lens gives:

$$\frac{1-n}{R_2} = -\frac{n-1}{R_1} + \frac{1}{s_2'} \quad \Rightarrow \quad \frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{4.5.3}$$

This gives us a prescription for building a lens – given the index of refraction of the material, we can grind the two spherical surfaces to radii of R_1 and R_2 to achieve a focal length given by f. For this reason, this is called the *lensmaker equation*.

The reader may be troubled that we assumed that the first surface is concave. Does a convex first surface result in a different lensmaker equation? The answer is no! If the first surface is convex, then the image of the incoming parallel light lands to the right of the lens. This is a positive-valued image distance: $s_1' > 0$. Making this image the object for the second surface gets confusing. With the image already to the right of the second surface, does the light even pass through the second surface? Does it turn around and go right-to-left? The answer is neither of these. While the idea may be difficult to visualize conceptually, mathematically the answer is simple – the light is still going left-to-right, and the object is *not on the incoming side*, so the object distance is *negative*: $s_2 < 0$. This is sometimes referred to as a *virtual object*, and it can only occur when light is diverted (reflected or refracted) twice, so that the image of the first diversion lands beyond the surface that causes of the second diversion. Notice that with $s_1' > 0$ and $s_2 < 0$, and their magnitudes equal, we once again get the $s_2 = -s_1'$ relation, and the same lensmaker equation results.

You may have noticed that the term "$\frac{1}{f}$" comes up a lot. It's clear that the smaller the value of f is, the more sharply the light is bent by the refraction, which means that the larger $\frac{1}{f}$ is, the more "focusing power" the device has. Frequently the focusing power

of a lens is measured with this inverse quantity rather than focal length. This quantity has the units of m^{-1}, which are designated their own name of *diopters* (D).

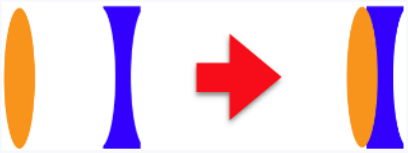
Stacking Thin Lenses

With our assumption that lenses are very thin, it's fair to also assume that if two lenses are placed together, the combination is also very thin. With the ability to refract through one lens, and then immediately the other, parallel rays that come into this combination of two lenses will focus at a different point than either of the lenses individually. In essence, two (or more) stacked lenses simply form the equivalent of a new single lens. If we follow the same process as above to determine the focal length of multiple lenses, we get sum where every term looks something like $\pm\left(\frac{n-1}{R}\right)$ for each refracting surface the light passes through. Rather than break down each lens into its two surfaces, however, we have a shortcut – just add the inverses of the individual lens focal lengths!

$$\frac{1}{f_{tot}} = \frac{1}{f_1} + \frac{1}{f_2} + \ldots \tag{4.5.4}$$

Example 4.5.1

You have a machine that grinds glass surfaces to a certain specified radius. You use it to make a double-convex lens out of a special glass with high index of refraction equal to 2.75. You then use it to make a double-concave lens out of a different type of glass with index of refraction equal to 1.32. Because the curved surfaces are the same radii, the two lenses fit together perfectly. When this is done, it forms a single lens that is half-convex, half concave, as in the diagram below.



a. *If you wanted to grind a third lens with the same machine that had the same optical properties of the combined lens above, but instead use a single type of glass, then would it be double-convex or double concave? Explain.*
b. *Compute the index of refraction needed to accomplish the task described in (a).*

Solution

a. The radius used everywhere is the same, so we will call it "R." The lens maker equation for these two lenses gives:

$$\frac{1}{f_{convex}} = (n_{convex}-1)\left(\frac{1}{R} - \frac{1}{-R}\right) = 2\left(\frac{n_{convex}-1}{R}\right)$$
$$\frac{1}{f_{concave}} = (n_{concave}-1)\left(\frac{1}{-R} - \frac{1}{R}\right) = -2\left(\frac{n_{concave}-1}{R}\right)$$

When the two lenses are pushed together, the diopter strengths (inverses of the focal lengths) are added to get the new diopter strength, so since the index of refraction is greater for the converging lens, from the equations above we see that its diopter strength has a greater magnitude than its diverging counterpart, and the sum will come out positive. The combined lens will converge light.

b. The combined diopter strength is the sum of the diopter strengths, so we want to make a lens with a diopter strength equal to that sum.

$$\frac{1}{f_{tot}} = \frac{1}{f_{convex}} + \frac{1}{f_{concave}} = 2\left(\frac{n_{convex}-1}{R}\right) - 2\left(\frac{n_{concave}-1}{R}\right) = 2\left(\frac{n_{convex}-n_{concave}}{R}\right)$$

Setting this equal to the diopter strength of a single lens with the same radius and unknown index of refraction, we have:

$$2\left(\frac{n-1}{R}\right) = 2\left(\frac{n_{convex}-n_{concave}}{R}\right) \quad \Rightarrow \quad n = n_{convex} - n_{concave} + 1 = 2.43$$

Objects and Images

Rays that pass through lenses follow paths that have the same characteristics as ray paths for single refracting surfaces. Rays parallel to the optical axis all cross the axis at the same focal point on the other side of the lens. One would therefore expect the object and image distances to be related to the focal length by a formula similar to Equation 4.4.6. The only question is what happens to the indices of refraction. Well, with lenses, the object and image *are in the same medium*, which means that n_1 and n_2 are the same. It is true that the lens itself has a different index of refraction, but we have packaged all the refractions within the lens into a single change of direction at a single point, so those details don't come into play. Setting the indices of refraction in Equation 4.4.6 equal to each other gives a simple relation, known as the *thin lens equation*:

$$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f} \tag{4.5.5}$$

If this looks familiar even without the n's, it's because it is identical to Equation 4.3.5, which we used for spherical mirrors! It is important to note that if we dig into the details hidden within the variable f, we find that for spherical mirrors the focal length is a simple function of the radius ($f=\frac{R}{2}$), while for lenses the focal length comes from the lensmaker equation (two radii and an index of refraction). But as along as the specific details of the focal length are not an issue, then this same equation works for both cases. What is more, we have been very careful in the wording of our sign conventions (all that language about "incoming/outgoing" and "not incoming/not outgoing" sides was carefully chosen), so that the sign conventions work perfectly well for both mirrors and lenses. And finally, the geometry of magnification also carries over to lenses, which means that Equation 4.3.7 also holds for lenses.

Principal Rays

While we have greatly economized the mathematics of geometrical optics to incorporate both mirrors and lenses, the physical processes involved are obviously different. We can express these differences by examining ray traces for lenses. We will see that while they obviously are a bit different from those we did for mirrors, there are many similarities in the logic.

While for the mirror there were four principal rays, for lenses there are only three. As we saw with mirror ray traces, there are differences for the diverging and converging cases. We will start with the converging lens. [*Note: In our diagrams, we will depict a converging lens with a double-convex structure, but it can have any combination of surfaces that results in a converging lens ($f>0$). Similarly, all diverging lenses are depicted as being double concave, though they too can be any combination of surfaces that results in a diverging lens ($f<0$). It should also be mentioned that although the diagrams will give these lenses an apparent thickness, we will be treating them as "thin," which means that the bending of rays will all occur at a single vertical plane through the center of the lens.*]

A major difference between lenses and mirrors is that lenses can allow light to travel in either direction. Therefore, while a mirror has a single focal point on its concave side (no matter which side is reflecting), lenses have focal points on both sides. Two of the principal rays behave like we have seen already – one comes into the lens parallel to the optical axis and exits the lens toward (for converging lenses) or away from (for diverging lenses) the focal point. The other comes in to the lens through (for converging lenses) or toward (for diverging lenses) the focal point, and exits the lens parallel to the optical axis. The third principal ray passes directly through the vertex of the lens without diverting its path.

Figure 4.5.3 – Principal Rays, Converging Lens (Object Outside Focal Point)

Figure 4.5.4 – Principal Rays, Converging Lens (Object Inside Focal Point)

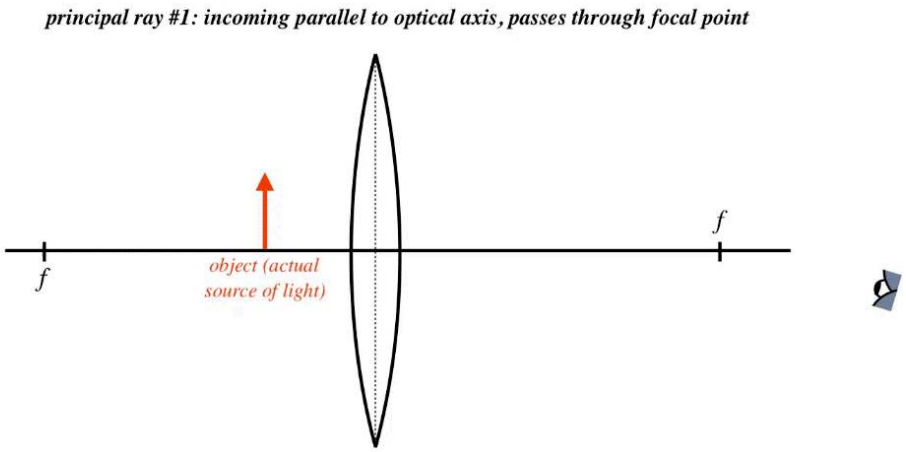
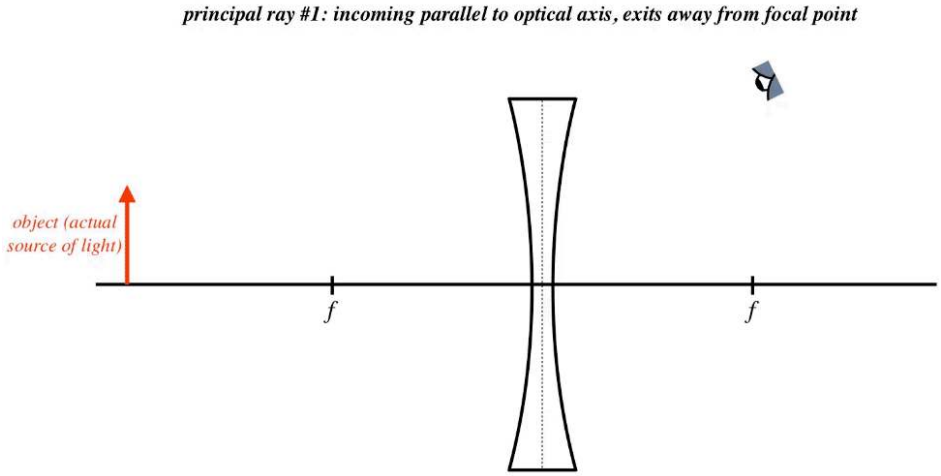


Figure 4.5.5 – Principal Rays, Diverging Lens



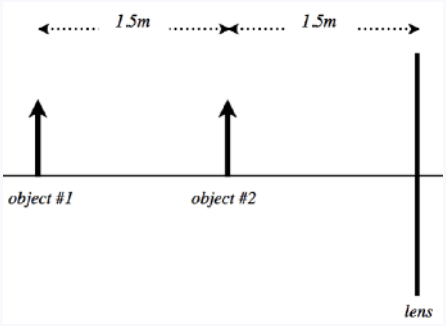
Checking Ray Traces with the Thin Lens Equation

It is worthwhile to confirm that the ray traces shown above give results predicted by the thin lens equation (Equation 4.5.5) and our sign conventions.

1. **converging lens ($f > 0$) with distant object ($s > f$)** – With $s > f$, we have $\frac{1}{s} < \frac{1}{f}$, and combining this with the thin lens equation, we can conclude that $s' > 0$, which means that the image is on the outgoing side of the lens, making it a real image (lies at the intersection of actual light). Furthermore, with both s and s' positive, the lateral magnification $M = -\frac{s'}{s}$ is negative, which means that the image is inverted. From the thin lens equation, we can deduce that when $s > 2f$, then the image comes out closer to the lens than the object ($s' < s$), and the lateral magnification formula tells us that this means that the image is diminished. All of these features are reflected in Figure 4.5.3.
2. **converging lens ($f > 0$) with close object ($s < f$)** – This time the thin lens equation results in an image distance that is negative, because $\frac{1}{s} > \frac{1}{f}$. This indicates that the image will *not* be on the outgoing side of the lens, and is therefore virtual. The lateral magnification comes out positive in this case, indicating that the image must be upright. With a positive object distance and a negative image distance, for the focal length in the thin lens equation to come out positive, it must be true that $\left|\frac{1}{s}\right| > \left|\frac{1}{s'}\right|$. This means that the image must be farther from the lens than the object, and again looking at the lateral magnification equation, this requires the image to be larger than the object. All of this is in perfect agreement with Figure 4.5.4.
3. **diverging lens ($f < 0$)** – With a positive object distance and a negative focal length, the only way that the thin lens equation can be satisfied is for the image distance to be negative, which means that the image is not on the outgoing side of the lens, and the image is virtual. Again the object and image distances having opposite signs means that the image is upright. With the positive object distance and negative image distance needing to produce a negative focal length in the thin lens equation, it must be true that $\left|\frac{1}{s}\right| < \left|\frac{1}{s'}\right|$, which means that the image is closer to the lens than the object, and according to the lateral magnification equation, the image is diminished in size compared to the object. The math agrees with Figure 4.5.5.

Example 4.5.2

Two identical objects are placed with the same orientation, separated by a distance of $1.5m$. A lens is then placed along the line formed by the two objects such that all three of these items are equally-spaced, as in the diagram below. When one views the light coming from the objects through the lens, one sees two images of equal size (though they do not look like it, as the images are not the same distance away). One of these images is upright, and the other inverted



a. *Find the focal length of the lens, including the sign (indicating whether it is converging or diverging).*
b. *Determine which of the two images **appears** larger to the observer.*

Solution

a. For an image to be upright for a single lens (where the object distance is always positive), it must be virtual ($s' < 0$), and for it to be inverted, it must be real ($s' > 0$). Single diverging lenses only create virtual images, so the fact that a real (inverted) image exists here means this lens must be converging. Real images for single converging lenses occur when the object is outside the focal length, and virtual images occur when the object is inside the focal length. Therefore the focal point of the lens must land between the two objects, giving it a range of between $+1.5m$ and $+3.0m$. Okay, so let's do the actual calculation:

With images of equal size but opposite orientation, their lateral magnifications must be negatives of each other. This gives us information about their image distances (we already know their object distances, which we can use):

$$M_1 = -M_2 \quad \Rightarrow \quad -\frac{s_1'}{s_1} = \frac{s_2'}{s_2} \quad \Rightarrow \quad \frac{s_1'}{s_2'} = -\frac{s_1}{s_2} = -2$$

We also have the thin lens equations for both objects. Both involve the same lens, so f is the same in each case:

$$\frac{1}{s_1}+\frac{1}{s_1'}=\frac{1}{f} \qquad \frac{1}{s_2}+\frac{1}{s_2'}=\frac{1}{f}$$

Plugging in $2s_2$ for s_1 and $-2s_2'$ for s_1' in the thin lens equation for object #1, we get:

$$\frac{1}{2s_2}+\frac{1}{-2s_2'}=\frac{1}{f} \quad \Rightarrow \quad \frac{1}{s_2}-\frac{1}{s_2'}=\frac{2}{f}$$

Adding this equation to the thin lens equation for object #2, we get:

$$\left(\frac{1}{s_2}+\frac{1}{s_2'}\right)+\left(\frac{1}{s_2}-\frac{1}{s_2'}\right)=\frac{1}{f}+\frac{2}{f} \quad \Rightarrow \quad \frac{2}{s_2}=\frac{3}{f} \quad \Rightarrow \quad f=\frac{3}{2}s_2=2.25m$$

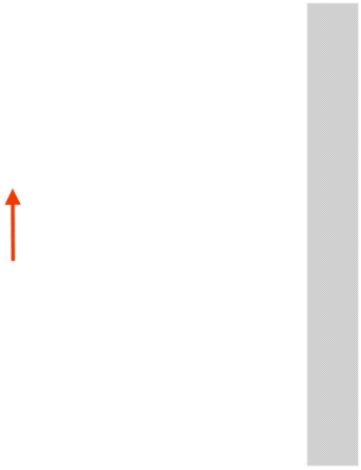
b. The lateral magnifications are equal, but the one that appears bigger is the one that results in a larger angular magnification. This honor belongs to the image that is closer to the observer. Since one of the images is real, it moves to the other side of the lens, where the observer is located. The virtual image remains on the same side of the lens, so clearly the real image must look larger to the observer. For a converging lens, all real images require objects be farther from the lens than the focal length, and virtual images closer than the focal length. Therefore object #1 produces the real image, and image #1 appears larger.

4.6: Multiple Optical Devices

Image Becomes Object

In the previous section we built a lens from two successive spherical refractions. The trick we used to get the lensmaker equation can be extended to many more applications. Indeed, anytime light is affected by more than one optical device on its way from the object to the observer, the image that results from the first optical device in the path is effectively the source for the second optical device. The location of this first image can then be used to compute an "object" distance for the second device. As usual, the best way to visualize the physical basis for this assumption is to think about what is happening to the light waves as it exits the original object, is affected by the first device (which can be a mirror, a lens, or even just a single surface interface between two media), and eventually converges to the image.

Figure 4.6.1 – Image Becomes Object



Alert

Clearly "image becomes object" does not mean that an item suddenly materializes in empty space! Rather, as the light continues past the convergence point, the waves that emerge are indistinguishable from light waves that leave an actual object, which means that the image can subsequently be used as an "object" for further adventures of that light, such as passing through a lens or bouncing off a mirror.

It should be noted that the image-used-as-an-object has an important difference with an actual object. The light that leaves the tip of an actual object arrow propagates out in all directions. The light that leaves the image can only be light that took a path from the original object and was altered by the optical device. This limits the outgoing light of the image to a specific cone. In the figure above, the optical device is clearly a converging lens, and the cone of light that leaves the image represents the region where one is able to look at the object *through the lens*. If one does not look through the lens, of course this image cannot be seen.

Example 4.6.1

A vessel, the bottom of which is a flat mirror, contains water that is $30.0cm$ deep. Someone looks down into the container at their reflection, from a height of $50.0cm$ from the surface of the water. Find how far the image they see is from their actual face. The index of refraction of water is 1.33.

Solution

This problem includes an extra level of thought that has not yet been introduced in this text, but which will play an important role later, so you should not be discouraged if you found this especially challenging. To see how this works, one has to track what happens to the light, and what the "apparent" source of the light is every time that the light encounters a change (reflection or refraction). We will keep track of where images and objects are throughout the calculation, so there will be no need to track the signs of the values (we will use absolute values throughout).

First, the light leaves the face of the person, on its way to the mirror. We know that the mirror will send back the light in a manner exactly symmetric to how it came into the mirror, but the apparent source of the light for the mirror is not positioned

*where the person's face is, because the water refracts the light before it gets to the mirror. We therefore need to compute the distance from the surface of the water that the face **appears to be**, according to the mirror. In essence, the position of the image created by the light passing from the air into the water becomes the position of the object for the mirror. The light is coming from air ($n = 1$ into the water, so according the the mirror the image of the face is farther away from the surface of the water than the actual face is:*

$$s' = \frac{1.33}{1}(50.0cm) = 66.5cm$$

The "object distance" to the mirror equals the apparent distance of the face to the surface of the water, plus the depth of the water:

$$s = 66.5cm + 30.0cm = 96.5cm$$

Now after the light reflects, it behaves as if it is coming from the other side of the mirror, making the apparent source of light a distance of $96.5cm + 30.0cm = 126.5cm$ from the surface of the water.

This light passes back through the surface of the water and is refracted into the air, which means that the apparent source of the light for the person observing it is closer to the water surface than the $126.5cm$ computed above:

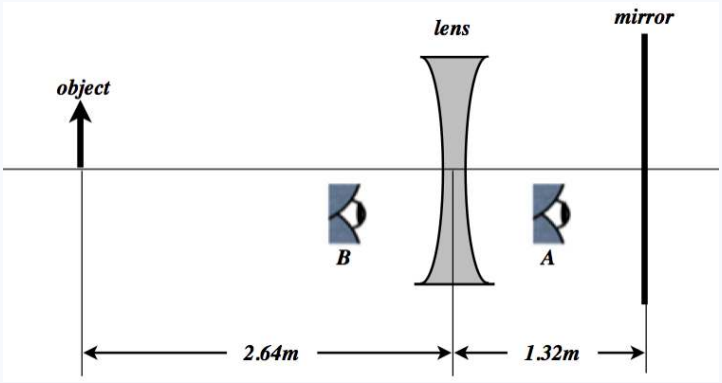
$$s' = \frac{1}{1.33}(126.5cm) = 95.1cm$$

This is the distance from the surface of the water of the image seen by the person. Adding this to the distance of the face from the water gives our answer – the distance between the face and its image: $145cm$

If the water had not been present, then because it is $80.0cm$ from the mirror, the image of the face would have been on the opposite side of the mirror, $160cm$ from the face. So the presence of the water has the effect of bringing the image closer.

Example 4.6.2

A diverging lens is placed in front of a plane mirror as in the diagram to the right. The separation of the lens and the mirror exactly equals the magnitude of the lens's focal length, which is $-1.32m$. An object is placed twice this distance ($2.64m$) on the other side of the lens (see the diagram below).



a. Find the position of the image seen by eye *A* (which is looking into the mirror). Express your answer as a distance measured from the mirror, and indicate which side of the mirror the image appears on (left or right).
b. Find the lateral magnification (relative to the original object) of the image seen by eye *A*. Indicate whether the image is upright or inverted.
c. Repeat (a) and (b) for the image seen by eye *B* (which is looking through the lens into the mirror).

Solution

a. The object distance is positive, so it is -2 times the focal length of the diverging, negative-focal-length lens. Plugging this into the lens equation gives the position of the first image:

$$\frac{1}{-2f} + \frac{1}{s'} = \frac{1}{f} \quad \Rightarrow \quad s' = \frac{2}{3}f$$

The image distance is the same sign as the focal length, so it is negative, placing it on the left side of the lens. This image becomes the object for the next stage, which is reflection off the plane mirror. The distance of this new object to the plane mirror is the distance to the lens (remember it is to the left of the lens), plus the distance between the lens and the mirror, which is given to be the focal length. Therefore the distance of the first image to the mirror is:

$$s = \frac{5}{3} f$$

The image formed by the plane mirror is exactly the same distance behind it as the object is in front of it, so the image viewed by eye A is $\frac{5}{3} f$ behind the mirror. Plugging in for f gives the result: The image seen by eye A is $2.2m$ to the right of the mirror.

b. The only lateral magnification comes from the lens, because plane mirrors provide no lateral magnification (i.e. $M = +1$. This is easily calculated:

$$M = -\frac{s'}{s} = -\frac{\frac{2}{3} f}{-2f} = +\frac{1}{3}$$

The positive value indicates that it is upright, and plane mirrors don't invert images, so the image seen by eye A is upright.

c. We use the result for eye A, as the image for the previous refraction and reflection becomes the object for the second refraction. The distance of the reflected image from the lens is the distance it is behind the mirror plus the focal length of the lens (since that is the separation between the lens and the mirror). Note that the object distance is positive and the focal length negative, so we need to include the minus sign as before. Using this value and the lens equation gives the position of the image seen by eye B relative to the lens:

$$\frac{1}{-\frac{8}{3} f} + \frac{1}{s'} = \frac{1}{f} \quad \Rightarrow \quad s' = \frac{8}{11} f$$

This is the distance from the lens, so the distance from the mirror is $\frac{3}{11} f = 0.36m$ to the left of the mirror. The lateral magnification is the product of the three lateral magnifications (the second one being that of the plane mirror, which is just 1), so we just need to calculate the final lateral magnification:

$$M_3 = -\frac{s'}{s} = -\frac{\frac{8}{11} f}{-\frac{8}{3} f} = +\frac{3}{11} \quad \Rightarrow \quad M = M_1 M_2 M_3 = \left(+\frac{1}{3}\right)(+1)\left(+\frac{3}{11}\right) = +\frac{1}{11}$$

The magnification is positive and smaller than 1, which means the final image is upright and diminished.

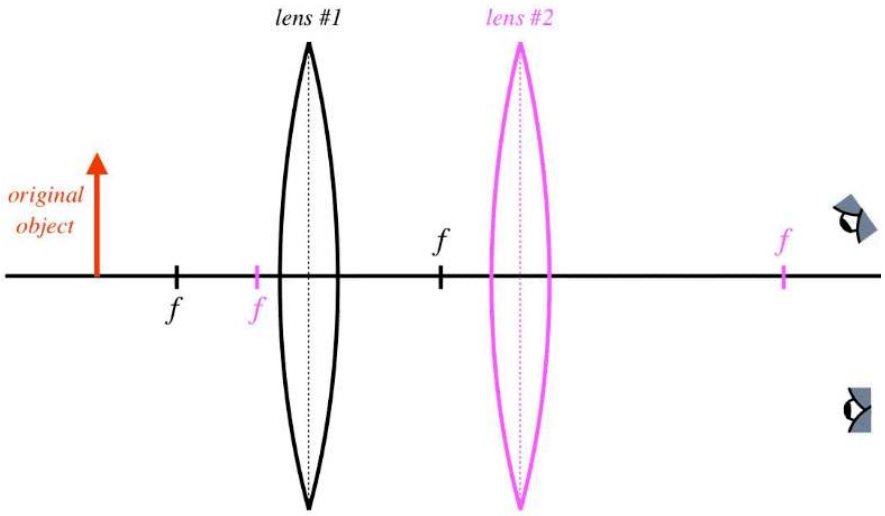
Virtual Objects

Another thing that came up in our derivation of the lensmaker equation was the possibility that the light could pass through a second surface before it was able to converge due to the first surface. The same can happen with two devices. This seems to cause a problem with the idea of the image the first optical device being an object for the next optical device, but mathematically we get around this by giving the object distance for the second device a negative value, in accordance with our sign conventions. While this mathematics does the trick and worked perfectly for the lensmaker equation, it is somehow not as satisfying as seeing how it works with a ray trace. While it is a bit tricky to do, this can be achieved as well. Here is the method:

1. Ignore the second device and use a couple of principal rays for the original object and the first optical device to locate the image of the first device.
2. Keep in mind that these are not the only two rays that we can sketch from the object. All of the rays that leave the object converge toward the first image, even if the light doesn't actually make it there. Of all these rays, select two that *start at the first device and are principal rays for the second device*. Note that one of these may be a principal ray of the first device (the one that emerges parallel to the optical axis), but others will not be.
3. Use two of these principal rays of the second device to sketch the rays that converge to the final image.

The trick here is that the virtual object is not the *origin* of rays for the second device, but rather is the *target* of rays emerging from the first device before they reach the second device. Let's see an example of how this would work for two converging lenses, where the real image of the first converging lens lands behind the second.

Figure 4.6.2 – Ray Trace with a Virtual Object

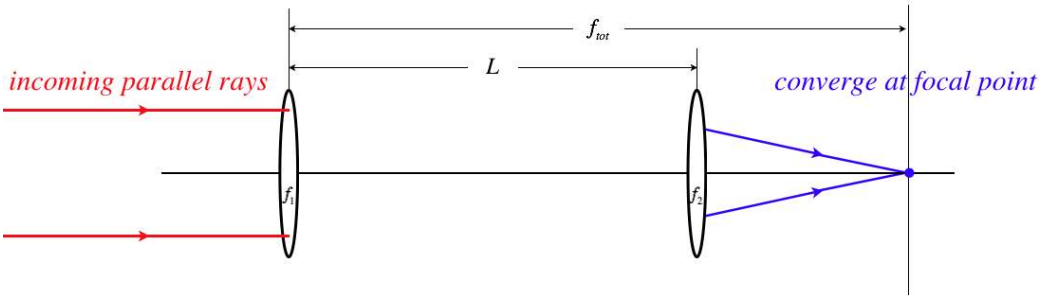


Note that there is an infinite number of rays coming from the first lens, all of them heading for the first image. We just happen to choose two that conveniently happen to be principal rays for the second lens.

Stacking Thin Lenses (Again)

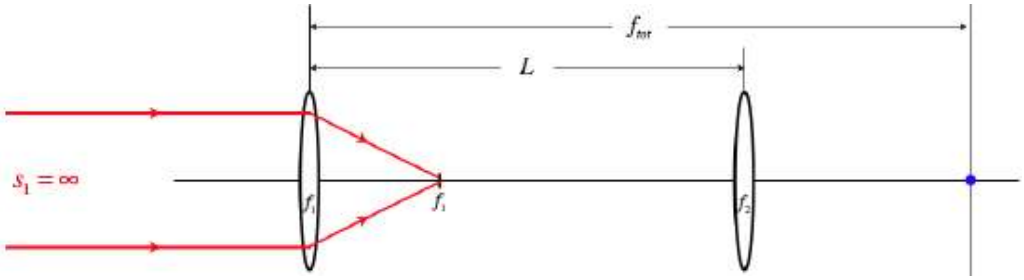
In Section 4.5 we used the lensmaker equation to deduce that multiple thin lenses placed back-to-back result in a single lens with a diopter strength that is the sum of the diopter strengths of the individual lenses (Equation 4.5.4). Here we will show this same equation follows from the first-image-becomes-second-object idea. We start with two lenses (for the sake of this discussion, we will use converging lenses, but this is not required) that are separated by a large distance l. We'll define the "total focal length" f_{tot} as the distance from the first lens to the image of parallel rays after the second lens.

Figure 4.6.3a – Deriving the Lens-Stacking Formula



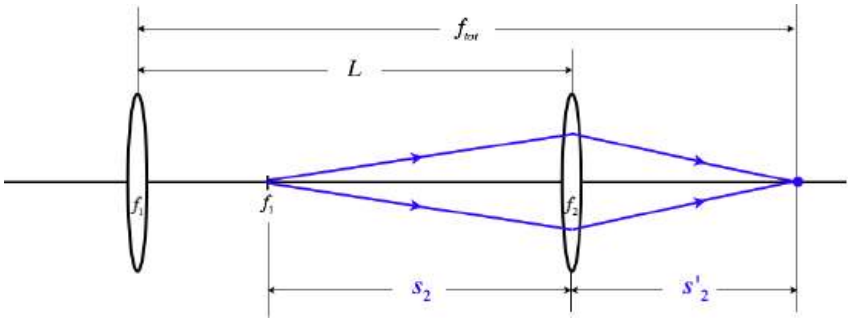
To get the parallel incoming rays, we follow the usual method of placing a point object on the axis very far away from the first lens (just a single point is all we need to get distances we are looking for, no arrow is necessary). According to the thin lens equation, this will result in the image point landing at the focal point of the first lens.

Figure 4.6.3b – Deriving the Lens-Stacking Formula



This image point of the first lens now becomes the object for the second lens, and we can compute the image distance for the second lens.

Figure 4.6.3c – Deriving the Lens-Stacking Formula



Now we need to use the thin lens equation and do a bit of math to write f_{tot} in terms of the other variables:

$$\left.\begin{aligned} &\frac{1}{s_2}+\frac{1}{s_2'}=\frac{1}{f_2} \\ &s_2 = L - f_1 \\ &s_2' = f_{tot} - L \end{aligned}\right\} \quad \frac{1}{L-f_1}+\frac{1}{f_{tot}-L}=\frac{1}{f_2} \tag{4.6.1}$$

Now to get the stacked thin lens result, all we need to do is take the limit where the separation of the lenses goes to zero ($L \to 0$), and sure enough, Equation 4.5.4 is the result.

4.7: Wrap-Up

Simple Magnifier

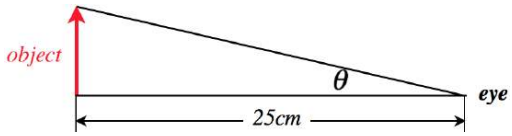
When one thinks about lenses in the "real world," one of the first things that leaps to mind is the magnifying glass. With all of our discussion of lateral magnification in geometrical optics, it's tempting to think that explaining how a single lens facilitates seeing details of small objects should be simple. But it turns out that a full explanation is more complicated than one would think.

The first step in understanding the use of a "simple magnifier" (a single converging lens) is to return to the idea that seeing something better is a function of the fraction of one's field of view the object occupies. The angle subtended by the object is directly related to how big it looks. The angle subtended is a function of two things: the size of the object and its proximity to the viewer. Viewing objects through lenses results in images that differ from the object in *both* of these elements, so we cannot draw a conclusion about the effect of a magnifying glass on the basis of lateral magnification alone.

One way to "magnify" a small object does not even involve a lens – just bring it closer to your eye. The trouble is, the human eye can only observe a clear image that is beyond a certain distance – too close and it gets blurry. The ability to see close items varies from one person to the next, but the generally-accepted standard for the *near point* for human beings (the minimum distance where a clear image can be seen) is $25cm$. So whatever lateral magnification we can artificially create optically, we can always make the image easier to see by moving it closer to our eye, up to a distance of $25cm$.

Let's suppose we maximize this "natural" form of magnification for a naked eye. The biggest angle the item can subtend in our view can be easily calculated from the size of the object y and the $25cm$ separation:

Figure 4.7.1 – Maximum "Magnification" for Naked Eye



$$\tan\theta = \frac{y}{25cm} \tag{4.7.1}$$

So our goal is to use a lens to create a situation where the image subtends a larger angle than is indicated by the equation above, *without that image being closer than $25cm$ to the eye* (which would make it impossible for a human to see clearly). There are three single-lens scenarios to choose from: The diverging lens (Figure 4.5.5), the real image converging lens (Figure 4.5.3), and the virtual image converging lens (Figure 4.5.4). Let's consider each possibility in turn.

The diverging lens actually has the opposite of the desired effect. The image is actually *closer* to the observer than the object, which means that the object must be placed at a distance greater than $25cm$ from the observer. The ray trace in Figure 4.5.5 reveals that the angles subtended by the image and object is the same. This is evident from the fact that the ray that passes straight through the vertex of the lens also passes through the point of both image and object, creating similar triangles. With the object height still being y and the horizontal distance greater than $25cm$, it's clear that the angle subtended is smaller for this case than if we simply look at the object directly, without the lens.

The real image created by a converging lens can laterally magnify the image as large as we like, and if we then observe this image from a distance of $25cm$, the angle subtended can be arbitrarily large. While this works as a magnifier, it comes with some impractical aspects to it. The first is that the image is inverted, which can cause problems for applications like reading fine print. More importantly, to create a real image, the lens must be a distance from the object that it greater than the focal length of the lens, *and* the observer must be a distance from the lens that is at least the object distance plus $25cm$. This is best illustrated with an actual calculation...

Suppose we have a lens with a focal length of $f = +20cm$ (this is a "focusing power" of $\frac{1}{0.2m} = 5\ diopters$), and we want to magnify it so that it subtends twice the angle that it subtends when we look at it directly. Whether we look directly at the object or look at its real image, we will be doing so at a distance of $25cm$, so what we need to do is laterally magnify the image by about a factor of 2. [Actually doubling the height doubles the *tangent*, but we still assume that the tangent of the angle is approximately equal to the angle.] The lateral magnification is 2 when the image distance is twice the object distance, $s' = 2s$. Putting these criteria into the thin lens equation gives:

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \quad \Rightarrow \quad \frac{1}{s} + \frac{1}{2s} = \frac{1}{f} \quad \Rightarrow \quad s = \frac{3}{2}f = 30cm \quad \Rightarrow \quad s' = 3f = 60cm \tag{4.7.2}$$

The observer must be an additional $25cm$ from the lens to see this image clearly, which places the observer $85cm$ from the lens and $115cm$ from the object. Okay, so this can be accomplished, but the idea of getting an inverted $2\times$ magnification from a distance of more than a meter from the object is not very practical. Let's look at the virtual image case of the converging lens.

Like the diverging lens, the object and image both subtend the same angle, but unlike the case of the diverging lens, it is the *object* that is closer to the observer. The observer's ability to focus on what it is seeing depends upon the *image* being at least $25cm$ away. This means we can actually move the object *closer than* $25cm$ from the observer, and since the angle subtended is determined by the height of the object and its distance from the observer, this allows the angle to get larger than it can without the lens present. What is more, unlike the case of the real image, the observer doesn't need to be any distance at all from the lens, as the image forms behind it.

Let's re-do the calculation for the $5diopter$ converging lens, this time using a virtual image. We want the image to be a distance of $25cm$ from the observer, who can place their eye right up to the lens. This means that the image distance we want is just $s' = -25cm$. From this we can solve for the object distance, and with that, the lateral magnification:

$$\frac{1}{s} + \frac{1}{-25cm} = \frac{1}{20cm} \quad \Rightarrow \quad s = 11.1cm \quad \Rightarrow \quad M = \frac{s'}{s} = \frac{25cm}{11.1cm} = 2.25 \tag{4.7.3}$$

Given that the object and image form similar triangles, the angular magnification comes out approximately ($\tan\theta \approx \theta$) the same. So this gives us a magnification that is greater than $2\times$, the image is upright, and the lens is only $11cm$ from the object.

There is one advantage that comes with the real image method that is not available in the virtual image case, though it is not easy to implement. By adjusting the object distance closer to the focal length, the size of the image can be made as large as we like, which means that the observer can just look from a distance of $25cm$ farther, and the amount of angular magnification can be made arbitrarily large for a given lens. For the virtual image case, we can increase the angle subtended by the image by moving the object closer to the lens, but doing so also moves the image closer to the lens (and observer). So this magnifier is limited to images that are $25cm$ from the lens.

Converging lenses used for magnification are often labeled according to their magnifying power. When they are, the standard is usually the angular magnification when the object is placed at the focal point of the lens. That is, the *magnifying power* of a lens is the ratio of the angle subtended when the object is at the focal point and the angle subtended when no lens is used (and the object is at $25cm$). This comes out to just a simple ratio of the two object distances:

$$M_{magnifier} = \frac{\theta_{image\ with\ magnifier}}{\theta_{image\ at\ 25cm}} \approx \frac{\tan\theta_{image\ with\ magnifier}}{\tan\theta_{image\ at\ 25cm}} = \frac{\frac{y}{f}}{\frac{y}{25cm}} = \frac{25cm}{f} \tag{4.7.4}$$

Magnification Devices

Not surprisingly, a lens has greater magnifying power when its focal length is shorter. So suppose we have a rather powerful magnifier that we would like to use to view something very small or something very far away. In both cases, for the magnifier to work, we need to get the object a very small distance (less than the focal length) from the lens. This is difficult to do for very small objects because the tolerances are so small, and because it might be difficult to get light to the small object with the lens looming so close to it. And it is obviously impossible for distant objects.

The solution is to use *two* lenses. The first lens (known as the *objective*) has as its primary function the creation of a real image at a convenient position. Then the second lens (the *eyepiece*) functions as a simple magnifier that acts on that real image to increase the angle it subtends.

The device that magnifies small objects is of course a *microscope*. As a secondary function, the objective lens creates a laterally magnified real image, because the objective is pretty close to the object, while the real image it creates is farther away. Then the real image becomes an object for the eyepiece, which then magnifies it further (in the usual angular sense of a simple magnifier).

The device that magnifies distant objects is a *telescope*. Its objective lens's secondary function is to collect as much of the light as possible coming from the object (this explains why objective lenses are so much larger than the eyepieces). The object distance for the objective lens is very large and the image distance very small, so obviously it exhibits a tiny fractional lateral magnification and diminishes the image (stars are very large compared to their images). But this real image is then located very precisely at a position where the eyepiece can magnify the angle it subtends.

Odds and Ends

There are a number of interesting observations and side discussions about optics that don't fit well into the development of the subject that we will now address.

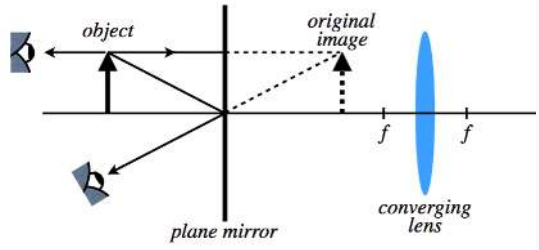
1. **light intensity** – With all the geometry we do with straight-line rays, it's easy to forget that we are talking about light *waves*. Figure 4.6.1 expresses this wave nature in the context of geometrical optics fairly well, and just reminding ourselves of the wave nature in this way leads to some interesting ideas. Consider light intensity. If we have a plane wave of light coming into a converging lens parallel to the optical axis, what comes out the other side is a spherical wave that is *collapsing* down to the focal point. We already know that outwardly-moving spherical waves lose intensity as they move outward (the inverse-square law), so waves that are collapsing into a single point *grow* in intensity. This is why one can cause a piece of paper to light on fire with a magnifying glass in the sunlight. By

placing the lens a distance equal to its focal length from the paper, all of the light energy passing through the lens is concentrated to a small region, raising the temperature to the combustion point.

2. **wave forms and standing waves** – For many applications we discussed in physical optics, we required plane waves. We can get these by looking at a very distant source like the sun, but we can also do this locally by placing a point source at the focal point of a concave mirror. All the light that reflects off the mirror will come away from it parallel to the optical axis – plane waves. Lasers also figure prominently in our study of physical optics (thanks to their coherence), and one of the important elements in the design of lasers is something called a *resonant cavity*. This is essentially two spherical mirrors that cause the light to reflect back-and-forth, forming a standing wave, into which energy is pumped, and from which the beam emerges (like sound coming from a standing wave in an organ pipe).
3. **images have no substance** – We have mentioned this already, but it bears repeating. When we do geometrical optics, the object and image arrows are both drawn in the ray trace, and in some sense are given equal status. But the object is an actual thing, made of atoms, while an image is a designated point in space where the light appears to be coming from. The thing that can get lost is that light only appears to be coming from that place if the *actual* light can be seen. Put in more visceral terms, an image created by a lens can only be seen if the observer looks through the lens! It sounds silly to have to say this, but in a ray trace diagram, one can see from any angle the image arrow that is drawn, and the notion that one must be looking in a specific direction is lost.
4. **projection onto a screen** – Everyone knows that images can be projected onto a screen, but exactly how does that work? The first thing to note is that every position on the screen is associated with a single point on the image, which means that for a sharp image to occur, *all* the light from a single point on the object must converge to a single point on the screen, and from there it is reflected. This means that the screen must be placed exactly at the position of the image (the point where the rays converge is the definition of the position of the image). If the distance is off slightly, then light from nearby points on the object land at the same point. At any given point, most of the intruding light comes from nearby points, and this is perceived as blurring of the image. There is one other very important thing about projected images: *It is not possible to project a virtual image onto a screen*. The reason should be obvious – if a screen is placed where the rays converge for a virtual object, there is not actually any light there, so it cannot reflect off that screen.
5. **lens size** – If we have two lenses of equal focal lengths, what is different about them if they are different sizes? One thing we don't tend to do is discuss the light that misses the lens and continues in its straight line. This light does not contribute to the image, which means that "collecting" more light with a larger lens only affects the brightness of the image. It perhaps seems strange that coating the left half of a lens with an opaque paint will not cut the image in half, but a dimmer version of the same image is what is seen.

Example 4.7.1

A ray trace for an object near a plane mirror is shown below. If a converging lens with the focal length shown is placed at the position indicated (two focal lengths from the position of the image), find the distance that the new image is from the original image.



Solution

*This problem points out how easy it is to get caught up in the geometry while losing sight of what is physically going on. The lens has been placed **behind the mirror**. Clearly this will not change in any way the image seen in the mirror. The "new" image will be the same as the original image.*

Summary of Geometrical Optics Mathematics

To close this chapter, its useful to pull together all of the mathematical relations into one place, as it allows us to see familiar patterns that run through all of them. In particular, it is interesting to note how defining variables in certain ways has allowed us a great deal of economy, especially in the sign conventions.

reflecting surface:	$f=\frac{R}{2}$	$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f}$	$M=-\frac{s'}{s}$
refracting surface:	$f=\frac{n_2}{n_2-n_1}R$	$\left(\frac{n_1}{n_2}\right)\frac{1}{s}+\frac{1}{s'}=\frac{1}{f}$	$M=-\frac{n_1}{n_2}\frac{s'}{s}$
lens:	$\frac{1}{f}=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$	$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f}$	$M=-\frac{s'}{s}$

(4.7.5)

name	symbol	positive when	negative when
object distance	s	*object on incoming side*	*object on other side*
image distance	s'	*image on outgoing side*	*image on other side*
lateral magnification	M	*image upright*	*image inverted*
radius of curvature	R	*center of sphere on outgoing side*	*center of sphere on other side*
focal length	f	*focal point on outgoing side*	*focal point on other side*

(4.7.6)

Looking at these, we can draw the following conclusions for cases involving *single optical devices only* (when an image for one device can become an object for a second device, these don't necessarily apply):

- Object distances are positive.
- When an image is virtual, its image distance is negative, which makes the magnification positive, so virtual images are upright and real images are inverted.
- Converging devices can produce real or virtual images (real when $s > f$, virtual when $s < f$), while diverging devices can only produce virtual images. This follows from the sign of the focal length and the fact that object distances are positive.
- Diverging devices (diverging lenses and convex mirrors) produce diminished images (which, as stated above, must be upright and virtual), since the negative value of f and positive value of s requires that s' is smaller in magnitude than s.

CHAPTER OVERVIEW

5: Fundamentals of Thermodynamics

5.1: Temperature

Thermal Energy

Back in Physics 9A, the idea of *thermal energy* first arose in the context of energy conservation. We concluded that work done by non-conservative forces would convert mechanical energy into this form of energy, which became internal to the system, and didn't spontaneously return to mechanical form. We know intuitively that this form of energy reveals itself to our senses through temperature. We also can intuit that two objects at different temperatures that are brought into contact can exchange energy and change their temperatures (cold milk added to hot coffee both cools the coffee and warms the milk).

It therefore makes sense that we begin our exploration of this topic by examining the concept and the measurement of temperature, but before we do, two definitions are in order:

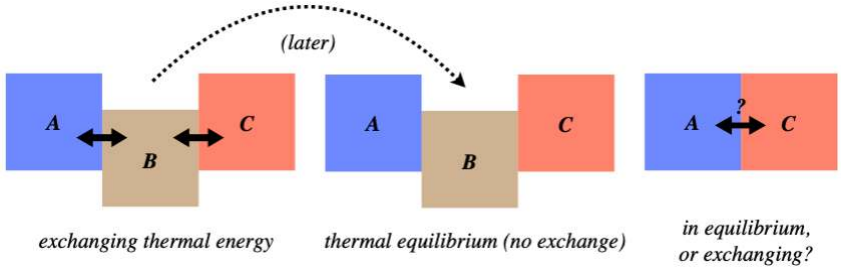
thermal contact: when two systems are allowed to directly exchange thermal energy with each other

thermal equilibrium: when two systems in thermal contact nevertheless exchange zero net thermal energy

Zeroth Law of Thermodynamics

Imagine the following scenario: Three systems, which we will call *A*, *B*, and *C*, are arranged as shown in the left figure below (in thermal contact), exchanging thermal energy. They are allowed to continue this until systems *A* and *B* are in thermal equilibrium, while systems *B* and *C* are also in thermal equilibrium. The question is, after these equilibria are reached (the middle figure), if system *A* and *C* are brought into thermal contact (the right figure), will they automatically be in thermal equilibrium?

Figure 5.1.1 – Thermal Equilibrium Thought Experiment



We won't beat around the bush on this – the answer is "yes." While worded in the arcane language of physics with all the references to thermal equilibrium, the answer is fairly obvious when worded differently: When two systems come to thermal equilibrium, they have reached the same *temperature*, and when two systems that have the same temperature come into contact, no thermal energy is exchanged between them. That is, we don't see one of them get spontaneously hotter than the equilibrium temperature while the other gets cooler than the same temperature.

The key thing to take away from this is that temperature is an *indicator* of whether a thermal transfer will spontaneously occur between two systems in thermal contact. We know this is true intuitively, and this fact explains the peculiar name of this law. Back in the 19th century when thermodynamics was being developed, the first and second laws of thermodynamics were expressed. In the 20th century there was a somewhat more rigorous treatment of the subject, and physicists realized that for completeness one should probably state that a property that we call “temperature” exists and can be uniquely defined. They didn’t feel like it was a good idea to make this a later law in the queue (there were three laws of thermodynamics by then, and in the 60’s another principle was dubbed the fourth law), so they went backwards to zero.

To co-opt a famous phrase from Supreme Court Justice Potter Stewart in 1964, temperature is not easy to define, but when it comes to determining if one thing is hotter than another, *you know it when you see it*. What this means is that temperature has many easily-measurable consequences in the physical world, and we can define temperature *in a relative sense* in terms of those consequences. We will see exactly how we do this now, and in fact we will even come up with a clever way of defining it in an absolute sense (i.e. we will not have to compare temperatures).

Relative Temperature Scales

When it comes to defining a temperature scale, the fact that (for now) we are only interested in differences means that we can set our scale arbitrarily. One way to do this is to pick a phenomenon that we can refer to as our reference. One phenomenon that we

know is related to temperature is phase changes. We commonly see water in all three of its phases – ice, liquid water, and steam – and water is the most abundant substance on the surface of our planet, so it seems like a good choice. We know that it freezes at some low temperature and boils at some high temperature (well, we are actually *defining* temperature such that it is lower when water freezes than when it boils). Therefore we can arbitrarily set temperature equal to zero degrees at the freezing point, and set it equal to 7.5 degrees at the boiling point. We can then write these melting and boiling points as 0^oW and 7.5^oW. On this scale, the standard temperature of a healthy human body is 2.775^oW. This is not a very compelling scale, and one that makes more sense is the *celsius* (or *centigrade*) scale, which divides the temperature from water freezing to boiling into 100 equal degrees, starting at zero for freezing. Another is the *fahrenheit* scale, which starts at 32^oF for freezing and 212^oF for boiling.

Digression: Give the Fahrenheit Some Love!

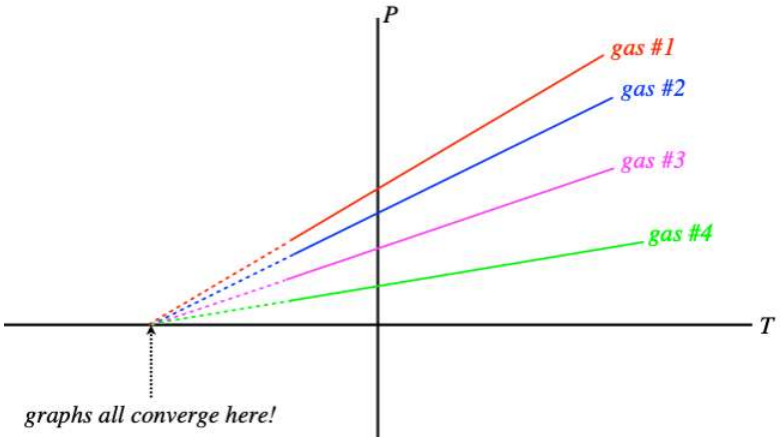
It may seem like the fahrenheit scale is incredibly arbitrary, but in fact it has many practical advantages over the celsius scale. These advantages are decidedly "human" in nature. Our ability to sense temperature differences is pretty sensitive (depending upon many factors), so having more degrees between two close temperatures provides us a level of precision that is easier to relate to (without having to resort to more decimal places). This is especially important if we are reading the temperature off a device (thermometer) that has only whole number gradations. Another human element is that the fahrenheit scale reserves 0^o and 100^o for (more-or-less) the extremes of human experience (at least in a cold northern climate, like where the German physicist Daniel Gabriel Fahrenheit lived). And there is one other nice feature, which interestingly is related to the advantage of using degrees instead of radians when measuring angles. The circle is divided into 360 degrees so that one can subdivide it an enormous number of ways, while still making the angle observable – 360 has prime factors of: $2^3 \times 3^2 \times 5$. Well, the number of subdivisions of the Fahrenheit scale from freezing to boiling water is also very high, since the difference is 180^oF. In a time before calculators, such advantages were always welcome.

We have two points on our temperature scale that we can measure (using water), and these two points define the line that is our scale, but the question arises, "*How do we measure the temperatures between the freezing point and the boiling point (and the temperatures beyond them in both directions)?*" We do this not by measuring temperature directly, but rather by measuring the *effects* of temperature changes. We will look at a few of these effects, and how they can be used to fashion a thermometer.

Gas Thermometers and the Absolute Temperature Scale

The first effect we will look at is how temperature changes cause changes in pressure in a gas (assuming that everything else is held fixed). Suppose we perform an experiment where we measure the pressure of a gas for many temperatures (but a fixed volume) and graph the results. We find that the relationship is linear, but the graphs of different gases (or different volumes of the same gas) have different slopes. When this was first analyzed, the technology was not available to bring the temperatures down extremely low, but extending these graphs reveals the remarkable result that when the pressure hits zero, all the graphs hit the axis at the same point!

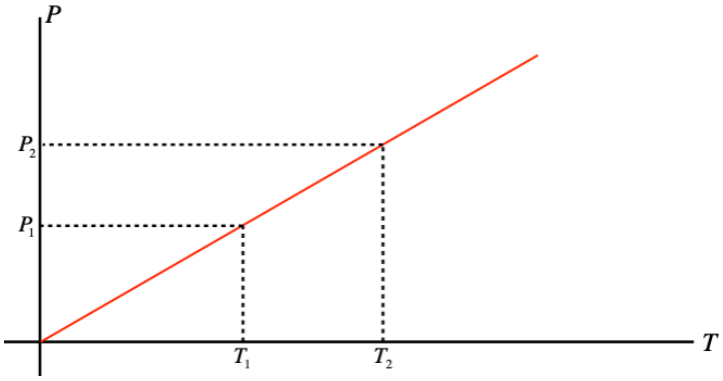
Figure 5.1.2 – Extrapolating Absolute Zero



What this suggests is that we can redefine the zero point for temperature, and it will hold for all gases in all circumstances – an absolute scale is born! All that remains is to determine what temperature differences between neighboring integers should be used, and Celsius was the choice. So essentially the absolute temperature scale (a unit of measurement called *kelvins*, no "degrees") is the same as Celsius, with the zero point shifted backward. How far? $0^oC \approx +273.15K$.

Notice that this result also provides us an opportunity to simply relate pressures and temperatures of a gas in a confined volume. Re-drawing the P vs. T graph with the T axis measured in kelvins, we get a line that intercepts the origin:

Figure 5.1.3 – Gay-Lussac's Law



Using similar triangles, we can immediately derive a relationship between pressures and temperatures for gases confined to a fixed volume:

$$\frac{P_1}{T_1} = \frac{P_2}{T_2} \quad \Rightarrow \quad \frac{P_1}{P_2} = \frac{T_1}{T_2} \tag{5.1.1}$$

5.2: Thermal Expansion

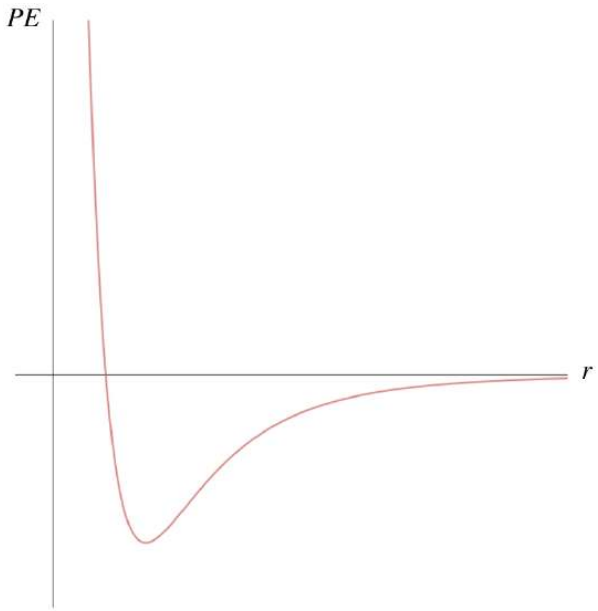
The Microscopic Model of Thermal Energy

Another phenomenon commonly seen that is a result of a changing temperature is the expansion/contraction of solids and liquids. While we may know that this occurs, understanding *why* it occurs requires a bit of review from Physics 9A.

Whenever one wants to talk about the fundamentals of thermal energy, two words immediately come up: *microscopic* and *random*. That is, we see thermal energy as the random motions of the tiny particles that comprise the macroscopic objects we see. A macroscopic indicator of the density of this random microscopic energy is what we call temperature. How does this energy manifest itself? Well, like any energy, it comes in basically two forms: potential and kinetic. We already know what the *KE* is, but on the microscopic level, what is the *PE*?

The potential energy comes from inter-particle forces that are at their core electromagnetic in nature. But there are many electric charges involved, and the potential energy function associated with this complicated combination of forces generally takes on a shape that looks like this:

Figure 5.2.1 – Molecular Potential Energy

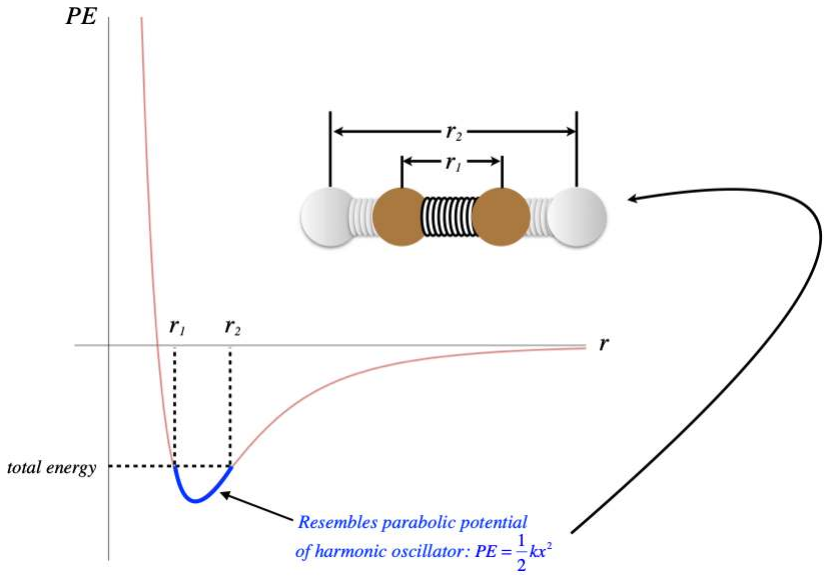


Recall that the force can be determined from the *PE* graph by taking the negative of the derivative (technically, the gradient):

$$F = -\frac{\partial U}{\partial r} \tag{5.2.1}$$

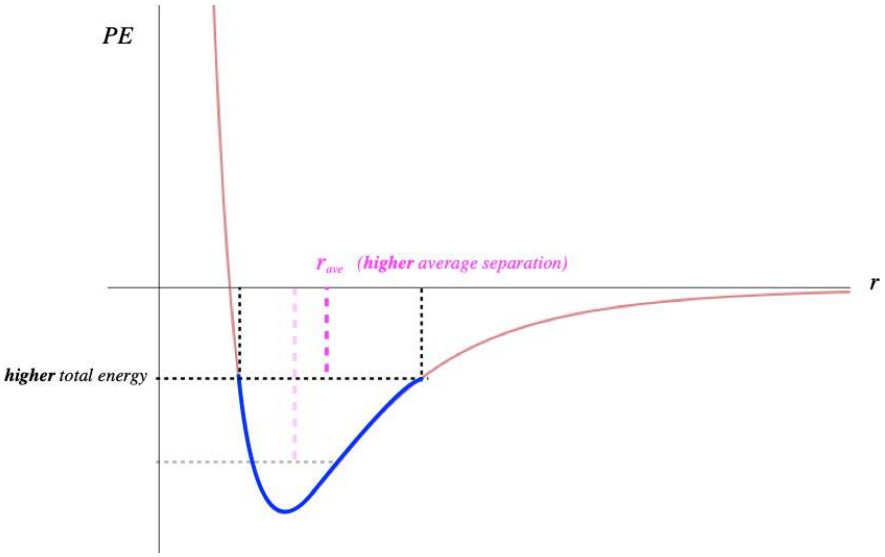
So the places on this curve that have negative slopes have positive forces, and the positive direction is in the $+r$-direction, which is a repulsive force. When the slope is positive, the force direction is negative, which is in the $-r$-direction and is therefore attractive. We therefore have a force that is alternately repulsive and attractive around a certain equilibrium position – near the bottom of the dip, usually called a *potential well*. The *PE* curve down there is closely approximated by a parabola, so the *KE* and *PE* of the particle combine to result in a physical system that looks very much like two particles bound together with a spring, with the equilibrium point of the spring being the separation of the particles associated with the bottom of the well. [Go here for a more complete review of this topic from Physics 9A.]

Figure 5.2.2 – Modeling Two-Particle Interactions with a Spring



We know that when we add energy to a spring system, the amplitude of the oscillations increases. The same is true here – the distance across the well widens. But *unlike* a spring system, this *PE* curve is not quite symmetric. When the energy of the particle-particle system is increased, the average separation actually *increases*.

Figure 5.2.3 – Increasing Energy Increases Average Separation



Keep in mind that when energy is added to a macroscopic system, it gets distributed through all the particles, raising all of their energies, and increasing all of their average separations. If we have N particles in a row, and the average separation increases by some small factor, then the overall length of this chain of particles increases by the same factor. While the amount of change on the microscopic level is incredibly small, it is measurable on the macroscopic level because there are so many particles.

Linear Expansion

The end result of this is that (for reasonably low temperature changes) the percentage expansion/contraction of the length of an object is proportional to the temperature change:

$$\frac{\Delta L}{L_o} = \alpha \Delta T \tag{5.2.2}$$

The constant α is called the *coefficient of linear expansion*, and it has units of K^{-1} (or, equivalently, ${}^oC^{-1}$). This is a constant that depends upon the type of material we are looking at.

Alert

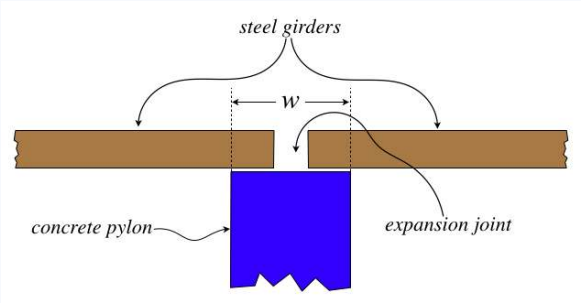
*We can use ^oC for temperature here, because we are only interested in temperature **changes**. The scale for kelvin and celsius are the same, only differing in their zero points. It will be a good idea as we go forward to keep close track of when these scales are interchangeable (temperature changes) and when they are not (absolute temperature related to pressure, for example).*

This thermal expansion formula can be rewritten so that it relates an initial length to a final length that occurs due to a temperature change:

$$\Delta L = \alpha \Delta T L_o \quad \Rightarrow \quad L_f - L_o = \alpha \Delta T L_o \quad \Rightarrow \quad L_f = L_o (1 + \alpha \Delta T) \tag{5.2.3}$$

Example 5.2.1

A structural engineer is designing the portion of a bridge where steel girders rest on top of concrete pylons, as shown in the diagram below. Space needs to be reserved between consecutive girders in the expansion joint, so that a thermal expansion of the girders will not cause them to butt against each other and buckle, and the width of the concrete pylon must be sufficient so that thermal contraction of the girder does not result in it falling off the pylon. If the distance between the centers of consecutive pylons is $60\ m$, find the minimum width of the pylons for no buckling or slippage of the girders to occur for a temperature range between the extremes of -50^oC and $+50^oC$. The amount of thermal contraction/expansion of the concrete pylons is negligible, and the coefficient of linear expansion of steel is $1.2 \times 10^{-5}\ {}^oC^{-1}$.



Solution

The longest that a girder can be is the separation of the pylon centers, or else consecutive girders would buckle with each other. When a girder contracts due to cooling, both ends get drawn-in, so the shortest the girder can be is the pylon separation minus the pylon width, so:

$$L_{min} = d - w \quad \Rightarrow \quad w = d - L_{min}$$

The maximum expansion/contraction occurs when the temperature swings from the minimum to the maximum value (a 100^oC difference), so:

$$\frac{\Delta L}{L_{min}} = \alpha \Delta T \quad \Rightarrow \quad \frac{w}{d - w} = \alpha \Delta T$$

Solving for the width, we get:

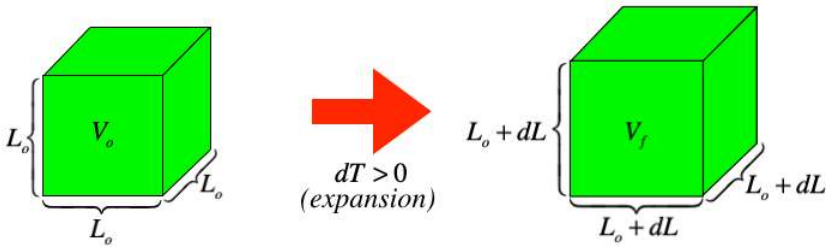
$$w = \frac{\alpha \Delta T}{1 + \alpha \Delta T}\ d = \frac{0.0012}{1.0012}(60m) = 7.2cm$$

Naturally in practice one would want to provide much more tolerance than this, but this is the absolute lower limit.

Volume Expansion

Of course, solids and liquids expand in more than just one dimension when their temperatures increase, and we can extend our previous analysis to determine the amount of *volume expansion*. For simplicity, we will consider a cube (this works with any shape, such as a sphere, but the calculus is easiest in cartesian coordinates):

Figure 5.2.4 – Infinitesimal Volume Expansion



We already know how to handle the linear expansion of each of the dimensions of this cube, so now we want to determine how much the *volume* expands as a result. The difference of the final and initial volumes is:

$$dV = V_f - V_o = (L_o + dL)^3 - L_o^3 = \left(L_o^3 + 3L_o^2 dL + 3L_o dL^2 + dL^3\right) - L_o^3 = 3L_o^2 dL + 3L_o dL^2 + dL^3 \qquad (5.2.4)$$

This is an infinitesimal change in the volume, so we expect an infinitesimal on the right-hand-side of the equation, but the product of multiple infinitesimals as we see in the last two terms are vanishingly small compared to the single infinitesimal, so we throw them away. We can now do a bit of algebra on what remains, to get the effect of temperature change on the volume:

$$\Delta V = 3L_o^2 \Delta L = \frac{3L_o^3 \Delta L}{L_o} = 3V_o \frac{\Delta L}{L_o} = 3V_o (\alpha \Delta T) \quad \Rightarrow \quad \frac{\Delta V}{V_o} = 3\alpha \Delta T \qquad (5.2.5)$$

This looks very similar to the equation for linear expansion, with the coefficient α replaced with 3α. We can rename this constant "β" and call it the *coefficient of volumetric expansion*, and for all but anomalous materials (such as crystals that expand different amounts along different axes) or large temperature changes (where those other terms we threw out might become significant), it simply equals three times the coefficient of linear expansion.

Digression: The Weirdness of Water

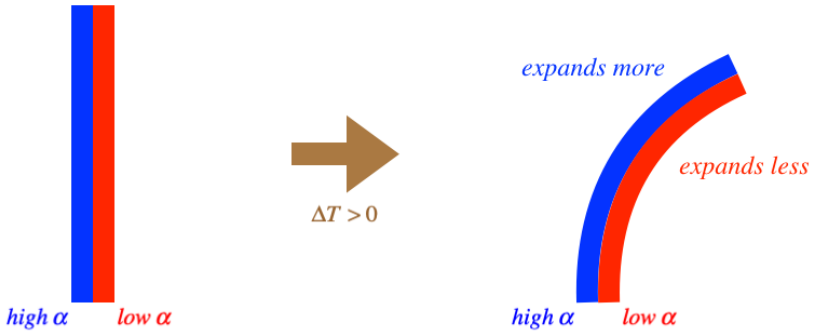
There is one notable exception to this phenomenon, and it happens to apply to the most common substance on the Earth's surface: water. It turns out that water near its freezing point actually expands as it cools. In addition, it expands significantly more when it changes phase to ice. If it did not, then ice would sink rather than float (which means, for example, that ponds would freeze from the bottom-up, making things significantly less comfortable for fish).

Expansion Thermometers and Thermostats

Now that we know that the expansion of a substance is proportional to its temperature change, we can get back to our idea of using temperature-based phenomena to measure temperature by using this phenomenon to fashion a *thermometer*. The simplest is the mercury thermometer. As the mercury changes temperature, it expands or contracts, so if we confine its expansion to be along a straight line, we can mark the container (a thin glass tube) with a linear scale and the temperature will follow that scale. Another type of thermometer involves what is called a *bimetallic strip*.

If we take two flat strips of different metals with different coefficients of thermal expansion and adhere them to each other, then raising their temperature will cause their lengths to expand by different amounts, resulting in the combination bending in the direction of the strip that reacts less to the temperature change (i.e. the one with a lower coefficient of linear expansion).

Figure 5.2.5 – Bimetallic Strip



The temperature scale can therefore be measured in an *angular* manner – placing a scale that records the amount of bending will mirror the temperature change. An interesting application of this is the *thermostat*. The strip is made of a conducting material, so electric current will run through it. If we position electrical leads so that when it curls too much, the strip closes a circuit with the air conditioning, it will come on automatically. If it cools to much (curls back the other way), it can close a circuit that kicks-on a heater.

5.3: Heat Capacity and Phase Transitions

Definition of Heat

In the pantheon of misused physics terms, at the top of the list must be the word *heat*. Standard usage would have us believe that it is a quantity of energy stored within a system, measurable by temperature. But we already have a term for that – thermal energy. Heat is a much more subtle concept.

In Physics 9A, we learned about the idea of work as a means of transferring energy between systems, or between types (e.g. from potential energy to kinetic energy, or mechanical energy to thermal energy). Clearly it would make no sense to say that two systems contain a certain amount of work, which then can be transferred between them. The definition of work makes it clear that it *only* exists as a transfer – the contained energy is kinetic, potential, or thermal, and when energy is moved or converted through a force acting through a distance, the energy is transferred through the process of work.

Heat is precisely like work in this regard. It is a process of the transfer of energy between systems, which necessarily requires an interaction between the systems. In the case of work, the interaction involved is a macroscopic net force. The emphasis here needs to be on the word "net," which indicates an imbalance. For heat transfer to occur, there also needs to be an imbalance, but in that case the imbalance needs to be in temperature. Microscopically, this means that when two systems come into thermal contact, the random motions of particles naturally tend toward bringing the average energies of the particles in both systems into balance. In short, work is energy transfer resulting from a (macroscopic, non-random) force imbalance, while *heat is energy transfer resulting from a temperature imbalance*, which is expressed microscopically as an imbalance in average particle energy.

Heat Capacity

We know that work can convert energy back-and-forth between kinetic and potential energy (when the force is conservative), and can transfer energy into thermal form (when the force is non-conservative), but because heat involves temperature differences, it is obviously intimately tied to thermal energy, and not mechanical energy. The most natural first question to ask is therefore, "How is the amount of heat transferred into or out of a system related to the temperature change that results from that transfer?"

We can answer this question through direct experiment, and the results of such an experiment are quite intuitive. Given that the temperature is a measure of average particle energy, then it stands to reason that if we add a known quantity of heat to a sample and the temperature goes up (say) 1^oC, we would expect that adding the same amount of energy again to the same sample would have the same effect. After all, the energy added is divided among the same number of particles each time, so the average energy per particle will go up the same each time. We express this result as a statement of proportionality:

$$Q \propto \Delta T \tag{5.3.1}$$

Another fairly obvious factor is the *amount of matter* in the sample. Again, the added heat is distributed amongst the particles in the sample, and the temperature change is a measure of how much their average energies go up. If we put in a known quantity of heat, then the more particles there are in the sample to share this energy, the less each particle's temperature changes. This means that the heat needed to raise the temperature a certain amount is proportional to the amount of matter in the sample. Now the "amount of matter" can be measured in many ways, and we will use two different ones, depending upon which is more convenient. Here we will look at the most obvious choice – the mass of the sample. We therefore extend the proportionality above to a factor of the amount of mass:

$$Q \propto m\Delta T \tag{5.3.2}$$

It's clear that the units don't match here, so there will need to be a constant of proportionality to turn this into an equality. But there may be additional information that needs to be contained within that constant as well. In this case we really do need to appeal to experiment, and we find that different substances require different constants for the equality to hold. We therefore introduce the *specific heat capacity, c*, which differs from one substance to the next, and which completes the relationship between heat transferred and temperature change:

$$Q = mc\Delta T \tag{5.3.3}$$

The mass is always a positive quantity, and we define the specific heat capacity as a positive value, which means that $Q > 0$ when $\Delta T > 0$. The temperature of a system rises when heat *enters* it, so in defining heat capacity as a positive value, we set the following sign convention:

$Q > 0$ whenever heat enters a system, and $Q < 0$ whenever heat exits a system

Digression: Specific Heat Capacity

Perhaps the reader is wondering where the name for this constant comes from. The word "specific" generally comes into scientific vernacular whenever one wants to factor out the role of a sample's mass. For example, the "specific gravity" of a sample is the ratio of its mass to the mass of an equal volume of water. So two samples of the same substance that have different masses have the same specific gravity. The specific heat capacity factors out the role of mass as well – i.e. the heat capacity per unit mass.

The second question is how the word "capacity" applies. The idea is that the heat capacity (the product of specific heat capacity and mass) is a measure of a sample's "capacity" to take in heat energy for a given temperature increase. Samples with a large heat capacity can take in a lot of heat energy without changing their temperature very much. There is a similar use of a derivative of the word "capacity" (called "capacitance") in the field of electromagnetic theory, which you will study in Physics 9C. In this case, it measures a system's "capacity" to separate charge across a given voltage difference.

The mention of specific gravity in the digression points out a common theme in chemistry and physics – using water as a standard for units of measurement. We already have a standard set of units (SI) that we have been using since day 1 of Physics 9A, but it is useful to point out a very common unit of energy – the *calorie*. This is defined as the amount of heat energy that must be added to 1 gram of water to raise its temperature by 1 degree Celsius. In other words:

$$c_{water} = \frac{Q}{m\Delta T} = \frac{1\ calorie}{(1\ g)\,(1^oC)} \tag{5.3.4}$$

We can, of course, convert between calories and joules, and it turns out that:

$$1cal = 4.184J \quad \Rightarrow \quad c_{liquid\ water} = 4184\frac{J}{kg^oC} \tag{5.3.5}$$

Alert

Because calculations involving heat capacity involve changes in temperature, it's okay to use Celsius in the units rather than Kelvins – changes in one of these equals changes in the other.

This provides us with a powerful tool for solving problems. If we put a hot piece of metal into a container of cool water, we can relate the starting temperatures to the final temperature reached when the two substances come to equilibrium. All we need to do is invoke conservation of energy (the heat that leaves one substance enters the other – and be careful to follow sign conventions properly), and use the above expressions. We can also measure the change in temperature in a reservoir of water that surrounds a chemical reaction to determine how much energy goes into or comes out of a chemical reaction. This process is commonly referred to as *calorimetry*.

Example 5.3.1

A $50.0\ g$ chunk of aluminum is left in a pot of boiling water at 100^oC until it comes to thermal equilibrium with the water. It is then placed into an insulated vessel containing $200\ g$ of water at 20.0^oC, and the system is allowed to come to thermal equilibrium. Find the final temperature this system reaches. The specific heat capacity of aluminum is $0.215\frac{cal}{g\,^oC}$.

Solution

The starting point is noting that the insulated container allows for no heat transfer to or from outside the system, which means that all heat exchanged is between the aluminum and the water. All the heat that leaves the hot aluminum enters the cool water, so energy conservation (and our sign convention for heat) demands (the "a" subscript refers to aluminum and the "w" to water):

$$Q_a = -Q_w$$

Now plugging in Equation 5.3.3 for Q, we have:

$$m_a c_a \Delta T_a = -m_w c_w \Delta T_w$$

Now we use the fact that the changes in temperature go from the starting temperatures to the same final temperature (thermal equilibrium), and solve for the final temperature:

$$m_a c_a (T - T_a) = -m_w c_w (T - T_w) \quad \Rightarrow \quad T = \frac{m_a c_a T_a + m_w c_w T_w}{m_a c_a + m_w c_w} = 24.1^o C$$

Note that the final expression before the numerical answer is exactly a weighted average of the two temperatures, with the weightings of each temperature being the heat capacity of each substance.

The second version of heat capacity we will work with changes the proportionality into terms of moles rather than mass. In this case, we refer to the constant as *molar heat capacity*, and we denote it with an upper-case C:

$$Q = nC\Delta T \tag{5.3.6}$$

The n in this formula is the number of moles in the sample. We will find this version to be very useful for gases in particular. It also fits well with some of the more general thermodynamic concepts that we will get to later.

Phase Transitions and Latent Heat

When it comes to heat transferring into or out of substances, changing the temperature of the sample is not all that can occur. The sample can also undergo a *phase change*. By "phase" we mean the solid, liquid, and gaseous states of matter. These differ primarily in how the particles involved interact with each other. Adding heat to a sample changes part or all of it from solid into liquid (*melting*), or from liquid into gas (*boiling*). Having heat exit a sample can result in the phase changing in the opposite direction: liquid to solid (*freezing*) or gas to liquid (*condensing*).

What is interesting about phase changes is that they *occur at a fixed temperature*. The temperature at which solid/liquid transitions occur is called the *melting point* of that substance, while the temperature at which liquid/gas transitions occur is called the *boiling point*. Note that the temperature of melting is the same as the temperature of freezing, and the temperature of boiling is the same as the temperature of condensing. A transitions that occurs at a fixed temperature requires a different relationship than above between the heat transferred and the change it invokes. Rather than being proportional to the amount of temperature change, the heat added/lost is proportional to the amount of mass made to change phase. We write it this way:

$$Q = \pm L\, \delta m \tag{5.3.7}$$

The value δm employs a lowercase δ to express a the amount of mass that changes phase which is assumed to always be a positive quantity, without invoking the "after-minus-before" that one generally associates with use of the uppercase Δ – the amount of mass present doesn't change (it doesn't go anywhere), there is just some amount of mass that is changing phase. The sign $\pm$ is positive when heat is entering the system, $Q > 0$, and the substance is either melting or vaporizing. The sign is negative when heat is leaving the system, $Q < 0$, and the substance is either freezing or condensing. The constant L is called the *latent heat*. This is a rather confusing name for this quantity, on a couple of counts. First, it is not heat – it doesn't even have the units of energy. And second, "latent" seems to imply that there is some heat sitting around stored in a sample, a misconception we have already addressed above. For any given substance, the constant L is different if it involves a phase change between solid and liquid than if it involves liquid and gas. The constant associated with the former we call the *latent heat of fusion* (L_f), and the the constant that deals with the latter is called the *latent heat of vaporization* (L_v).

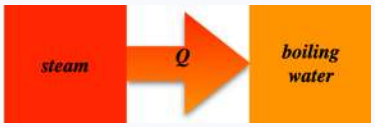
Naturally processes where heat is exchanged can involve both temperature change and phase change at the same time (such as ice melting in a cup of warm water until equilibrium is reached), but assuming the system where the process is occurring is insulated, energy conservation (heat lost by one substance is gained by the other) shows the way.

Example 5.3.2

*A sealed insulated vessel containing a quantity of water is heated until it is brought to a boil, and the heating continues until all of the water is steam at a temperature of $245^o C$. One **fifth** as much boiling (liquid) water is then injected into the vessel and the contents are held at constant pressure and are allowed to come to equilibrium. Find the final state of the contents of the vessel – the percentage of water in each phase (if mixed), or the temperature of the contents (if all in one phase). The specific heat capacity (at constant pressure) of steam is about $1900\ \frac{J}{kg^o C}$, and the latent heat of vaporization of water is $2.256 \times 10^6\ \frac{J}{kg}$.*

Solution

Let's start with a sketch to clarify what is happening here:



Energy is leaving the hotter steam and entering the cooler water. At least for awhile, this will result in the steam cooling off while some of the water changes phase to steam. We won't know, until we do the calculation, if all of the water will change to steam (and then get hotter than the boiling point), or if the steam will come down to the boiling point, thereby reaching thermal equilibrium. We'll start by calculating the energies required to do both of these things – unfortunately there is no shortcut to avoid having to do this.

$$\text{cooling the steam to } 100^oC: \qquad Q_1 = mc_s\Delta T = m\left(1900\ \tfrac{J}{kg^oC}\right)(-145^oC) = m\left(275,500\ \tfrac{J}{kg}\right)$$

$$\text{water to steam:} \qquad Q_2 = L_v\delta m = \left(\frac{m}{5}\right)\left(2.256\times10^6\ \tfrac{J}{kg}\right) = m\left(451,200\ \tfrac{J}{kg}\right)$$

We see that it takes more energy to turn all of the liquid water into steam than to reduce the temperature of the steam to the boiling point, so this heat transfer will end with a mix of liquid water and steam at 100^oC. We now calculate the mass of the water changed to steam in terms of the starting mass of the steam, using energy conservation:

$$0 = Q_1 + Q_2 = -m\left(275,500\ \frac{J}{kg}\right) + \delta m\left(2.256\times10^6\ \frac{J}{kg}\right) \quad \Rightarrow \quad \delta m = 0.122m$$

The total amount of steam at the end can now be divided by the total amount of steam + liquid water to get the percentages:

$$\left.\begin{aligned} m_{steam} &= m + \delta m = 1.122m \\ m_{total} &= m + \frac{m}{5} \end{aligned}\right\} \frac{m_{steam}}{m_{total}} = 0.935$$

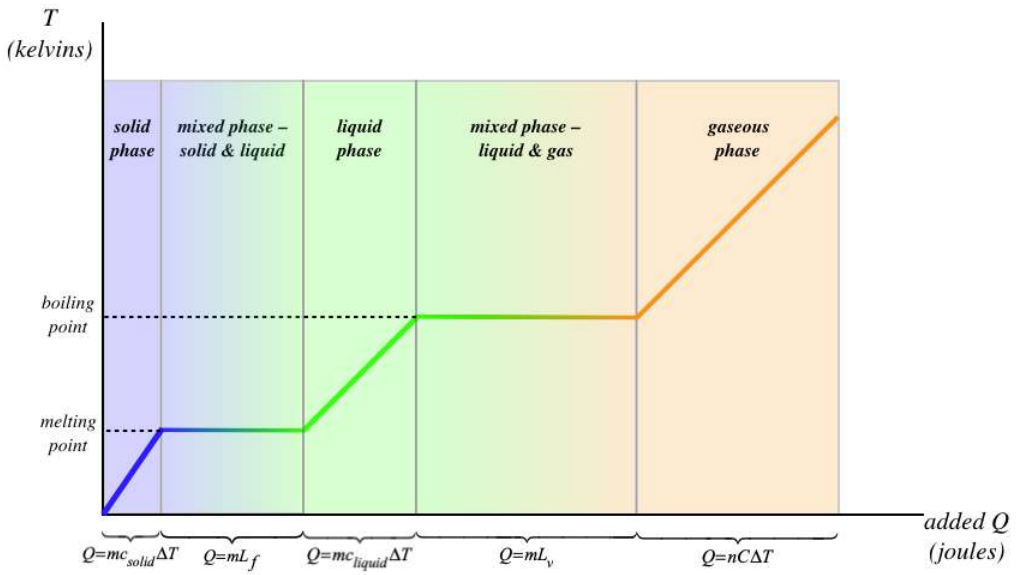
So the final result is 93.5% steam and 6.5% liquid water, at a temperature of 100^oC.

Suppose we wanted to transition a very cold solid all the way to a very hot gas. There would be several steps involved:

- add energy to raise the solid's temperature to the melting point
- add energy to change the phase from solid to liquid (while not changing the temperature)
- add energy to raise the liquid's temperature to the boiling point
- add energy to change the phase from liquid to gas (while not changing the temperature)
- add energy to raise the gas's temperature

Every one of these steps involves a different constant. The specific heat capacities of the solid, liquid, and gaseous phases of the same substance are not the same, and the latent heat of fusion is not the same as the latent heat of vaporization for the same substance. The full process is depicted in the graph below.

Figure 5.3.1 – Effect of Heat Transfer on Temperature and Phase



[Note: In the final section of the graph, you'll note that the molar heat capacity is used. This is because it is more common to measure the amount of a gas in moles rather than in units of mass.]

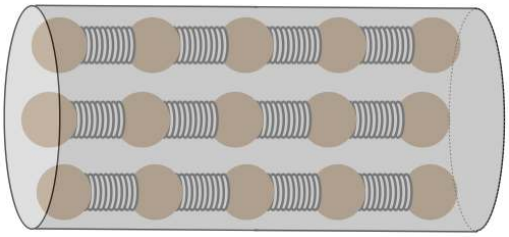
5.4: Modes of Heat Transfer

Conduction

We know the effects of heat being transferred into or out of systems, but now we are going to take a look at the *ways* in which this transfer can occur. As we stated earlier, “heat” is a rather generic description of energy transfer due to a temperature difference between two systems, and we will see there are three modes through which this transfer can occur. The first is the most intuitive, and as it turns out, the one we can most easily deal with mathematically. It is called *conduction*.

As we saw with thermal expansion, the trick to understanding conduction is to consider what is happening on a microscopic scale. Consider a solid cylindrical object that connects two systems at different temperatures. This cylinder acts as the conduit for heat energy to flow from the hotter system to the cooler one. We model this cylinder microscopically using parallel chains of particles joined by springs.

Figure 5.4.1 – Heat Conductor Model

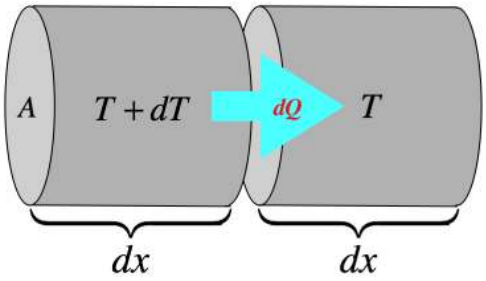


[*Technically, these particles should be attached to all of their nearest neighbors by springs, but we will only be looking at the transfer of the heat along the length of the cylinder, so we have simplified the model accordingly.*]

It should be clear from this model how heat energy can be transferred from one end of the cylinder to the other: If we vibrate the particles on one end of the cylinder, they will vibrate their nearest neighbors, and the effect will carry its way down to the other end. This is easy to see, but what is tougher to understand is that we will be considering only steady-state circumstances, which means that the particles on each end of the cylinder vibrate with amplitudes that have energies that match the systems with which they are in contact (these regions are called *thermal reservoirs*, because during the heat transfer process their temperatures don’t change appreciably). Every particle between the ends vibrates with an amplitude between the two extremes defined by the hot and cold reservoir.

Okay, so our task now (as it will be with all forms of heat transfer) is to determine the rate at which energy is transferred from one system to another in terms of the conditions provided. We'll do this by considering each element of this model in turn. As with any analysis of a continuously-changing phenomenon, we start with differential elements. In this case, we have two small segments of the cylinder at slightly different temperatures, across which some heat is transferred.

Figure 5.4.2 – Differential Heat Conduction



The more chains of spring-connected particles we can use, the faster the energy can be transferred. The number of chains is proportional to the cross-sectional area of the cylinder, so the rate of heat transfer is also proportional to the cross-sectional area:

$$\frac{dQ}{dt} \propto A \tag{5.4.1}$$

The next factor in determining the rate of heat flow between these two segments is the temperature difference. It should not be surprising that heat will flow faster when the difference in temperature is greater. It turns out that the rate of heat flow is directly proportional to the temperature difference. This phenomenon is often referred to as *Newton's law of cooling*, and works fairly well as an approximation in more general circumstances, though it is only strictly applicable to this one. We have to be careful about the sign we use; recall that the sign for dQ is positive when heat is flowing into a system, but in this case the heat is flowing out of the system with the higher temperature:

$$\frac{dQ}{dt} \propto -dT \tag{5.4.2}$$

In order to get all the energy in the first segment to be transferred into the second segment, energy in the left end of the segment has to traverse the length of that segment, dx. The longer this segment is, the longer it will take, so the rate of heat transfer is inversely-proportional to that distance:

$$\frac{dQ}{dt} \propto \frac{1}{dx} \tag{5.4.3}$$

Different substances will be structured differently (different springs, different masses of particles, etc.), so we have to take into account the type of substance. We do that by incorporating that into the constant of proportionality (*thermal conductivity, k*) that turns the proportional relationships into an equality:

$$\frac{dQ}{dt} = -kA\frac{dT}{dx} \tag{5.4.4}$$

The derivative of the temperature is called the *temperature gradient*, and can be thought of as the steepness at which the temperature tapers-off from the origin of the heat transfer (the hotter thermal reservoir) to its destination (the cooler thermal reservoir), which in the most general cases (e.g. in non-steady-state situations) will not be constant. This is known as the *heat equation*, but really it is a specific example of the *diffusion equation*, which applies to many other phenomena as well.

It turns out that this equation is overkill for our purposes (we are not about to start solving differential equations), and in fact in all the cases we address we will deal with steady-state situations with the temperature gradient being linear. When the temperature changes linearly from the hot thermal reservoir to the cool one, the gradient is a constant, equal to simply the temperature difference of the two reservoirs (ΔT), divided by the distance separating them (L):

$$\frac{dQ}{dt} = -\frac{kA}{L}\Delta T \tag{5.4.5}$$

Example 5.4.1

An iron bar with a square cross-section is used to melt square holes in a slab of ice. The bar is then cut in half and the two halves are welded together on their sides to make a new shorter bar with a rectangular cross-section. If the mass of ice melted in 10 minutes by the original bar is M, how much ice will be melted in 10 minutes by the new configuration? Assume the temperature on the hot end of the bar is the same in both cases.

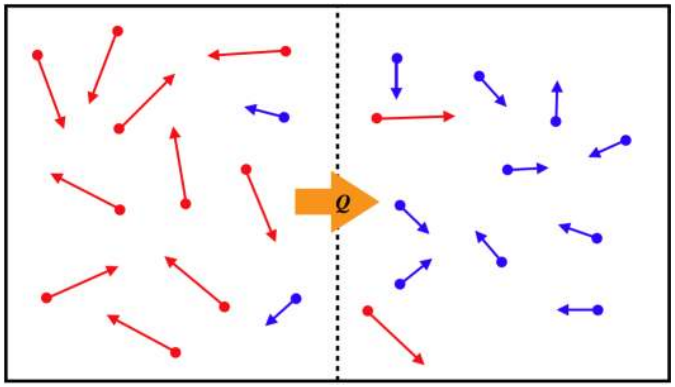
Solution

The rate of heat conduction is proportional to the cross-sectional area and inversely proportional to the length of the material through which the heat passes, and in this case the area doubles while the length is divided in half. The material is the same (thermal conductivity is unchanged), so the rate of heat flow is quadrupled. With four times as much energy transferred into the ice in the same period of time, four times as much ice is melted, so the answer is $4M$.

Convection

Convection is another form of heat transfer, that operates through a completely different mechanism from conduction. Rather than particles interacting with each other, the energy is transferred by simply having particles with more *KE* move (thanks to random motion) from the hotter region to the cooler one, while lower-*KE* particles take their place in the hotter region (there is no net exchange of particles), resulting in a transfer of energy.

Figure 5.4.3 – Convection Mechanism



Applying rigorous analysis to derive a mathematical model for the rate of heat flow via convection is well beyond the scope of this course. But an approximate relation based on the temperature difference of the two reservoirs is:

$$\frac{dQ}{dt} \propto [\Delta T]^{\frac{5}{4}} \tag{5.4.6}$$

Two things to note here:

- While this doesn't quite follow "Newton's Law of Cooling" like conduction, it comes very close.
- We don't have the constant of proportionality, or even what this constant depends upon. That doesn't mean this isn't useful, because we can still compare the convection rates for two scenarios where "all else is equal."

Without being more specific about the proportionality constants for conduction and convection, it is nevertheless a safe bet to say that convection in general is a faster mode of heat transport. For example, if you have a plastic bag full of hot water and a plastic bag full of cold water, and you want two plastic bags of warm water (so you want to transfer heat from the hot bag of water to the cold bag of water), the fastest way to achieve that is by mixing the water from the bags together, rather than by bringing the bags in contact with each other.

Radiation

The third way in which heat transfer can occur should become obvious when one thinks about how our Earth stays warm. After all, it is in contact with the vacuum of space (which in the absence of a nearby sun is at a temperature of $3K$), so heat should be transferring out of it at an alarming rate. The source of the Earth's incoming heat transfer is of course the Sun. But the space between the Sun and Earth is not conducting heat (it's empty space - no particles connected with springs), and the Sun isn't firing really hot particles to mix with the Earth's atmosphere (well actually it is, but that "convection" is not doing much to heat our atmosphere, though it puts on a nice light show at the poles). Instead, the Sun is transferring energy to the Earth via *radiation*.

We already know that radiation is just light waves. We also know that light waves are driven by vibrating electric charges (electrons). One source of vibrating charges is atoms in a sample with thermal energy. The hotter the sample gets, the more energetically the charges vibrate, which means more energy is sent out in the form of radiation, so we would expect the electromagnetic power output of a sample to grow as its temperature grows, but it is by no means obvious in what way the power output will mathematically depend on the temperature.

These light waves don't know where they are going, they only know that some vibrating electrons are driving them, so it is not the temperature *difference* that is causing this heat transfer, but rather the *absolute* temperature. A fellow named Boltzmann derived the dependence of the power output on temperature, and a guy named Stefan measured it. It turns out that the power output goes as the *fourth power* of the absolute temperature. The actual power output also depends upon the surface area (more space for the radiation to come out of), and a property called *emissivity*, which measures how well the surface emits light (how rough/smooth the surface is, how it is shaped, etc.) into the region just outside the object (the surface is the border between these two regions):

$$\left|\frac{dQ}{dt}\right| = \sigma e A T^4 \ , \qquad \sigma \equiv 5.67 \times 10^{-8} \frac{W}{m^2\ K^4} \tag{5.4.7}$$

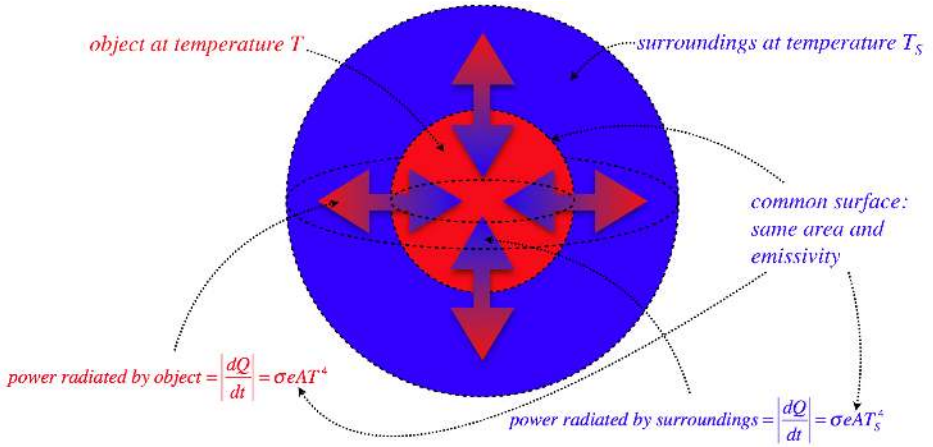
The constant σ is called the Stefan-Boltzmann constant, e is the emissivity, and A is the surface area of the emitting body. The absolute value is included here because this equation involves the absolute temperature rather than a temperature difference. The sign we put to this equation depends upon whether we are talking about the rate at which heat that is exiting the object at

temperature T (in which case the sign is negative), or the rate at which heat is entering the region surrounding the object at temperature T (in which case the sign is positive).

Nowhere here have we mentioned the frequency of the light emitted. It turns out that all of the frequencies (up to a certain maximum) are emitted, but the energy transferred is not uniform across frequencies. Remember, these thermal electrons are vibrating *randomly*, although that randomness has a non-uniform distribution, making some frequencies more common than others. Most of the light emitted in this manner at "everyday" temperatures (say, hundreds of kelvins) is in a part of the spectrum that we refer to as *infrared*, a frequency range we are unable to see with the naked eye, though we can see it with the help of special devices (e.g. infrared cameras). The part of the power output that is in the visible spectrum is too low for us to be able to see when the temperature is at "typical" temperatures in the region of 300*K*. But if something gets significantly hotter, the power output of every frequency goes up, and power in the visible spectrum can reach a level that we can see – the object "glows hot."

Recall we said that heat transfer requires a temperature *difference* to occur, but here we seem to be saying that heat is transferred out of an object at an absolute temperature. Well, any object that can emit light can also *absorb* it. So let's consider an object sitting in an environment which is at a different temperature.

Figure 5.4.4 – Heat Transfer to/from Surroundings Via Radiation



The emissivity is a property of the boundary surface between the two realms exchanging heat (which in this case we are calling the object and its surroundings), so naturally it is the same value going in both directions (we'll see another reason that this must be the case shortly). Obviously the surface area of the boundary is also the same going both ways as well. So the only thing that makes the heat energy exiting the object different (and entering the surroundings) from the heat energy going the other way (from surroundings to object) is the difference in temperature. We can now employ the sign convention for heat and conclude that the net rate of heat entering the surroundings is:

$$\frac{dQ}{dt} = \sigma e A \left(T^4 - T_S^4\right) \tag{5.4.8}$$

Once again we see that net heat flow is induced by a difference in temperature, though like convection, this mode does not obey Newton's law of cooling.

Example 5.4.2

A typical red giant star is so big that it can fit about 1,000,000 stars the size of our sun inside of it (i.e. red giants occupy a volume about 1 million times greater than the volume occupied by our sun). For our sun to radiate energy at the same rate as such a star, how would their temperatures need to compare?

Solution

One million times the volume translates into 100 times the radius, which in turn translates into 10,000 times the surface area. The rate of energy transfer due to radiation is proportional to the surface area, so if the temperatures were equal, the red giant would radiate energy at a rate 10,000 times that of our sun. The rate of energy transfer due to radiation also goes as the 4th power of the temperature, so if the sun was 10 times hotter than the red giant (it turns out it is not, though it is close to twice as hot), then that would exactly compensate for the much greater surface area of the red giant.

Digression: Temperatures Near Stars

As seen in the example, a common application of the Stefan-Boltzmann law comes from the study of stars. But such cases do not involve two regions sharing a common surface border at different temperatures. That is, the space immediately outside the surface of the sun is not the same near-absolute-zero temperature that it is far outside the solar system. Intuitively it makes sense that at steady-state the temperature would gradually decrease from what it is at the surface of the sun, down to the ~3K temperature we see in deep space, but is there some way to compute this temperature gradient? It's clear it can't be linear as it is for conduction, because drawing a straight line from the 5800K temperature of our sun down to 3K many billions of light years away would mean that the earth is residing in space that has a temperature that is essentially the same temperature as the sun. Figuring out this temperature gradient is essential to finding planets around other stars that could support life, because the planets will be in approximate thermal equilibrium with the space around them, and we assume life can only be sustained within a certain temperature range (the so-called "Goldilocks zone").

To solve this problem, let's consider a star with radius R_o and surface temperature T_o. Next, construct an imaginary spherical surface centered at the center of the star, with a radius R. We wish to compute the temperature T evaluated at this surface. Treating the star as a blackbody, the rate at which energy is radiated from it is:

$$\frac{dQ}{dt}(star) = \sigma \left(4\pi R_o^2\right) T_o^4$$

Now let's imagine that we treat our imaginary surface as a radiator of energy – we don't even know about the star inside of it. Naturally it behaves like a blackbody, because none of the radiation that strikes it from outside is reflected (it is imaginary!), and a perfect absorber is exactly the definition of a blackbody. We can therefore compute the rate at which it radiates energy outward:

$$\frac{dQ}{dt}(sphere) = \sigma \left(4\pi R^2\right) T^4$$

But of course, all of the power that comes from the star passes through this surface, so the power emitted by the surface is the same as the power emitted by the star. Setting them equal and solving for T gives:

$$T(R) = \sqrt{\frac{R_o}{R}} T_o$$

So we see that the temperature drops as the inverse-square-root of the distance from the star.

5.5: Thermodynamic States of Ideal Gases

State Variables

One of the most important concepts we will use in this course is the idea of a *thermodynamic state*. There are two key elements to this:

- In thermodynamics, we only deal with *equilibrium states*. By this we mean that if the physical conditions imposed on the system are not changed, then none of the macroscopically-measurable properties of that state will change. An example in terms of what we have discussed already is the temperature of a sample. We have been assuming that the sample is a uniform temperature throughout its volume. If it were not, then even if we allow no heat to enter or exit the sample (don't change the physical conditions), heat transfers *within* the sample would still occur, and we would be able to measure temperature changes in various regions of the sample. A sample with differing temperatures occurring in separate regions is not in an equilibrium state.
- For any given equilibrium state, we can completely describe its condition with just a few macroscopically-measurable quantities. So for example, if we have a volume of gas, we can completely define its equilibrium state by measuring its *temperature* (T), *volume* (V), and *pressure* (P). This may not seem to be all that amazing, but the point is that we will see there are many other quantities that can be measured as well (*number of particles* (N), *internal energy* (U), etc.) that we can compute from our three measured values, making it unnecessary to measure these new quantities separately. That is, the thermodynamic state is completely defined by the measurements of temperature, volume, and pressure, and every other measurable quantity is then uniquely defined by the fact that we know what state the system is in.

All of the measurable quantities like those mentioned above that define a thermodynamic state are called *state variables*. What is interesting is that we are not required to measure the three specific state variables mentioned earlier in order to completely define the system – we can mix-and-match them! Using the example above, we could measure the number of particles, the pressure, and the internal energy of the gas, and those measurements would be sufficient to compute the other unmeasured state variables like volume and temperature.

Heat and work are also important concepts in thermodynamics, but ***they are not state variables***! We could have guessed this, since they involve *transfers* between systems. That is, heat and work are responsible for *changing* thermodynamic states – they do not serve to define them. As we continue in this subject, we will see how both of these quantities can morph one state into another, thereby changing one or more of the state variables.

Pressure

Many of the state variables mentioned above require no explanation – we have already discussed temperature, and the meanings of particle number and volume are obvious. We will get to internal energy in due course (as well as other state functions not yet mentioned), but pressure requires a brief introduction. Like temperature, the formal definition of pressure is mathematically complicated, but also like temperature, we can get a sense of what it is by the macroscopic effects it causes. Fluids (and in particular, gases, which we will be studying here) consist of many particles moving in a random fashion. When these particles are confined, they bounce off the walls of the container. The transfer of momentum caused by all these collisions is manifested as a force on the confining wall by the fluid. The amount of force is of course a function of how many particles are hitting the surface at any given moment, which means it is proportional to the area of that surface.

We want to define pressure as a property of the fluid, and not dependent upon the container, so we therefore define pressure of the fluid as the amount of force it exerts on a surface *per unit area* of that surface. What makes pressure tricky is that it takes into account only the disordered (random) motion of the particles. If the fluid is moving in a macroscopically-measurable way (like a river flowing), then the collisions of the particles that result from this macroscopic motion, while it does account for a force, it does not contribute to the pressure of the fluid. For example, your hand is constantly being pressed in all directions from randomly-moving particles in the air, and the force from these collisions can be used to compute the air pressure. But when you hold your hand out the window of your car as the car moves, the force of the air that you feel on your hand is *not* air pressure. This is because the motion of the air that contributes to this force is not random – it is ordered because it acts in a single direction.

The units of pressure in SI are: $[P] = \frac{N}{m^2} \equiv Pa$. The renamed units are called *Pascals*. There are several other units that are used as well, such as psi (lbs per square inch), and torr (aka millimeters of mercury). We will discuss pressure in greater detail, including the origin of this last odd-sounding unit of measurement in Chapter 7.

Equations of State

Given that the state of a system is defined by three state variables, with all the other state variables thereby well-defined, there must be some equations that get us from the values of the state variables we measure to the values of the others. These formulas that relate state variables to each other are called *equations of state*. For example, the pressure of a state could be computed from (i.e. is a function of) volume, the number of particles, and the temperature:

$$P = f(V, N, T) \tag{5.5.1}$$

These relationships between state variables (functions) are not the same in all cases – the relationships depend upon the physical system we are talking about. The type of system we will examine extensively (because it is the simplest one for which we can derive useful information, and because it works very well as an approximation) is that of an *ideal gas*. Physically, an ideal gas is easy to define: It is a system of particles that are free to move (within the confined space defined by the volume), that never interact with each other. This physical system results in an equation of state (called the *ideal gas law*) that relates the four variables mentioned above through the following functional dependence:

$$P = f(V, N, T) = \frac{N k_B T}{V}, \qquad k_B = 1.38 \times 10^{-23} \frac{J}{K} \tag{5.5.2}$$

where k_B is called the *Boltzmann constant*. [*Yes, this is the same Boltzmann mentioned in a previous section on radiative heat transfer. To say that he was a giant in this field would be an understatement.*]

Alert

Note that this equation of state involves temperature as an absolute quantity, rather than just the change of temperature, as we saw previously. For this reason, it is necessary to use an absolute scale (kelvins) rather than a comparative scale (celsius) in calculations involving this formula.

A common alternative to using the particle number as a state variable is using the number of *moles* (n). A mole of a gas is simply defined as a specific number of particles of that gas. That "specific number" is known as *Avogadro's number*: $N_A \equiv \frac{N}{n} = 6.02 \times 10^{23}$.

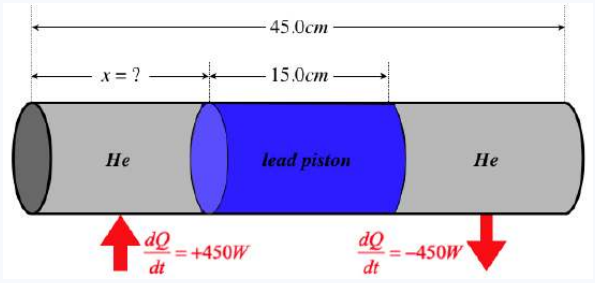
Writing the ideal gas law in terms of the number of moles is therefore a simple conversion:

$$P = \frac{nRT}{V}, \qquad R = N_A k_B = 8.31 \frac{J}{mol\ K} \tag{5.5.3}$$

The constant R is known as the *gas constant*.

Example 5.5.1

One mole of helium gas is injected into each side of a sliding, airtight lead cylindrical piston of radius $8.00cm$ that separates two chambers of a sealed cylinder. The outside of the cylinder is insulated everywhere except where noted below. The cylinder and piston have lengths and cross-sectional areas as labeled in the diagram below. The helium in one of the chambers is heated from outside at a rate of $450\ W$, and the helium in the other chamber expels heat into a cool region. The full system eventually comes to steady-state, so that cooler chamber of gas is losing heat at the same rate as the hotter chamber is receiving it. At this steady state, the pressures of the gases in both chambers are equal to $9.00 \times 10^5 \frac{N}{m^2}$, creating a balance in the forces exerted on the two ends of the piston. The piston therefore remains stationary at an equilibrium point that is a distance x from the end of the cylinder with the chamber receiving the heat. Find the value of x. The thermal conductivity of lead is $35.0 \frac{W}{m\ K}$.



Solution

Heat is being conducted from the left chamber to the right one through the lead piston. We are given the rate of heat transfer, the length of the piston, its cross-sectional area, and the thermal conductivity of lead. We can plug all of these into the steady-state heat equation to find the temperature difference:

$$\frac{dQ}{dt} = -\frac{kA}{L}\Delta T \quad \Rightarrow \quad \Delta T = \frac{0.150m}{\left(35.0 \frac{W}{m\ K}\right)\pi(0.0800m)^2}(450W) = 95.9K$$

We are given the pressure of the gases in the two chambers, and the number of moles of gas in each chamber. Using the ideal gas law, we can determine the difference in the volumes of the chambers in terms of the difference in their temperatures:

$$\left.\begin{array}{l} PV_1 = nRT_1 \\ PV_2 = nRT_2 \end{array}\right\} \quad \Rightarrow \quad V_1 - V_2 = \frac{nR}{P}(T_1 - T_2) = \frac{(1mol)\left(8.31\frac{J}{mol\ K}\right)}{9.00\times10^5\frac{N}{m^2}}(95.9K) = 885cm^3$$

Now that we know the difference in the two volumes, we can combine this with the sum of the two volumes to get the volume of the left side:

$$V_1 + V_2 = \pi(8.00cm)^2\,(30.0cm) = 6030cm^3 \quad \Rightarrow \quad V_1 = \frac{1}{2}\left[(V_1 - V_2) + (V_1 + V_2)\right] = \frac{1}{2}\left[\left(885cm^3\right) + \left(6030cm^3\right)\right] = 3460cm^3$$

The length of this chamber is its volume divided by its area:

$$x = \frac{3460cm^3}{\pi(8.00cm)^2} = 17.2cm$$

As with everything else in physics, the idea of an ideal gas is a *model*. It works pretty well for gases in most real-world circumstances, but it is by no means the only model. Another model treats the gas particles as though they are tiny hard spheres that can bounce off each other. This model may work better in cases of larger molecules and/or higher densities, for example. With a new model comes a new equation of state, and in this case the governing equation is known as the *van der Waals equation*, which is significantly more complex than the ideal gas law:

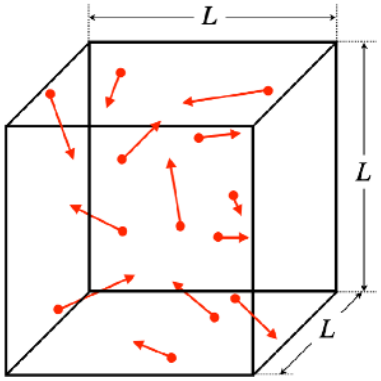
$$\left(P + \frac{an^2}{V^2}\right)(V - nb) = nRT \tag{5.5.4}$$

The constant b takes into account the "hard sphere" aspect of the particles. The volume available to the gas is the volume of the vessel minus the volume occupied by the particles themselves. The constant a accounts for attractive forces (unsurprisingly called *van der Waals forces*) between the particles. When the particles attract each other, they don't strike the walls of the container so hard, and the pressure measured by force-per-area on the walls is lower. Notice that taking the limit as $a,\ b \to 0$ returns the equation of state back to the ideal gas law. When the particle radii are negligible and interactions forces vanish, then the conditions for an ideal gas are met.

Kinetic Theory of Gases

One of the most impressive aspects of the study of thermodynamics lies in how it is possible to use a simple microscopic model of how a large number (10^{23}) of particles in a gas behave to derive some very specific relationships between *macroscopically* measurable quantities. We see this in action in the following application of the *kinetic theory of gases*. We will assume a gas is ideal – that the particles do not interact with each other – and that the gas is trapped within a cubical enclosure.

Figure 5.5.1 – Ideal Gas in a Cubical Container



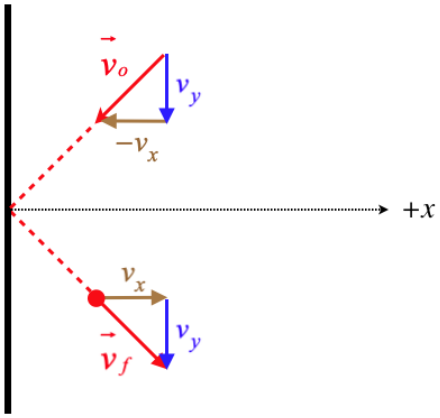
Here are some further assumptions we will make, beyond that of the particles not interacting with each other:

- particles have random positions, speeds, and directions of motion
- no energy lost to the walls (particles collide elastically with walls)
- walls are smooth, so force between particles and wall is perpendicular to wall
- gas is monatomic (individual atoms, not molecules)

We will eventually loosen the last of these constraints, but the others are reasonable and necessary to do the derivation that follows. The assumption that the walls are smooth is not necessary for the final result (nor is the use of a cubical container), but it does make the analysis that follows easier. We will not prove that these simplifying assumptions (smooth surfaces and cubical container) are unnecessary, but at a minimum it should be noted that experimental evidence confirms that the final result works for more general circumstances.

There are literally a trillion-trillion particles in this box, so looking at what they do individually might seem a bit pointless, but in fact we will have the powerful ally of *averaging* on our side, as you will see. So we forge ahead, looking at the effect of a single particle as it reflects off a wall of the container, hoping to parlay the information we gain into some understanding of pressure, which is manifested as the gas pushing on the walls of the container...

Figure 5.5.2 – Particle Reflection Off Container Wall



A particle that comes into a wall of the container and reflects elastically will exit with the same speed that it has coming in. With the wall being "smooth," there is no force on the particle parallel to the wall, so the component of the particle's velocity parallel to the wall remains unchanged ($v_y(before) = v_y(after)$). The elastic collision ensures that the total speed of the particle is unchanged ($\left|\vec{v}_o\right| = \left|\vec{v}_f\right|$). The combination of these two facts means that the component of the particle's velocity in the direction perpendicular to the wall (depicted in the figure above as the x-axis) is the same before and after the collision with the wall, but in opposite directions.

This particle therefore experiences a force from the wall in the $+x$-direction that results in a change of momentum equal to $2mv_x$. Newton's third law tells us that the wall experiences the same force in the $-x$-direction from the particle. If we average this force over a short time that spans the period from shortly before the collision to shortly after, we get:

$$(\text{average force on particle over timespan } \Delta t) = \frac{1}{\Delta t}\int_{t_o}^{t_f} \vec{F}\, dt \tag{5.5.5}$$

From the impulse-momentum theorem (a fancy version of Newton's second law), we can replace the time integral of the force with the change in momentum over that timespan:

$$(\text{average force on particle over timespan } \Delta t) = \frac{\Delta \vec{p}}{\Delta t} = \frac{2mv_x}{\Delta t}\hat{i} \tag{5.5.6}$$

This particle will strike walls other than the two that are perpendicular to the x-axis, but for now we will focus only on the component of the particle's motion along the x-axis. Suppose we wish to know the force that the left wall exerts on this particle averaged over all time. Well, after it bounces off the wall once, it will travel across the box, strike the other wall, and come back again. The x-component of the particle's velocity won't change at the other wall either, so we know exactly how long it takes the particle to make a round trip – the total distance divided by the speed:

$$\text{time of round trip} = \frac{2L}{v_x} \tag{5.5.7}$$

With the particle exerting the same force periodically, the average force exerted on this particle over all time is found directly from the last two equations above (we will remove the unit vector from here on, as the direction is clear):

$$\text{average force on particle over all time} = \frac{2mv_x}{2L\ /\ v_x} = \frac{1}{L}mv_x^2 \tag{5.5.8}$$

We now note that this is the average force exerted on just one of the many particles present, and with the particles randomly distributed (one of our assumptions), it is clear that the wall is constantly being pelted with particle collisions. The average force exerted on each particle by the wall is equal to the average force exerted on the wall by each particle (Newton's third law), and since some number of the randomly-distributed particles are constantly hitting the wall, there is no reason to expect that the total force on the wall will fluctuate over time. Therefore the force on the wall *by the gas* is just equal to the sum of the average forces exerted by all N particles:

$$\text{force on wall by gas} = F = \frac{1}{L}\sum_N mv_x^2 \tag{5.5.9}$$

We now define the pressure of the gas as the force it exerts on a surface per unit area of that surface. In the case of the wall of the container, the area is L^2, so we have for the pressure of the gas:

$$P = \frac{F}{L^2} = \frac{1}{L^3}\sum_N mv_x^2 = \frac{1}{V}\sum_N mv_x^2 \ , \tag{5.5.10}$$

where V is the volume of the cubical container.

Nothing about what we have found here is unique to the walls of the container that are perpendicular to the x-axis, so we also have:

$$P = \frac{1}{V}\sum_N mv_y^2 = \frac{1}{V}\sum_N mv_z^2 \tag{5.5.11}$$

Now we can add the last three equations to each other, and to obtain:

$$3P = \frac{1}{V}\sum_N \left(mv_x^2 + mv_y^2 + mv_z^2\right) = \frac{2}{V}\sum_N \left(\frac{1}{2}mv^2\right) \tag{5.5.12}$$

The particles in this gas do not interact, so the only form of energy they possess is kinetic energy. The sum of the kinetic energies of all the particles is the total internal energy of the gas, which gives us a result that no longer includes any reference to individual particles:

$$3P = \frac{2U}{V} \quad \Rightarrow \quad PV = \frac{2}{3}U \tag{5.5.13}$$

Given that this is an ideal gas which also satisfies the idea gas law, we can write the internal energy in terms of the temperature:

$$U = \frac{3}{2}nRT \tag{5.5.14}$$

We have been saying for awhile that temperature provides a measure of thermal energy, and now we finally have a formula that gives us exactly how these quantities are related. It is truly remarkable that such specific conclusions can be drawn about the macroscopic state of a gas from such simple assumptions about the microscopic behavior of the particles.

Average Particle Speed in a Gas

We can use Equation 5.5.2 and Equation 5.5.14 to draw another conclusion about the particles in this gas:

$$U = \frac{3}{2}Nk_BT \quad \Rightarrow \quad u \equiv \frac{U}{N} = \frac{3}{2}k_BT \tag{5.5.15}$$

This tells us that the *average energy per particle* u is a constant times the temperature of the gas. Keep in mind that the particles' motions are randomly (but not uniformly) distributed, so while the average particle has this energy, the actual particles have a range of energies.

With the average kinetic energy per particle, we can determine a sort of average velocity of particles in the gas. There are many sorts of averages, and in this case the type we are referring to is called the *root-mean-square*, or *rms* velocity, so-named because its calculation involves taking the square root of the mean of the square of the velocity:

$$v_{rms} = \sqrt{\langle v^2\rangle} = \sqrt{\frac{2}{m}}\sqrt{\frac{1}{2}m\langle v^2\rangle} = \sqrt{\frac{2}{m}}\sqrt{\langle KE\rangle} = \sqrt{\frac{2}{m}}\sqrt{\frac{3}{2}k_BT} = \sqrt{\frac{3k_BT}{m}} \tag{5.5.16}$$

So the rms speeds of the particles in an ideal gas increase as the square root of the temperature. Also, if the gas is a mixture of particles of different masses, the heavier particles have a lower rms speeds.

Alert

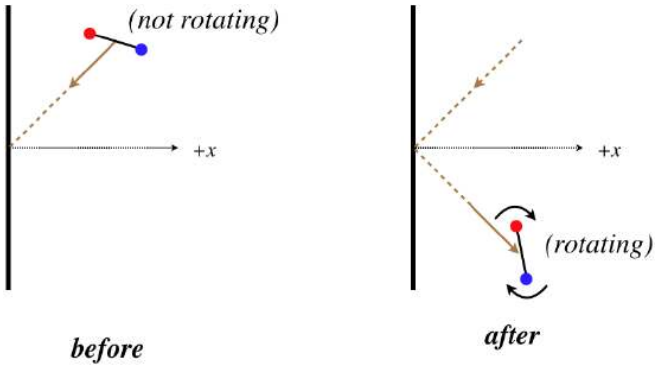
The root-mean-square speed of particles in a gas should not be confused with the "usual" definition of "average," where the speeds of all the particles are added together and the sum is divided by the number of particles. To see the difference, consider a "gas" consisting of two particles, one of which is stationary, and the other moving with a speed of $2v$. These particles have an average speed of v, and an rms speed of $\sqrt{2}v$. The main reason for preferring to use the rms value of speed is that the rms speed is well-defined by the total energy of the gas and the particle number. This is not true of the standard average speed – many average speeds are possible for the same total energy and particle number.

5.6: Equipartition of Energy

Energy Modes

Reviewing our work in the previous section in kinetic theory, one might ask, “Where did we use the assumption that the particles are monatomic?” The answer to this question certainly isn’t obvious. Imagine for a moment that we are instead talking about a *diatomic gas*, where the particles look more like dumbbells than tiny point particles. When one of these particles strikes the wall, one could easily imagine it going from a state of no rotation to one where it is rotating.

Figure 5.6.1 – Diatomic Molecule Bouncing Off a Container Wall



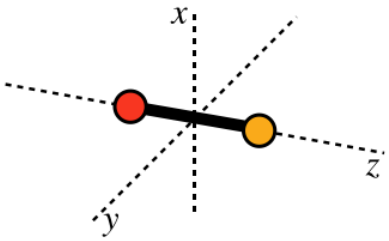
Why would this invalidate the derivation of the previous section? By assumption, the molecule cannot gain or lose energy when it strikes the wall, and we used that fact to show that the speed of the molecule stays constant, but if the molecule can now gain *rotational* kinetic energy upon striking the wall, then energy conservation requires that the kinetic energy that comes from its linear motion be reduced.

While we can't use the result of our previous work as it stands for diatomic molecules, we can modify things slightly so that we once again get a useful answer. This modification involves the concept of *modes* available to the energy. In the monatomic case, the motion of the particles involved three independent *degrees of freedom* – motion along the x, y, and z directions. Looking back at the derivation of the previous section, we see that each of those modes of freedom accounted for an equal amount of internal energy, equal to $\frac{1}{2}nRT$. It turns out that in general the internal energy of a system divides is itself (on average) equally amongst all the available modes. This phenomenon is known as the *equipartition of energy theorem*.

The best way to track these modes is to imagine energy going into one of them at a time, without changing the energy in the previous ones. So for example, a particle can be moving in the x direction only, and it has a kinetic energy. We could then give it a “kick” in the y direction, and this addition of energy doesn’t change the contribution to the kinetic energy by the motion in the x direction. Once we have kicked it in all 3 directions, there is no other way to kick it that doesn’t give it additional energy in one of the modes already used, so the number of *translational modes* ends at three.

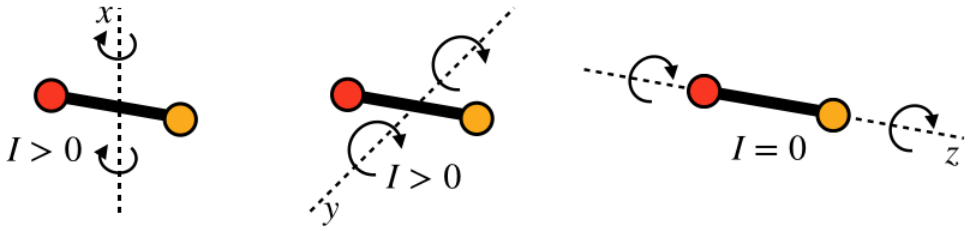
As we have seen, monatomic gases have only these translational modes, because they have negligible extension in space, which equates to a negligible moment of inertia, which in turn results in zero rotational energy. Diatomic particles clearly do not suffer from having zero moment of inertia, so they do have rotational modes available for kinetic energy. For any three-dimensional object, there are three axes around which any object can rotate, and it is generally convenient to choose their principal axes when discussing such rotational degrees of freedom. For a diatomic molecule which we model as two point particles joined by a rigid rod, these axes look like:

Figure 5.6.2 – Principal Axes of a Diatomic Molecule



Rotation around two of these axes features a non-zero moment of inertia, but around the z-axis rotation is similar to that of a monatomic particle – no extension in space relative to that axis, and therefore zero moment of inertia. Therefore there are really only *two* rotational modes into which kinetic energy can be distributed for diatomic gases.

Figure 5.6.3 – Two Rotational Modes for Diatomic Particles



So with this rigid-rod model of diatomic particles, the total number of modes into which energy can be distributed is five - three translational and two rotational. With each of these modes getting an average energy of $\frac{1}{2}k_BT$, the average energy per particle for this type of ideal gas is:

$$u = \frac{5}{2}k_BT \qquad \text{[rigid-rod model of diatomic ideal gas]} \tag{5.6.1}$$

If the particles in the gas are composed of enough atoms that the single axis of symmetry no longer applies, then the $I = 0$ axis no longer exists, and there are three rotational modes available, resulting in an average energy per particle equal to $3k_BT$. We therefore have the following relationships between internal energies and temperatures for various types of ideal gases:

$$\begin{aligned} U &= \tfrac{3}{2}nRT \qquad \text{monatomic} \\ U &= \tfrac{5}{2}nRT \qquad \text{diatomic with rigid bonds} \\ U &= \tfrac{6}{2}nRT \qquad \text{polyatomic with rigid bonds} \end{aligned} \tag{5.6.2}$$

Example 5.6.1

One mole of monatomic nitrogen (N) and one mole of diatomic nitrogen (N_2) have equal internal energies. How do their rms velocities compare? Assume the mass of N_2 is twice that of N, and that N_2 can be modeled with a rigid-rod dumbbell.

Solution

With equal internal energies, the two gases cannot have the same temperature. Their temperatures are related by:

$$\left.\begin{aligned} U_1 &= \tfrac{3}{2}RT_1 \\ U_2 &= \tfrac{5}{2}RT_2 \end{aligned}\right\} \quad \Rightarrow \quad T_2 = \frac{3}{5}T_1$$

The rms velocity of N_2 is:

$$v_2 = \sqrt{\frac{3k_BT_2}{m_2}}$$

We can use the temperature relationship above and the mass relationship $m_2 = 2m_1$ to write this rms velocity in terms of the rms velocity of N:

$$v_2 = \sqrt{\frac{3k_B\left(\frac{3}{5}T_1\right)}{2m_1}} \quad \Rightarrow \quad v_2 = \sqrt{\frac{3}{10}}\sqrt{\frac{3k_BT_1}{m_1}} = \sqrt{\frac{3}{10}}v_1$$

Vibrational Modes

As we have seen in a previous section, the bonds between atoms can be effectively modeled as springs. If we do this for diatomic particles, then there is yet another independent way to introduce energy into the system – through vibrations of the springs. What is especially interesting about vibrational modes is that they involve not only an additional kinetic energy degree of freedom, but also

one associated with *potential energy*. If we look at a large collection of masses vibrating on springs and do an accounting of the energy, then on average we will find that at any given moment half of the total energy in that collection is in the form of spring potential energy, and half is kinetic energy. Therefore the equipartition of energy theorem for a diatomic molecule that can vibrate allows for two more modes. In this case, we have another relationship between internal energy and temperature:

$$U = \frac{7}{2}nRT \quad \text{diatomic with vibration} \tag{5.6.3}$$

So if these modes exist, why even mention the case of "rigid bonds?" A detailed answer to this question lies outside the scope of this course, but the short answer is this: When we get into the realm of the very small (like discussing bonds between individual atoms), we are forced to turn the discussion to an entirely different (more appropriate) model of nature, called *quantum theory*. In this theory, we find that energy cannot be indefinitely subdivided to smaller and smaller units – it is "quantized." The upshot of this for vibrational modes is that a certain minimum amount of energy is required for this mode to be activated. If this energy is not available (i.e. if the temperature is too low), then we say that the vibrational modes are *frozen-out*.

It turns out that the amount of energy needed to activate these vibrational modes is quite high, compared to typical temperatures we encounter – this happens at temperatures on the order of *thousands* of kelvins. Whenever we deal with diatomic gases, we will use the rigid-rod model (and its associated number of available modes) as the default – if the vibrational modes are in play, then it will be made clear that this is the case, either by saying so explicitly, or noting that the temperature of the gas is extremely high.

Digression: More Quantum Weirdness

Lest the reader think that the freezing-out vibrational modes is where the quantum weirdness ends, it should be noted that when the temperature of a diatomic gas gets sufficiently low (roughly double-digit kelvins), the rotational modes also get frozen-out! It might be possible to conceptualize that a molecule requires a certain minimum amount of energy for it to vibrate, but the idea that the same is true for it to ***rotate*** *is quite bizarre. Quantum theory is full of these truth-is-stranger-than-fiction phenomena.*

Molar Heat Capacity

We now return to our discussion of heat. To keep things simple, we will consider the case when two samples are exchanging heat energy (from the hotter to the colder one, obviously), and *the volumes of both remain constant*. This would obviously be mostly true for solids and liquids (the thermal expansion is a negligible effect), while gases would have to be held in containers with fixed volumes. Under these conditions, all of the heat lost/gained by a sample comes out of/goes into the internal energy of the sample:

$$Q = \Delta U = U_{after} - U_{before} \tag{5.6.4}$$

For the sample that gains/loses heat, Q is positive/negative, and the internal energy goes up/down. In our discussion of heat capacity in Section 5.3, we related the change in a sample's temperature to the amount of heat that the same gains or loses. With our results from kinetic theory and the equipartition of energy theorem, we can determine this heat capacity per mole. For example, for a monatomic ideal gas:

$$Q = \Delta U = \Delta\left(\frac{3}{2}nRT\right) = n\left(\frac{3}{2}R\right)\Delta T \tag{5.6.5}$$

Comparing this to Equation 5.3.6, we see that the molar heat capacity (heat capacity per mole) is a simple constant. It turns out that the assumption we have made (that the gas is confined to a fixed volume) is important to this result. We will look at the more general case later, but for now we will make sure we keep this assumption in mind by adding a subscript to the symbol for the molar heat capacity:

$$C_V = \frac{3}{2}R \quad \text{monatomic ideal gas} \tag{5.6.6}$$

Despite the constraints, (the gas must be ideal, it must be monatomic, and the heat transfer must occur with the gas held at constant volume), the result is nonetheless quite remarkable. This same result works for all inert gases such as helium and neon (because they are always monatomic), as well as for other gases that can be found in a monatomic state, like hydrogen and nitrogen.

We can at least account for some of the constraints, by incorporating the number of available modes into the computation of C_V:

$$C_V = \frac{1}{2}R \times (number\ of\ available\ modes) \tag{5.6.7}$$

While this was approached from the case of ideal gases, the result given above is actually more general, since non-ideal gases, liquids, and solids can be taken into account by including the modes associated with interactions of the particles. For example, if we consider a monatomic solid at a high temperature (so that the vibrational modes are not frozen-out), then the lattice structure will result in 6 vibrational modes: 3 kinetic, and 3 potential (springs). With that the case, the molar heat capacity of a monatomic solid is simply $3R$. We don't even care what substance we are talking about!

How do we reconcile this with the fact that we already know that different substances have different *specific* heat capacities? The answer is that one mole of one substance does not have the same mass as one mole of another. If these two substances have the same number of modes into which energy can be distributed, then they have equal molar heat capacities, but when one wants to convert heat capacity per mole to heat capacity per kilogram, the new constants are no longer equal.

Finally, it should be noted that because C_V takes into account the number of modes available for any substance, we have a shorthand for the relationship between the internal energy and the temperature of a sample:

$$U = nC_V T \tag{5.6.8}$$

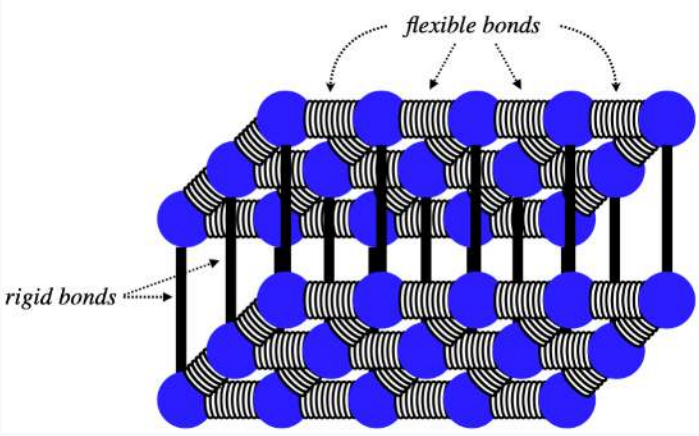
This applies no matter what type of gas (monatomic, diatomic, etc) we are talking about, because the molar heat capacity (at constant volume) takes into account the number of modes.

Alert

This is an equation of state, relating three state variables to each other (U, n, and T) through a physical constant (C_V) that depends only upon the number of modes available to the gas in question. While we got to this point by talking about a process (heat transfer with volume held constant), that process is irrelevant to the meaning of this equation. This is often a point of confusion, and we will return to it later, when we consider processes where volume is not held constant.

Example 5.6.2

Consider a 1-mole sample of a solid which forms a simple monatomic rectangular lattice, as in the diagram. The types of bonds formed and the temperature of the substance are such that the horizontal bonds are all free to vibrate, while the vertical bonds remain rigid. This sample is brought into thermal contact with 2 moles of a monatomic ideal gas at a different temperature (and the combination of the two are insulated from their surroundings). Assume that the gas and solid neither change phase nor volume during this heat exchange. Find the ratio of the temperature changes ($\Delta T_{gas} / \Delta T_{solid}$) of the gas and solid from the time when they are brought into contact to when they reach thermal equilibrium.



Solution

The heat that enters or exits the solid exits or enters the gas, so they differ only by a sign. Writing the heats in terms of the molar heat capacities and applying this fact gives:

$$Q_{into\ gas} = -Q_{into\ solid} \quad \Rightarrow \quad n_{gas} C_{gas} \Delta T_{gas} = -n_{solid} C_{solid} \Delta T_{solid} \quad \Rightarrow \quad \frac{\Delta T_{gas}}{\Delta T_{solid}} = -\frac{n_{solid} C_{solid}}{n_{gas} C_{gas}}$$

The number of moles of each is given, so all that remains is to determine the molar heat capacities, which we can get from Equation 5.6.7. For a monatomic ideal gas the number of modes is 3. This solid has horizontal vibrational degrees of

freedom, giving it 4 modes (two KE and two PE). It has no vertical degrees of freedom, nor does it have translational or rotational degrees of freedom, so its total number of modes is 4. Plugging all of this in above (including the numbers of moles of each), we get:

$$\frac{\Delta T_{gas}}{\Delta T_{solid}} = -\frac{(1)(4)}{(2)(3)} = -\frac{2}{3}$$

The temperature of the gas changes two-thirds as much as the temperature of the solid, and in the opposite direction.

While we can draw some conclusions about liquids and solids, as we move on into thermodynamics, we are going to focus on ideal gases, for a couple of reasons. First, gases (unlike liquids and solids) can readily change volume, which gives us a great deal more depth of study. And second, we have a state equation we can work with (the ideal gas law) that relates the various state variables.

5.7: Thermodynamic Processes

Definition of a Thermodynamic Process

Up to now, we haven't spent a lot of time on the *dynamics* part of thermodynamics. So far our notion of dynamics has been limited to talking about heat transfer, and how it relates to temperature change. Indeed, in Section 5.5 we concluded that we can only really deal with equilibrium states, which seems to directly contradict the notion of examining dynamics, which requires *changes* in the state.

We get around this apparent conundrum through the introduction of something called a *quasi-static process*. The idea is that a system can evolve from one equilibrium state to a neighboring one (i.e. one infinitesimally close) slowly, so that at any instant in time the state variables are in perfect balance are are not "leaning" toward change. A nice analogy for this idea of a quasi-static process is someone balancing on a large ball. When they are balanced, they are in equilibrium. But they would like to move the ball along, so they move their feet ever so slightly, and the ball rolls a tiny bit. The ball stops, and the person is again in equilibrium. They can continue this process to get across the floor, inching from one equilibrium state to another, never endangering themselves. Alternatively, they could just “go for it” and start moving their feet fast. In this case, they will have to lean, and cannot simply stop whenever they like at an equilibrium position. The slow, equilibrium-to-equilibrium process is what we call quasi-static.

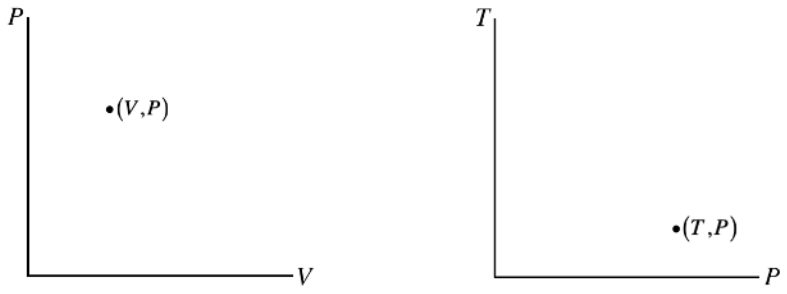
Another aspect of processes that we need to define is *reversibility*. A process is reversible if it needs to be "coaxed" into occurring – if the process occurs spontaneously from the initial conditions, then it is said to be *irreversible*. It's certainly true that if a process is not quasi-static, then it is irreversible – the person that is off-balance on the ball has no ability to stop the ball or reverse its direction at any instant, so the system evolves from a given state without coaxing. So a process being quasi-static is *necessary* for it to be reversible, but it is not *sufficient*.

To see this, suppose we have a large imbalance between two adjacent systems. For example, one system may be significantly hotter than the other. If these are put into direct contact, then the heat will transfer very fast – not quasi-statically – and since this process proceeds spontaneously, it is irreversible. But now suppose we introduce a transfer conduit to conduct the heat between the two systems. The thermal conductivity of the conduit medium (and/or the length of the conduit) can be made arbitrarily-small, slowing the heat transfer process to a trickle into the colder system. This would be a quasi-static process, because at any moment if we cut off the transfer (insert insulation), the systems don't have to "settle into" equilibrium – they are already there, because the changes have been so slow. However, if we again remove the insulation, these systems do not remain in the same state – they spontaneously start evolving (albeit slowly) again. They are not equally-likely to go in either direction – the heat spontaneously transfers from hot to cold. We will discuss this idea of reversibility further in future sections.

Process Diagrams

One way to analyze processes is with *process diagrams*, which are depictions (graphs) of processes through equilibrium states. We already said that in general, equations of state involve four state variables (one state variable dependent upon three others), so if we want to graph general processes, we will need many dimensions. We will not be dealing with processes during which the particle number changes, so all of our diagrams hold that number fixed. That leaves three variables with two of them independent, and the third being derived from the other two through the equation of state. The two independent variables that we choose can be plotted on a pair of axes, and the point thus plotted will represent a specific thermodynamic state. As we must use absolute temperature, none of the state variables (such as pressure, volume, and internal energy) can ever be negative, so these plots only require one quadrant of the axes.

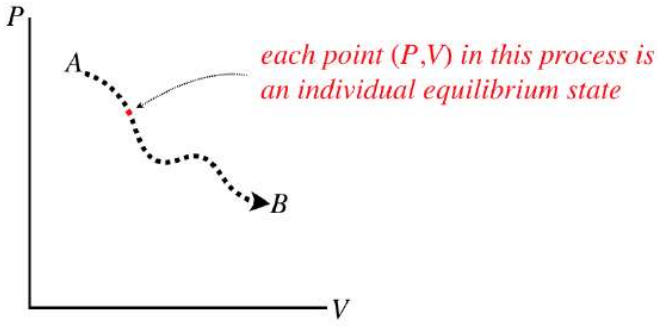
Figure 5.7.1 – Equilibrium Thermodynamic States Plotted on Two Axes



The points on these diagrams, in conjunction with separate information about the number of particles and an equation of state (like $PV = nRT$) completely define an equilibrium state (i.e. the values of all of the state variables) of the system.

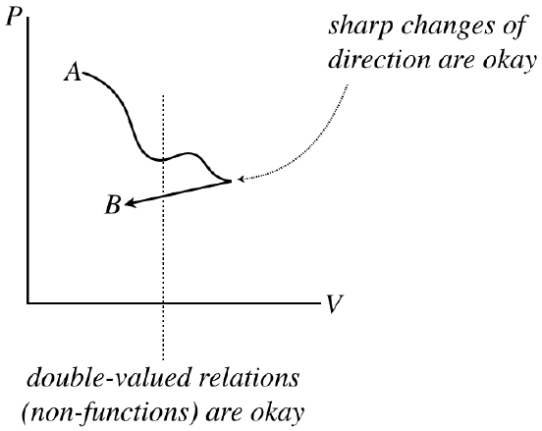
A quasi-static process would be represented on one of these diagrams as a continuous curve (along with an indicated direction), because in such a process a system changes from one equilibrium state to another that is infinitesimally close. If we want to know what state the system is in at any point during a process, we just read off that point's values. The process does not have any "momentum," meaning that if we stop it at any point, it is not inclined to continue to the next point – it is an equilibrium state.

Figure 5.7.2 – Every Point on a Process Diagram is an Equilibrium State



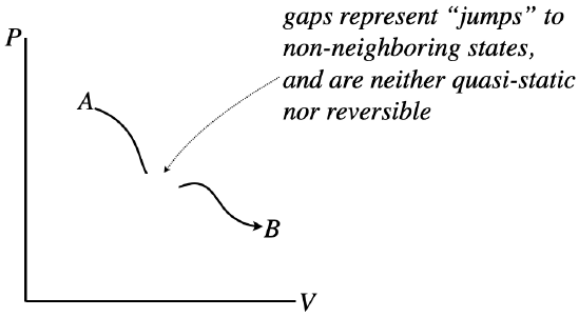
Note that these processes *are not necessarily functions* – it is perfectly acceptable for a process to circle back on itself. They also don't need to be differentiable (smooth), because each state along the way has no memory of the state that comes before, or anticipation of the state that follows.

Figure 5.7.3 – Process Diagrams Do Not Need to be Differentiable Functions



However, the graph of the process *does* need to be continuous. Sudden jumps represent states suddenly changing to other states that are not nearby, which can only be achieved by going through a non-equilibrium state – such a process is not quasi-static. Another way to think of this is in terms of reversibility. Gaps involve jumps between states, and without any "bread crumbs," showing the system the way back, there is no way to retrace steps, and the process is therefore not reversible.

Figure 5.7.4 – Process Diagrams Must be Continuous to be Quasi-Static



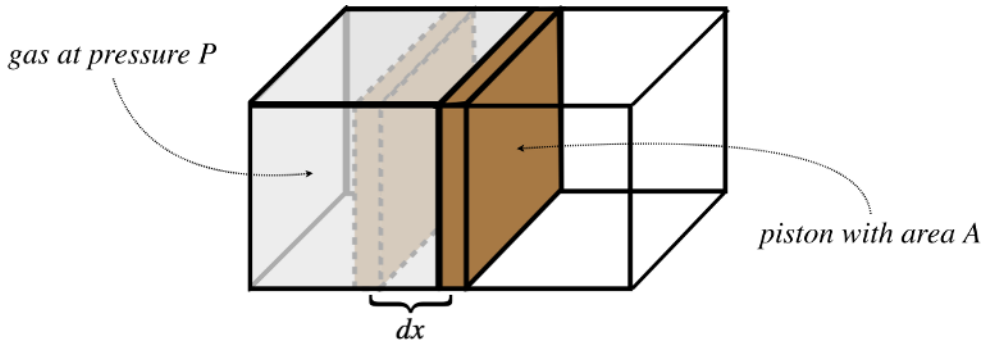
Work Done By a Gas

We said earlier that the two “oddballs” in the field of thermodynamics are heat and work, and we’ve talked a bit about heat, so it’s time we explore the topic of work a bit more. We already know something about it from Physics 9A:

$$W\,(A \rightarrow B) = \int_A^B \vec{F} \cdot \vec{dl} = \int_A^B \left|\vec{F}\right| \left|\vec{dl}\right| cos\theta \tag{5.7.1}$$

How does this apply to our study of the thermodynamics of ideal gases? Well, gases exhibit pressure, which can result in the exertion of a force, so all we need to do is conceive of a case where gas pressure moves something. To this end, consider a gas confined by a container with a piston:

Figure 5.7.5 – Work Done on a Piston by a Confined Gas



The confined gas exerts a force on the piston that equals the pressure of the gas multiplied by the area of the piston. The forces acts to push the piston outward a small distance, thereby doing a small amount of work on it, giving:

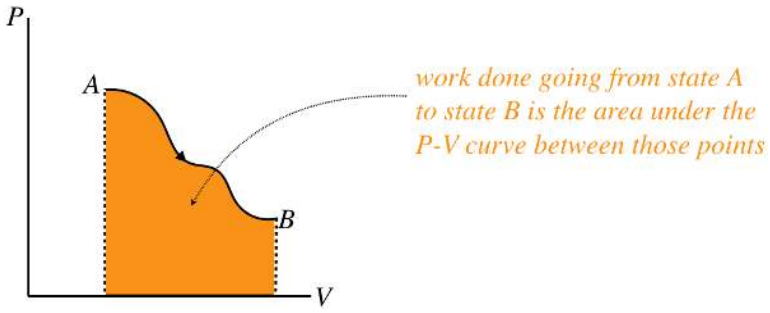
$$dW_{by\ gas} = F\ dx = PA\ dx \tag{5.7.2}$$

But the product $A\ dx$ equals the small change in the gas's volume dV, allowing us to write the work done in terms of two state variables. As along as the piston expands slowly, the gas will go from one equilibrium state to another in a quasi-static manner, giving a total amount of work equal to:

$$W_{by\ gas} = \int_A^B P\ dV \tag{5.7.3}$$

In terms of the process diagram of pressure vs. volume, this integral is simply the area under the curve.

Figure 5.7.6 – Work Done by Gas is Area Under PV Curve



Notice that the *sign* of the work is also important. If the gas is compressed rather than expanded, then the process goes right-to-left in the PV diagram, and the integral is negative. We therefore assert the following sign convention:

work done by an expanding gas has a (+) sign, while work done on a gas to compress it has a (–) sign

For comparison purposes, let's restate the sign convention for heat transfer:

heat transferred into a gas has a (+) sign, while heat transferred out of a gas has a (–) sign

The First Law of Thermodynamics

Now that we have placed work and heat into the big picture of thermodynamics, we can apply a principle we have known since early in Physics 9A. We do this by taking an accounting of all the energy associated with a thermodynamic system. We know that all of the energy *within* the system (all kinetic and potential energy) is accounted-for in the internal energy U, which is a state function. Work and heat both either bring energy into or take energy away from the system. According to the law of conservation of energy, whatever the net result of the energy transfer is must be reflected in the change of internal energy. Since work is done by the gas, a positive amount of work done is energy that comes out of the gas, while a positive amount of heat is energy that enters the gas. The change in the gas's internal energy is positive when net energy comes in and negative when net energy exits, so it is given by:

$$\Delta U = Q - W \tag{5.7.4}$$

This simple expression of energy conservation is known as the *first law of thermodynamics*.

Alert

*Chemistry and physics classes typically approach the sign conventions for the first law differently. The symbol W in the equation above represents the work done by the gas. Chemists generally use the symbol W to represent the work done **on** the gas, which has the effect of changing the sign of W in the equation for the first law. Obviously the physical meaning is the same in both cases, but the choice in sign convention likely comes from a difference in emphasis. The convention from chemistry has the advantage of symmetry between the sign conventions for heat and work (both are positive when energy is going into the gas). Also, chemists are more likely to concern themselves with the effects of work on a gas than the effects of gas on its surroundings. One advantage to the sign convention used here is that the work integral in terms of pressure and volume doesn't require a negative sign – the work done by a gas is more intuitive than the work done on it. Another is that we will be discussing engines, which have the role of converting heat taken in into work put out. The "physics sign convention" is more convenient for a process where the amount of heat in equals the amount of work out. Anyone who finds these differences in sign conventions confusing should just always think "work done **by** the gas" when they see W in the formulas encountered in this text.*

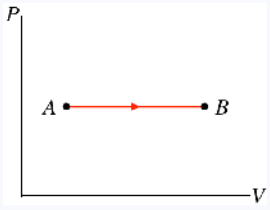
As simple as this law is (after all, it is just conservation of energy), it has some interesting properties. The most striking is that we have a change in a state variable (U) on one side of the equation (which depends only upon the starting and ending states of the process), while on the other side are two quantities that depend upon the path taken. This tells us that different paths between two states result in different amounts of each type of energy transfer, but the final energy change is the same, as it only depends upon the endpoints.

Example 5.7.1

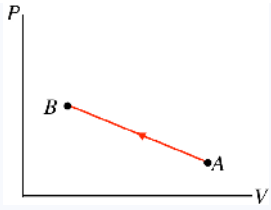
For each of the straight-line processes for an ideal gas shown below, answer the following questions:

- *Work done: on the gas, by the gas, or can't tell?*
- *Heat transferred: into the gas, out of the gas, or can't tell?*
- *Temperature change: gas gets hotter, gas gets colder, or can't tell?*

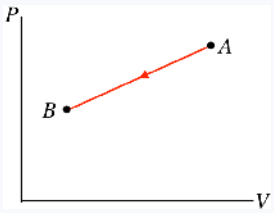
a.



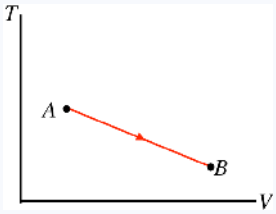
b.



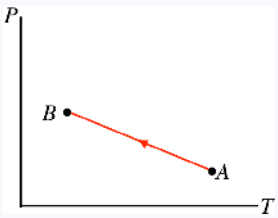
c.



d.



e.



Solution

For all of these graphs, we have the following tools to work with:

$$PV = nRT \qquad U = nC_V T \qquad W = \int_A^B P dV \qquad \Delta U = Q - W$$

a. The process goes from left to right, so the positive work is done by the gas. The pressure doesn't change during this process, but the volume goes up, so the quantity PV must increase from A to B. With the number of moles not changing, the ideal gas law tells us that the temperature must go up. The increase in temperature means that the internal energy goes up, or $\Delta U > 0$. We already said that $W > 0$, so from the first law we find that $Q > 0$, which means that heat enters the gas.

b. The process goes from right to left, so the work done by the gas is negative, which means work is done on the gas. With the pressure rising and the volume falling, there is no way to tell what happens to the temperature without more details of the endpoints. Without knowing whether the temperature goes up or down, there's no way to tell what happens to the internal energy, which gives us no way to use the first law to determine whether heat is entering or leaving the system.

c. Right-to-left process $\rightarrow$ work done on gas. Both the pressure and the volume go down, which means that PV goes down, allowing us to use the ideal gas law to conclude that the temperature also goes down. The decrease in temperature means $\Delta U < 0$, and since $W < 0$, the first law requires that $Q < 0$, so heat leaves the system during this process.

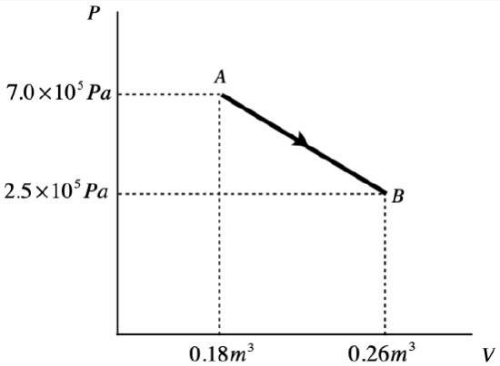
d. Note that this is process expressed in terms of temperature vs. volume. The area under this curve does not give us the work done! Though we don't know what happens to the pressure during this process, the volume nevertheless increases, which means that the gas is expanding and pushing a piston outward, so positive work is done. The temperature (measured on the

vertical axis) clearly drops during the process. We know that $W > 0$ and $\Delta U < 0$, so the first law can't tell us what happens with the heat without more details about the endpoints A and B.

e. This process is expressed in terms of pressure and temperature. It isn't immediately clear from the graph whether the volume is expanding or not, so we will have to wait on drawing a conclusion about work. The temperature (measured on the horizontal axis) is dropping because the process is right-to-left. The pressure is rising while this occurs, so from the ideal gas law, the volume must get smaller during this process. When the volume gets smaller, the gas is compressed, so work is done on the gas. With $U < 0$ (temperature drops) and $W < 0$ (volume decreases), the first law tells us that $Q < 0$, so heat is leaving the system.

Example 5.7.2

A monatomic ideal gas undergoes a quasi-static process from state A to state B, illustrated in the PV diagram below. The process forms a straight line on the P-vs-V graph.



a. Calculate the work done in this process.
b. Find the change in internal energy from state A to state B.
c. Find the quantity of heat transferred into or out of the system during this process and indicate whether the heat goes in or comes out.

Solution

a. The work done during a process is the area under the P-vs-V curve, so all we need to do is compute the area of the top triangle and the bottom rectangle and add them:

$$\left.\begin{aligned} top\ triangle\ area &= \tfrac{1}{2}bh = \tfrac{1}{2}\Delta P\Delta V = 1.8\times10^4 J \\ bottom\ rectangle\ area &= P_{min}\Delta V = 2.0\times10^4 J \end{aligned}\right\} \quad \Rightarrow \quad W = 3.8\times10^4 J$$

b. The gas is monatomic, so the change in internal energy is:

$$\Delta U = \frac{3}{2}nR\Delta T = \frac{3}{2}(nRT_B - nRT_A)$$

Putting in the ideal gas law gives:

$$\Delta U = \frac{3}{2}(P_BV_B - P_AV_A) = -9.2\times10^4 J$$

c. The amount of heat transferred comes directly from the first law of thermodynamics:

$$\Delta U = Q - W \quad \Rightarrow \quad Q = \Delta U + W = -9.2\times10^4 J + 3.8\times10^4 J = -5.4\times10^4 J$$

The negative sign indicates that this heat is lost by the system. Given that work is also done by the gas (work is positive), it isn't surprising that the internal energy goes down (and with it the temperature of the gas).

Quasi-Static vs. Non-Quasi-Static Processes

We know that work and heat only represent *exchanges* of energy, or *changes* to a system – they are not values stored in the state of a system. This means that a single point on a process diagram does not define an amount of work or heat (in the case of work, you cannot define an area under a point!). In fact it turns out that *every time* a process occurs, it occurs because of either an exchange of heat or work or both. Instead of thinking of heat and work exchange as a *result* of a process, we can think of them as the *cause* of a process.

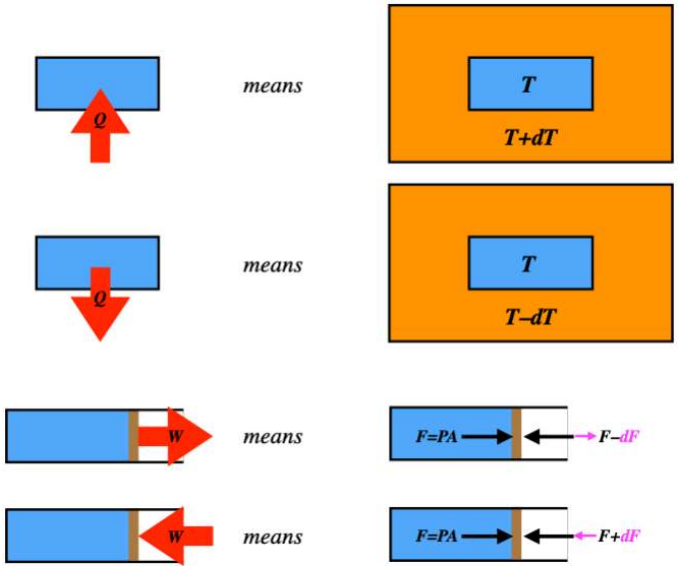
> **Alert**
>
> *Note that it is possible to change states without doing work, but in that case heat must be transferred, and it is possible to change states without heat being transferred, but in that case work must be done.*

Let's start with a piston at equilibrium, so the force on it due to the pressure of the gas is exactly balanced by a force pushing inward from the outside, holding it in place. Suppose we drop the force from the outside by a few Newtons – what happens? The piston expands outward until the pressure drops enough to rebalance the forces. The force imbalance causes the piston to *accelerate*. On its way to its final position, it is *not in equilibrium*, so this process is not quasi-static! We can't figure out the work, because the pressure of the non-equilibrium gas is not well-defined. Furthermore, the piston gains some kinetic energy, so some of the work goes into that. It's a total mess. This all came about because we didn't control the process from one equilibrium state to the next. This control is only achievable if the work is done *incrementally*. That is, any finite difference in forces on the piston causes the process to be non-quasi-static, but if we do just infinitesimal changes in the force, we are okay. Of course, this is not something we can do in practice, but it turns out that doing this kind of analysis is worthwhile nonetheless. For now it is important to understand that such a non-quasi-static process comes from a finite imbalance in the force.

It turns out that the same is true for heat. Recall that heat is the transfer of energy due to a temperature difference (analogous to the force difference for work). A process that involves heat transfer is only quasi-static if it occurs due to an infinitesimal temperature difference. Again, this is not something we can manage in the real world – we usually just put a hot system next to a cold one – but it is useful to use this analysis in the same way that it is useful to study frictionless motion in mechanics.

We will be drawing lots of diagrams that indicate work done and heat transferred, and since we are always assuming quasi-static processes, it is important to have a clear picture of what these diagrams are depicting.

Figure 5.7.7 – Interpreting Heat and Work Exchanges in Diagrams



5.8: Special Processes

Analyzing Special Processes

Next we will discuss some special processes, not only because they are themselves important, but because they are highly instructive. In this discussion, we will emphasize several aspects of each case:

- a physical picture of what is happening, using a gas trapped in a container with a moveable piston
- a schematic representation of what is happening with regard to heat transfer and work done
- an accounting of the fates of the various thermodynamic variables over the course of the process
- relationships between thermodynamic variables in the process due to state equations and the first law
- a PV diagram of the process

One must keep in mind that *all* of these processes are assumed to be quasi-static, which means that when arrows indicate heat transfer or work performed, this occurs due to an infinitesimal imbalance.

By "special processes," we simply mean processes that eliminate from consideration one (or more) of the thermodynamic variables we have discussed so far. These variables include, volume, pressure, temperature, internal energy, heat, and work (particle number is already eliminated for every case, since we have decided to always hold it constant).

Isochoric Process

A process that involves no change in volume is called *isochoric*. With no change in volume, $dV = 0$, there can be no work done on or by the gas, which means that the only exchange of energy possible is through heat transfer, giving one of two physical situations, both including a pegged piston, and one with heat entering and the other with heat leaving.

Figure 5.8.1 – Isochoric Processes – Physical Picture

Having already determined what happens with the non-state variables of heat and work in these processes, let's do an accounting of the changes that occur in the state variables for an ideal gas. We will consider the case of heat energy entering the system – the changes that occur for the case when the system is losing heat should be obvious once this case is understood.

We already know that the volume doesn't change:

$$\Delta V = 0 \tag{5.8.1}$$

With no work done, the first law requires that all of the internal energy change is due to heat transfer, and since the internal energy is proportional to the temperature, we get:

$$\Delta U = Q \quad \Rightarrow \quad \Delta U > 0 \;\; and \;\; \Delta T > 0 \tag{5.8.2}$$

What about the pressure? For this we can use the ideal gas law. We know that the temperature is increasing while the volume (and particle number) remains fixed, so the pressure goes up:

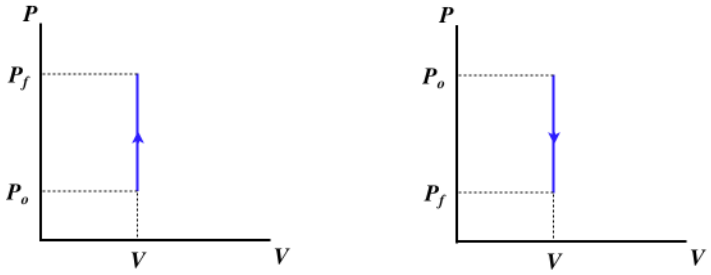
$$PV = nRT \quad \Rightarrow \quad \Delta P = \Delta\left(\frac{nRT}{V}\right) = \left(\frac{nR}{V}\right)\Delta T > 0 \tag{5.8.3}$$

Every process results in its own unique relationship between the non-state variables (work and heat) and changes in the state variables, and in this case we have:

$$\begin{aligned} &monatomic: && Q = \Delta U && = && \frac{3}{2}nR\Delta T && = && \frac{3}{2}(\Delta P)V \\ &diatomic\ (no\ vibration\ mode): && Q = \Delta U && = && \frac{5}{2}nR\Delta T && = && \frac{5}{2}(\Delta P)V \\ &general: && Q = \Delta U && = && nC_V\Delta T && = && \frac{C_V}{R}(\Delta P)V\ , \end{aligned} \tag{5.8.4}$$

where C_V is given by Equation 5.6.7. The PV diagrams for these processes confirms that the work done (the area under the curve) is zero:

Figure 5.8.2 – Isochoric Processes – PV Diagram



A curve for an isochoric process is called an *isochor*.

Isobaric Process

A process that involves no change in pressure is called *isobaric*. We cannot draw a simple conclusion about work as we did in the isochoric case, because in this case the piston is free to move. However, we can use the ideal gas law and the first law to determine the physical picture. Let's take the case where the volume expands at constant pressure. We find from the ideal gas law that this requires the internal energy to go up:

$$\left.\begin{array}{l} P = constant \\ PV = nRT \\ \Delta V > 0 \end{array}\right\} \quad \Rightarrow \quad \Delta T > 0 \quad \Rightarrow \quad \Delta U > 0 \tag{5.8.5}$$

Now applying the first law and noting that the work done to expand the volume is positive, we find:

$$\left.\begin{array}{l} \Delta U = Q - W \\ W > 0 \\ \Delta U > 0 \end{array}\right\} \quad \Rightarrow \quad Q > 0 \tag{5.8.6}$$

Therefore for an isobaric expansion of an ideal gas, heat must pass *into* the system. All of this can of course work in reverse as well, so we have the following physical pictures:

Figure 5.8.3 – Isobaric Processes – Physical Picture

In determining this physical picture, we determined all the state variable changes that occur. All that is left is to relate the heat and work to the state variable changes. The work we can compute easily from the pressure and volume, since the pressure is constant:

$$W = \int_{V_o}^{V_f} P dV = P \int_{V_o}^{V_f} dV = P\Delta V \tag{5.8.7}$$

From the ideal gas law, we can also write this in terms of the temperature change:

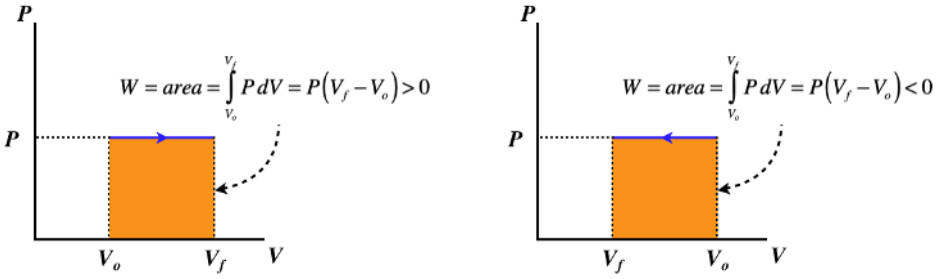
$$W = nR\Delta T \tag{5.8.8}$$

We already have the relationship between the internal energy and temperature, so we can use the first law to write the relationship between the heat transferred and the change in state variables:

$$Q = \Delta U + W = nC_V \Delta T + nR\Delta T = n\left(C_V + R\right)\Delta T = \left(\frac{C_V}{R} + 1\right) P\Delta V \tag{5.8.9}$$

where again C_V is given by Equation 5.6.7. The PV diagrams for these processes are simple enough – constant pressure translates to a horizontal graph:

Figure 5.8.4 – Isobaric Processes – PV Diagram



A curve for an isobaric process is called an *isobar*.

Heat Capacity

In Section 5.6, we noted that the relationship between the heat added and the change in temperature defines the molar heat capacity. It was noted then that this quantity is not a fixed constant for a given gas, but instead depends upon the process. With the the first law of thermodynamics and the fixed relationship between internal energy and temperature, we have:

$$Q = \Delta U + W = nC_V \Delta T + W = nC\Delta T \tag{5.8.10}$$

The molar heat capacity C is defined from the process by determining how the work done in that process is related to the temperature change. The quantity $nR\Delta T$ has units of energy, and is scaled for n moles of a gas, so if we measure the work done by a process as α of these units, we can write the heat transferred in terms of the temperature change as:

$$W = \alpha\,(nR\Delta T) \quad \Rightarrow \quad Q = nC_V\Delta T + n\alpha R\Delta T = n\,(C_V + \alpha R)\,\Delta T \quad \Rightarrow \quad C = C_V + \alpha R \tag{5.8.11}$$

So if we can determine the value of α for the process, we can compute the molar heat capacity. Since heat capacity is a measure of how much heat can be added to a system for a given temperature increase, it makes sense that if energy comes out of the system in the form of work, then more heat can be transferred for the same temperature change than if the work did not remove energy.

We have now seen two special processes, and we have the values of α for both of them. For the isochoric process, there is zero work done, so $\alpha = 0$, giving simply $C = C_V$. Now we see why we originally appended the "V" in the subscript – because this constant happens to equal the *molar heat capacity at constant volume*. For the isobaric process, we got a different value for α. From Equation 5.8.8, we see that for this process $\alpha = 1$, which gives us the *molar heat capacity at constant pressure*:

$$C_P = C_V + R \tag{5.8.12}$$

Notice that this takes into account the number of modes, because that information is contained in C_V. So for monatomic ideal gases, $C_P = \frac{5}{2}R$, for diatomic ideal gases, $C_P = \frac{7}{2}R$, and so on.

Example 5.8.1

One mole of a diatomic ideal gas is confined to a piston and undergoes a quasi-static process where $\frac{1}{4}th$ of the added heat is converted into work. Find the molar heat capacity for this process.

Solution

Plugging $W = \frac{1}{4}Q$ into the first law gives:

$$\Delta U = Q - W = Q - \frac{1}{4}Q \quad \Rightarrow \quad Q = \frac{4}{3}\Delta U = \frac{4}{3}(nC_V\Delta T) \quad \Rightarrow \quad C = \frac{4}{3}C_V$$

Now we plug in for the constant C_V for a diatomic ideal gas (for which, as always, we assume that the vibrational modes are frozen out) to get our answer:

$$C_V = \frac{1}{2}R \times (number\ of\ modes) = \frac{5}{2}R \quad \Rightarrow \quad C = \frac{10}{3}R$$

Gases can be mixtures of varieties (monatomic, diatomic without vibration, etc.), and it can be cumbersome computing the contributions of modes by each, so a more efficient method has been devised to characterize gases in the form of a constant γ, defined as:

$$\gamma \equiv \frac{C_P}{C_V} = 1 + \frac{R}{C_V} \tag{5.8.13}$$

In the cases where the gases are of a single type, this constant tells us exactly what type it is:

$$\begin{aligned} monatomic: &\quad \gamma = \frac{5}{3} \\ diatomic: &\quad \gamma = \frac{7}{5} \\ polyatomic: &\quad \gamma = \frac{4}{3} \end{aligned} \tag{5.8.14}$$

When the gas is a mixture, this constant lands between these values. Note that γ drops in value as more modes become available.

Isothermal Process

A process that involves no change in temperature is called *isothermal*. In this case, a constant temperature means a constant internal energy, and from the first law we get that the heat transferred equals the work done:

$$T = constant \Rightarrow \quad 0 = \Delta U = Q - W \quad \Rightarrow \quad Q = W \tag{5.8.15}$$

From our sign conventions for Q and W, we know that this means that if heat is entering the system, then the gas is doing work, and if heat is exiting the system, then work is being done on the gas. This gives the same basic physical picture as given above for the isobaric process, though the details are distinctly different, as is seen in the effects on the state variables. Let's consider the case of the expanding gas: The volume increases ($\Delta V > 0$), and the temperature and internal energy remains constant ($\Delta T = 0$, $\Delta U = 0$), so from the ideal gas law, the pressure must be going down ($\Delta P < 0$).

Next we want to write the work and heat in terms of the changes in state variables. Fortunately, these two quantities equal each other in this case, so if we compute one, we immediately have the other. We can compute the work done with the integral, because the constant temperature and ideal gas law enables us to write the pressure as a function of the volume:

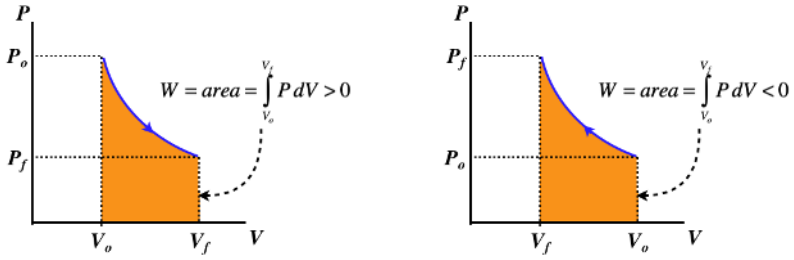
$$\left.\begin{aligned} &Q = W \\ &W = \int_{V_o}^{V_f} P dV \\ &P = (nRT)\frac{1}{V} \end{aligned}\right\} \quad \Rightarrow \quad Q = W = nRT \int_{V_o}^{V_f} \frac{dV}{V} = nRT \ln\left[\frac{V_f}{V_o}\right] = nRT \ln\left[\frac{P_o}{P_f}\right] \tag{5.8.16}$$

There are a number of differences between this process and the others we saw before it:

- The natural logarithm function makes an appearance in the calculation of the heat or work from the state variables, and either the volumes or pressures can be used in the calculation.
- The relationship between heat or work and the state variables does not depend upon whether the gas is monatomic, diatomic, etc. This makes sense, because this process involves no change in the internal energy, so the number of modes available for energy distribution is irrelevant.
- The molar heat capacity is infinite, since any amount of heat can be added and the temperature never changes.

The PV curve is no longer so trivial:

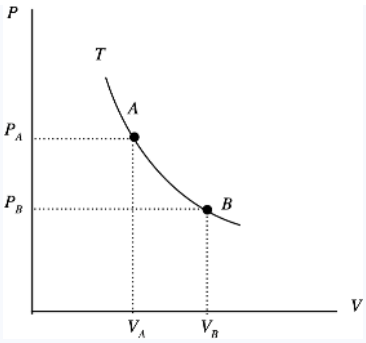
Figure 5.8.5 – Isothermal Processes – PV Diagram



A curve for an isothermal process is called an *isotherm*.

Example 5.8.2

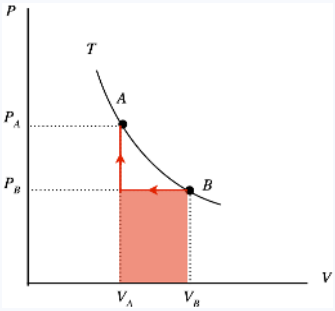
The PV diagram below depicts two equilibrium states (“A” and “B”) of two moles of an ideal gas confined by a piston which both lie on the same isotherm. You are supplied with a meter stick and a thermometer, so you can determine the volume of the piston and the temperature of the gas at any stage. Suppose you take the system from state B to state A. You do this in two stages: First you compress the gas at a constant pressure to the final volume, then you raise the pressure to the final pressure with a fixed volume. You measure the two volumes to be $V_A = 0.130m^3$ and $V_B = 0.160m^3$, and before you start the compression, your thermometer measures the temperature of the gas to be $T = 292K$.



a. Sketch this two-stage process (including an arrow to indicate the direction) on the diagram and use the sketch to compute the total work done on or by the gas (specify which) in joules.
b. Find the total heat that enters or exits the system (specify which) during this two-stage process.
c. Suppose you repeat this two-stage process with two different gases – He and N_2. During the process, which gas will exhibit the greatest difference between its maximum and minimum internal energy, or will both be the same?

Solution

a.



The work done during the second process is zero, so the total work done by the two processes is the shaded area shown under the curve. The process goes right-to-left, so the work done is negative, which means it is work done ***on the system****:*

$$W = P_B\left(V_A - V_B\right)$$

The starting temperature T_B is known, as is the number of moles, and the volume so the starting pressure can be computed, giving us the work:

$$P_B = \frac{nRT_B}{V_B} = \frac{(2mol)\left(8.31\frac{J}{mol\ K}\right)(292K)}{0.160m^3} = 3.03\times10^4 Pa \quad \Rightarrow \quad W = \left(3.03\times10^4 Pa\right)\left(0.130m^3 - 0.160m^3\right) = -910J$$

b. The system begins and ends with the same temperature, since both states are on the same isotherm. Therefore it begins and ends with the same internal energy, which means that all the energy that comes in as work goes out as heat: $Q = -910J$.

c. In both cases the gases pass through the same pressures and volumes, which means they will exhibit the same temperatures throughout (they both satisfy the same ideal gas law). The internal energy, on the other hand, will be different for the two, since its relationship with the temperature is different. For the same temperature change ΔT, the gas with the higher heat capacity will experience the greater change in internal energy. Therefore N_2 (which is diatomic and therefore has 5 modes per particle compared to 3 for monatomic He) exhibits the higher internal energy change.

Adiabatic Process

A process that involves no heat exchange is called *adiabatic*. Like the isochoric case, this makes the physical picture an easy one to describe. One difference that is added to the diagram is insulation of the container in the form of thicker walls, to represent the inability of heat to transfer into or out of the system.

Figure 5.8.6 – Adiabatic Processes – Physical Picture

The first law tells us immediately that the internal energy change is entirely caused by work done on or by the system:

$$\Delta U = Q - W = -W \tag{5.8.17}$$

In the case of the gas expanding, positive work is done, which means that $\Delta U < 0$ and $\Delta T < 0$. With volume increasing and temperature decreasing, the ideal gas law tells us that the pressure must be going down during this expansion. The relationship between the work and the change of state variables is easy to compute from its relationship to internal energy:

$$W = -\Delta U = -nC_V \Delta T = -nC_V \Delta \left(\frac{PV}{nR} \right) = -\frac{C_V}{R} \Delta (PV) = \frac{C_V}{R} (P_o V_o - P_f V_f) \tag{5.8.18}$$

We see that the work done on or by the gas depends upon the type of gas. It therefore becomes useful to express the work done in terms of the constant γ for the gas. Plugging Equation 5.8.13 into Equation 5.8.18 gives:

$$W = \frac{1}{1-\gamma} (P_f V_f - P_o V_o) \tag{5.8.19}$$

Example 5.8.3

Each of the four "special processes" involves its own unique graph on the PV diagram of the quasi-static process. The graph of the isochoric process is a vertical line ($V = const$); the graph of the isobaric process is a horizontal line ($P = const$); and the graph of the isothermal process is a hyperbola ($P = \frac{const}{V}$). Use the result for the work done in an adiabatic process to show that the graph of this process on the PV diagram is ($P = const\ V^{-\gamma}$).

Solution

Plugging the function $P = const\ V^{-\gamma}$ into the work integral does the job:

$$W = \int_{V_o}^{V_f} P dV = const \int_{V_o}^{V_f} V^{-\gamma} dV = const \left[\frac{V^{-\gamma+1}}{-\gamma+1} \right]_{V_o}^{V_f} = \frac{const}{1-\gamma} \left[V_f^{-\gamma+1} - V_o^{-\gamma+1} \right] = \frac{1}{1-\gamma} \left[const\ V_f^{-\gamma} V_f - const\ V_o^{-\gamma} V_o \right]$$

$$= \frac{P_f V_f - P_o V_o}{1-\gamma}$$

The graph for the PV diagram for an adiabatic process (called an *adiabat*) looks similar to that of an isotherm – the only difference is the magnitude of the exponent of the volume. In the example above, we showed that the graph is defined by:

$$adiabatic\ process: \qquad P = const\ V^{-\gamma} \tag{5.8.20}$$

The constant γ is always greater than 1 for any gas, so it is not difficult to show mathematically that for a single system, the adiabat that passes through a given point in the PV plane is steeper than the isotherm (defined by $P = const\ V^{-1}$) that passes through the same point. This makes sense, since one would expect that the pressure would drop faster as a gas expands when there is no heat entering the system.

CHAPTER OVERVIEW

6: Applications of Thermodynamics

6.1: More Processes

Real Processes

We have assumed throughout the previous section that these processes occur quasi-statically. But in real world processes, this is virtually never the case, so one could reasonably wonder what good all of this is. That is, to what extent can we apply results found here to the real world? The answer lies in the idea of the thermodynamic state.

Consider a system in a thermodynamic equilibrium state, which then goes through a real-world (not quasi-static) process, and finally comes back to equilibrium in a different thermodynamic state. It is important to remember that the beginning and ending states are perfectly defined, even if the journey that the system took to get from one to the other is not. This means that changes in state variables from the starting state to the ending state are knowable, because their values are defined at both endpoints. Our knowledge of these special processes provides us with a very useful trick for finding these state variable changes. Let's look at a very simple example.

Example 1 – Computing Volume Change of an Ideal Gas After a Non-Quasi-Static Process

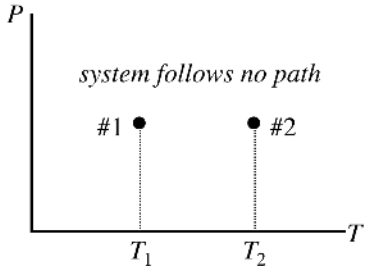
A piston confines n moles of an ideal gas at a pressure P, and is held in place with a constant force. The gas is heated quickly with a flame, and the gas expands non-quasi-statically to a new volume, where a short time later the gas comes back to equilibrium. The force on the piston is the same before and after the expansion, and the starting and ending temperatures are T_1 and T_2, respectively. Find the change in volume of the gas.

We don't need any "tricks" to solve this, because we have the ideal gas equation of state, which we know applies to the starting and ending equilibrium states. The equal force on the piston before and after means that the starting and ending pressures are equal to P, so we can easily find the change in volume:

$$\left.\begin{aligned} PV_1 &= nRT_1 \\ PV_2 &= nRT_2 \end{aligned}\right\} \quad \Rightarrow \quad V_2 - V_1 = \frac{nR}{P}(T_2 - T_1) \tag{6.1.1}$$

Suppose now that we didn't have this direct route to an answer available to us. A PT process diagram of this situation gives us an idea of what we are up against:

Figure 5.8.7 – A Non-Quasi-Static Process with Equal Starting and Ending Pressures



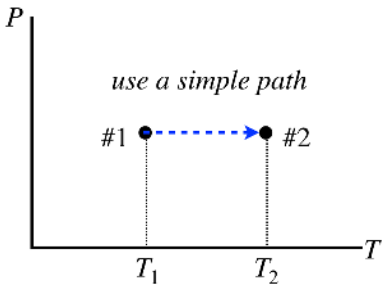
What are we supposed to do with this? Okay, so here is the trick:

Invent any quasi-static process whatsoever that connects the endpoints and use it to compute the change in the state variable.

The value of every state variable at the endpoints is well-defined, so if you can compute these values with a specific process, *it doesn't matter if that process didn't actually occur, the resulting values will be correct*! Notice that we cannot draw a conclusion about the amount of work done or the amount of heat transferred – these *are* dependent upon a path through equilibrium states, so picking a path arbitrarily doesn't help with those quantities.

What makes this trick so powerful is that we can often find a very simple quasi-static process to work with. In this particular case, the fact that the starting and ending pressures are equal inspires us to select an isobaric process:

Figure 5.8.8 – Choose an Isobaric Process for the Given Endpoints



Consider the work done over this path. We already solved this for an ideal gas – it can be expressed in two ways: Equation 5.8.7 and Equation 5.8.8. Putting these together gives us the answer we reached above:

$$P\Delta V = W = nR\Delta T \quad \Rightarrow \quad \Delta V = \frac{nR}{P}\Delta T \tag{6.1.2}$$

Example 2 – Computing Temperature Change of an Ideal Gas After a Non-Quasi-Static Process

A piston confines n moles of a diatomic ideal gas in a container at a temperature of $300K$. The piston is compressed very suddenly to one-tenth of its original volume, after which the gas quickly comes back to equilibrium. Find the new temperature of the gas.

This problem cannot be solved in the same ideal-gas-law-only manner that we used first in the previous example. That approach was possible because there were only two state variables changing, but this time the pressure, temperature, and volume are all changing. We need to use our trick, but what quasi-static process is appropriate in this case?

The key is in the fact that the process occurs very quickly. As we know from our studies in Section 5.4, heat transfer takes time, and a quick process does not allow for appreciable heat flow to occur. Therefore the starting and ending states lie along a common adiabat, and we can use the adiabatic quasi-static process for our trick. We have a functional dependence between P and V for this curve, and we have the ideal gas law, so we can relate the starting and ending volumes to the starting and ending temperatures:

$$\left.\begin{array}{l} PV = nRT \\ P = const V^{-\gamma} \end{array}\right\} \quad \Rightarrow \quad V^{\gamma-1}T = \frac{const}{nR} \quad \Rightarrow \quad V_1^{\gamma-1}T_1 = V_2^{\gamma-1}T_2 \tag{6.1.3}$$

Solving for the final temperature in terms of the initial temperature and the ratio of volumes (and noting that γ for a diatomic gas is $\frac{7}{5}$) gives:

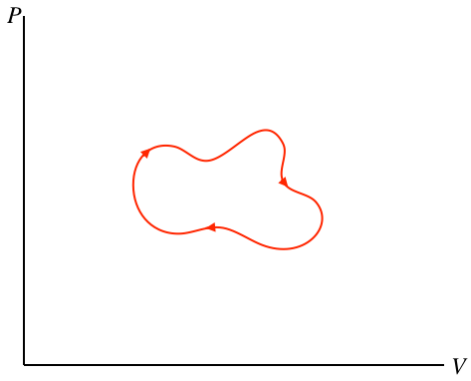
$$T_2 = \left(\frac{V_1}{V_2}\right)^{\gamma-1} T_1 = 10^{0.4}\,(300K) \approx 750K \tag{6.1.4}$$

Once again, it should be emphasized that the gas is not following a quasi-static process, so it does not go through the intermediate states of the adiabat, but as we are interested in the change of a state variable, we can choose any path that passes through the endpoints that results in no heat exchange. Interestingly, in this case, the fact that we know the change in internal energy and that there is zero heat exchanged means that we also know how much work is done. One might think that this is not supposed to be allowed, but knowing the work doesn't define the path in the PV diagram – any path between the endpoints that result in the same area-under-the-curve would work equally well.

Processes that Return to Previous States

One process that we will discuss in detail later is called a *cyclic process*. Simply put, this is a process that returns to the same thermodynamic state at which it started. In a process diagram, it forms a closed loop:

Figure 6.1.1 – A Cyclic Process

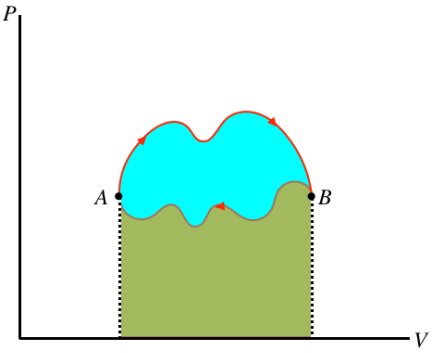


One of the state variables that returns to its original value when the cycle is complete is the internal energy. This means that for a full cycle we can use the first law to conclude:

$$0 = \Delta U = Q - W \quad \Rightarrow \quad Q = W \tag{6.1.5}$$

If there is a net amount of heat that comes into the system, it goes out of the system in the form of work – heat enters the gas, which in turn expands and does work on the piston. On the other hand, if a net amount of heat exits the system, it must have entered in the form of work. We can in fact even draw a general conclusion about which of these is occurring (and how much energy has been converted), based on the direction of the cycle (clockwise or counterclockwise) in the PV diagram. Whichever direction the cycle goes, part of the process is going left-to-right, and the rest is going right-to-left. The section of the process that is higher has a greater area under the curve, so if the top section is left-to-right there is net work done by the gas, and if it is right-to-left there is net work on the gas. Furthermore, the difference in the right-going and left-going processes is the net work done, and *this is the area bounded by the cyclic loop*.

Figure 6.1.2 – Work Done in a Cyclic Process is the Area Bounded by the Cycle

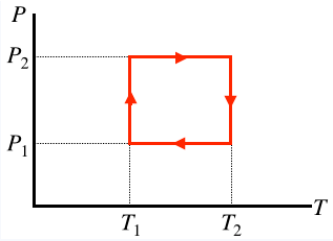


In the figure above, the area under the upper portion of the curve is positive, and the area under the lower portion is negative, so the total area for the full process is the difference, which is the blue region enclosed by the loop defining the cycle. Notice that for the case shown, the total area is positive, and if the cycle went the other way then the net work would be negative. We therefore conclude:

- *the area inside a closed cycle loop in a PV diagram represents the net amount of energy converted between work and heat*
- *clockwise cyclic processes on the PV diagram represent physical systems for which heat is entering and an equal amount of work is done by the gas*
- *counterclockwise cyclic processes on the PV diagram represent physical systems for which work is done on the gas and an equal amount of heat is expelled by the gas*

Example 6.1.1

The cyclic process of an ideal gas shown in the PT diagram below is rectangular, with vertical and horizontal sides at the positions labeled.



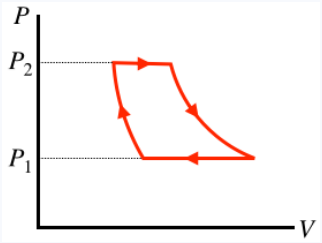
a. *Find the total heat transferred during the full cycle in terms of the quantities labeled, and indicate whether the heat is transferred into or out of the system.*
b. *Sketch the PV diagram of this cyclic process.*

Solution

a. As this is not a PV diagram, the total work done or heat exchanged is not simply the area inside the closed shape. We therefore need to look at each segment individually. The horizontal segments are processes that maintain a constant pressure, so they are isobaric. We know the work done during such a process in terms of the temperature change from Equation 5.8.9: $Q = nC_P\Delta T$. The two isobaric processes involve equal/opposite temperature changes, so these heat exchanges cancel each other out. The vertical segments are processes that maintain a constant temperature, so they are isothermal, and we can write down the work done for each segment using Equation 5.8.16:

$$\left.\begin{aligned} left\ segment: \quad & Q = nRT_1 \ln\left[\frac{P_1}{P_2}\right] \\ right\ segment: \quad & Q = nRT_2 \ln\left[\frac{P_2}{P_1}\right] \end{aligned}\right\} \quad \Rightarrow \quad Q_{tot} = nRT_2 \ln\left[\frac{P_2}{P_1}\right] + nRT_1 \ln\left[\frac{P_1}{P_2}\right] = nR\left(T_2 - T_1\right) \ln\left[\frac{P_2}{P_1}\right]$$

b. We need to sketch two isobaric and two isothermal processes that form a closed loop in a PV diagram. The isotherms need to be hyperbolas, which means the upper corners need to be at lower volumes than the corners directly below them on the PT diagram. Also, the work done in the upper isobaric process has to be equal in magnitude to the work done in the lower isobaric process (since we found that they cancel in part a). But the upper process is at a higher pressure, so the volume change has to be smaller (i.e. the horizontal segment is shorter on top than on bottom).

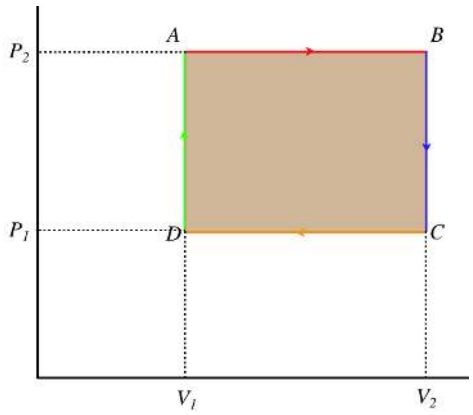


6.2: Engines and Thermal Efficiency

A Simple Engine

Cyclic processes provide a means to have repeatable ways to convert heat energy that comes into the gas into work energy that leaves the gas. In order for heat to be exchanged, we know there must be a temperature difference, and a properly-designed device can run in a cycle to exploit a temperature difference to deliver useful mechanical energy. Such a device is called a *heat engine*. Of course, this requires a cyclic process that runs clockwise on the PV diagram. We will now examine the simplest version of an engine – one that forms a rectangle in its PV diagram. Our emphasis will be to visualize each leg of the cycle as a physical process involving a piston that is exchanging heat with a thermal reservoir and/or work with its surroundings.

Figure 6.2.1 – A Simple Engine

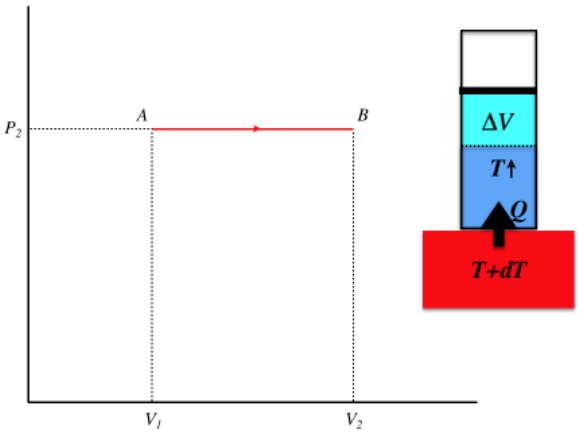


We'll start with what we already know about cycles – since the thermodynamic state returns to where it started, the internal energy doesn't change over the course of a cycle, which means that the work energy that comes out (equal to the area enclosed by the loop) equals the heat energy that goes in.

$$\Delta U = 0 \quad \Rightarrow \quad Q_{in} = W_{out} = (P_2 - P_1)(V_2 - V_1) \tag{6.2.1}$$

We will now compute the heat transferred during all four individual legs of the cyclic process to confirm this result. As we do, we will include a diagram of what is happening physically.

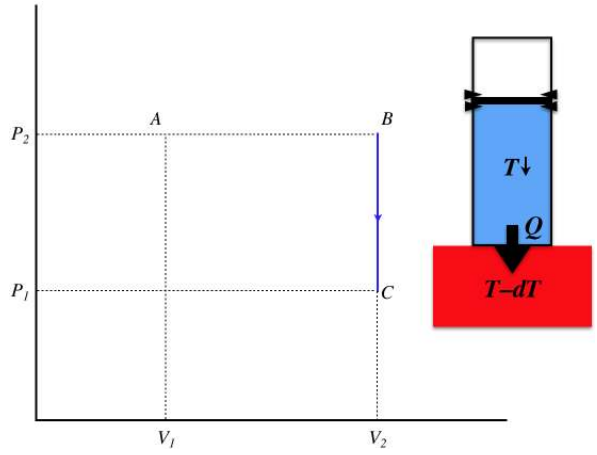
Figure 6.2.2a – Process A–B



This is a quasi-static isobaric process, which involves heat being transferred into the gas slowly (from a thermal reservoir that's barely warmer than the engine gas at every step of the process). The temperature of the gas rises during the process, and the volume increases, while heat comes into the system. The amount of heat transferred is:

$$Q_{AB} = nC_P \Delta T_{AB} = nC_P \left(\frac{P_2 \Delta V_{AB}}{nR} \right) = \left(\frac{C_P}{R} P_2 \right) (V_2 - V_1) \tag{6.2.2}$$

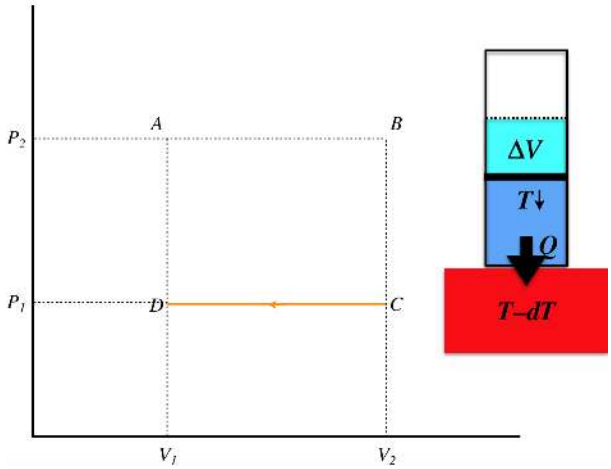
Figure 6.2.2b – Process B–C



This time we have an isochoric process, and since the pressure drops, it must be because the temperature is dropping. This can only occur with an unchanging volume when heat is leaving the system, and since the process is quasi-static, the temperature of the thermal reservoir is slightly lower than the temperature of the gas throughout the process. The heat lost during this stage is:

$$Q_{BC} = nC_V \Delta T_{BC} = nC_V \left(\frac{\Delta P_{BC} V_2}{nR} \right) = \left(\frac{C_V}{R} V_2 \right) (P_1 - P_2) \tag{6.2.3}$$

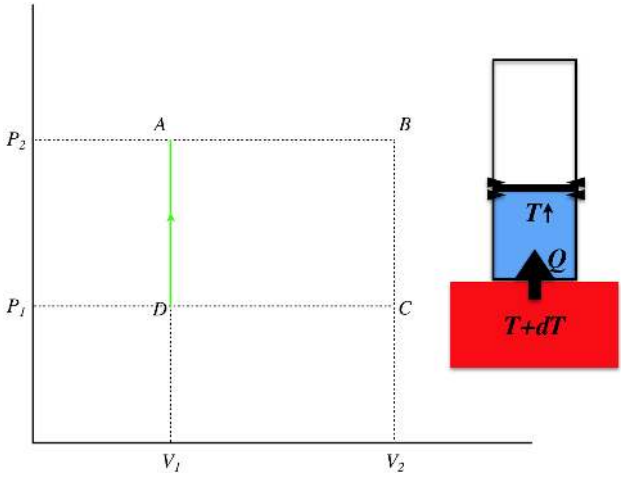
Figure 6.2.2c – Process C–D



This third leg is again an isobaric process, this time with the temperature and volume dropping. Again this quasi-static process requires that the temperature of the reservoir remain slightly lower than the temperature of the gas. The heat lost is:

$$Q_{CD} = nC_P \Delta T_{CD} = nC_P \left(\frac{P_2 \Delta V_{CD}}{nR} \right) = \left(\frac{C_P}{R} P_1 \right) (V_1 - V_2) \tag{6.2.4}$$

Figure 6.2.2d – Process D–A



The final leg is again isochoric, and the pressure is increased along with the temperature by heat added from the thermal reservoir that is slightly warmer than the gas. The heat transferred is:

$$Q_{DA} = nC_V \Delta T_{DA} = nC_V \left(\frac{\Delta P_{DA} V_1}{nR} \right) = \left(\frac{C_V}{R} V_1 \right) (P_2 - P_1) \tag{6.2.5}$$

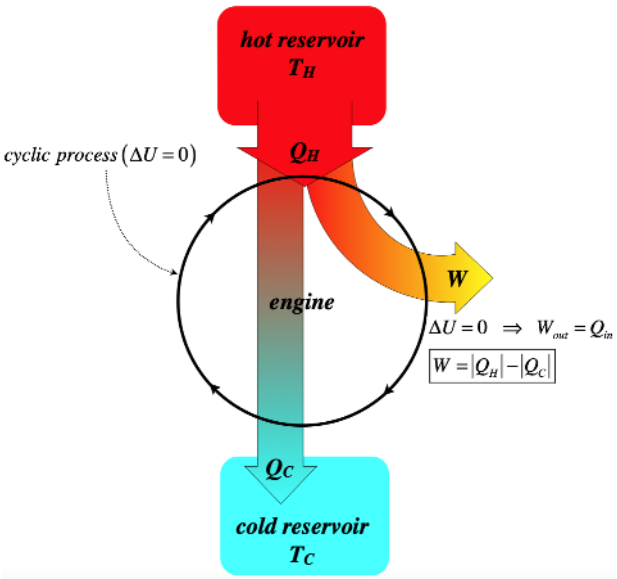
It is left as an algebra exercise to the reader to demonstrate that the sum of these four heat transfers equals the total heat transferred as given in Equation 6.2.1. When doing this exercise, it will be helpful to remember that $C_P = C_V + R$.

Real-World Engines

Throughout the calculation above, it may have occurred to the reader that there was one awkward requirement kept coming up – the thermal reservoir must always be an infinitesimal amount different in temperature from the gas in the engine. How exactly does one accomplish such a feat? The reservoir is very slightly warmer, increasing the gas's temperature until they are at thermal equilibrium, then the reservoir gets a little warmer again, so that it can again give a small amount of heat to the gas, and so on? This process is obviously not something that can be reasonably engineered, and even if it could, the fact that the rate of heat flow is related to the temperature difference means that it would be painfully slow.

In the real world, we typically have two thermal reservoirs at *fixed temperatures* to work with – one at a high temperature from which the engine receives heat, and one at a low temperature, where the engine dumps heat. Notice that in the simple engine above, the gas had to both receive and dump heat, even though it received a net amount of heat that it converted into work. This turns out to be a required feature of *all* engines (for reasons we will explore later on) – an engine cannot simply take heat in from a single hot thermal reservoir and convert it into work in a cycle without also dumping heat into another, colder thermal reservoir. A schematic of this general principle of engines is shown below.

Figure 6.2.3 – Real-World Schematic of a Heat Engine



Many elements of the engine are featured in the schematic. First, the process must be cyclic, which means that the overall change in the internal energy is zero, and the overall heat that comes in (the heat in from the warmer reservoir minus the heat out to the cooler reservoir) equals the overall work that goes out (technically there is also work that comes in, but this schematic includes only the *net* work, while dividing the "in" heat from the "out" heat, for reasons that will soon become clear). We have included the heats exchanged with the two reservoirs in terms of their absolute values, so that we don't have to concern ourselves with heat in/heat out sign conventions. Clearly the work produced is the difference of the total heat energy that comes in from the hot reservoir minus the total heat energy that goes out to the cold thermal reservoir.

Thermal Efficiency

Now it is true that in the real world when we take heat from one reservoir and give some to another, colder one, we make the two reservoirs a little closer in temperature. Ideally, we would like to avoid "wasting" any of that outgoing heat energy that does nothing but increase the temperature of the colder reservoir, and instead just convert all of the heat energy coming from the hot reservoir directly into work. Achieving this goal would mean the creation of a "perfectly efficient engine," and we would say that it has a *thermal efficiency* of 100%. Defining the percentage efficiency of any engine is therefore pretty obvious – simply take the ratio of the work extracted to the heat supplied:

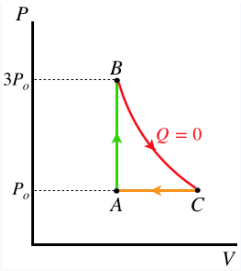
$$e = \frac{W_{net}}{Q_H} = \frac{|Q_H| - |Q_C|}{|Q_H|} = 1 - \frac{|Q_C|}{|Q_H|} \tag{6.2.6}$$

It should be noted that the "work extracted" is the net work – the work that comes out during the full cycle minus the work that is put in (i.e. it is the area inside the closed loop in the clockwise PV diagram). Notice also that for this engine the temperature is not just slightly above the temperature of the gas in the engine, and in fact the force due to the gas pressure is not slightly greater than the external force when work is being done, either. So none of these processes are quasi-static. But as we have seen, this will not stop us from making good use of the quasi-static process models.

We will see later that engines are maximally efficient when the processes they follow are reversible, but of course this requires for some processes that the thermal reservoir involved varies its temperature to remain infinitesimally larger or smaller than the temperature of the engine. This contradicts the entire notion of a "thermal reservoir," so it is clear that real engine efficiencies will be worse than those of reversible engines we may use to model them. Still, we can use the ratio of total work out to total heat in for a reversible model to compute the *maximum* possible efficiency for the engine modeled.

Example 6.2.1

In the cyclic process for an engine shown below, the process from A to B triples the pressure, the process from B to C is adiabatic, and the working gas in the engine is monatomic. Compute the maximum thermal efficiency of this engine.



Solution

Let's call the volume of the gas during the isochoric leg ($A \rightarrow B$) V_o. We can use the adiabatic process to relate the volume the gas occupies at C to V_o:

$$PV^{\gamma} = const \quad \Rightarrow \quad (3P_o)\,V_o^{\frac{5}{3}} = (P_o)\,V_C^{\frac{5}{3}} \quad \Rightarrow \quad V_C = 3^{\frac{3}{5}}V_o$$

The total area in the loop (the net work done in the cycle) is the area under the adiabat plus the (negative) area under the isobar. We have both of these quantities in terms of the pressures and volumes:

$$\left.\begin{array}{ll} \text{adiabat:} & W_1 = \dfrac{1}{1-\gamma}\left[P_fV_f - P_iV_i\right] = \frac{3}{2}\left[3P_oV_o - P_o\left(3^{\frac{3}{5}}V_o\right)\right] = 1.600P_oV_o \\ \text{isobar:} & W_1 = P\Delta V = P_o\left[V_o - \left(3^{\frac{3}{5}}V_o\right)\right] = -0.933P_oV_o \end{array}\right\} \quad \Rightarrow \quad W_{net} = W_1 + W_2 = 0.667P_oV_o$$

Heat exits the system during the isobaric process, and no heat is exchanged during the adiabatic process, so all the heat that comes into the engine does so during the isochoric process, and this is easy to compute for a monatomic ideal gas:

$$Q_{in} = \frac{3}{2}\Delta PV = 3P_oV_o$$

The efficiency comes from the ratio of net work out to heat in:

$$e = \frac{W_{net}}{Q_{in}} = \frac{0.667P_oV_o}{3P_oV_o} = 22.2\%$$

Otto Cycle

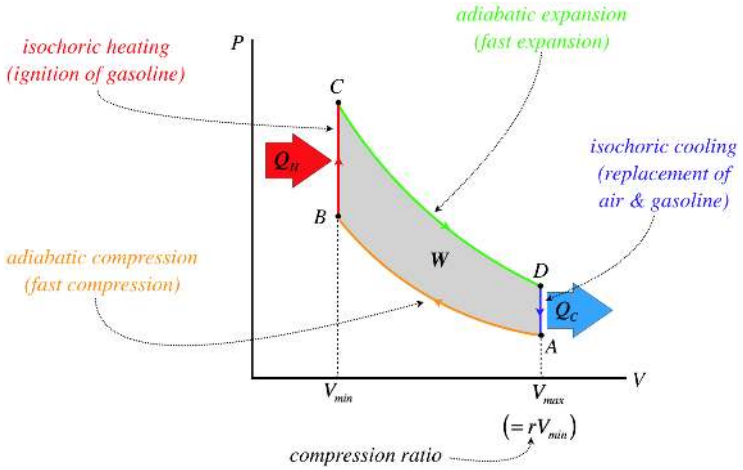
Our most recognizable type of engine is the internal combustion engine, and the most common cyclic process these follow is called the *Otto cycle*.

Alert

In what follows, when the word "gas" refers to the gas within the piston, which is mostly air. When referring to gasoline (the most common combustion fuel), we will refer to it in that long form – we will not use the shortened version of "gas."

We'll start by putting up a PV diagram that approximates the process, then explain each leg of the cycle.

Figure 6.2.4 – The Otto Cycle



process A-B (adiabatic compression)

Gasoline (or other combustible) vapor enters the chamber and mixes with air at a cool (ambient) temperature, at which point work is done on the mix to compress it. This happens very fast, so that the gas has no time to exchange heat with the surroundings, and this prompts us to treat this process as adiabatic.

process B-C (isochoric heating)

The gasoline is ignited, which rapidly changes the temperature of the gas inside the piston. Technically, the heat is not coming from outside the engine, but rather it comes from an exothermal chemical process, but it amounts to the same thing. This ignition happens very suddenly, before the gas has time to expand the piston, so we treat this process as isochoric.

process C-D (adiabatic expansion)

The heated gas is now at a very high pressure, and this pressure expands the piston, doing work. Once again, the speed of this process is so great that very little heat has time to escape the piston as it occurs, so we treat this process as adiabatic.

process D-A (isochoric cooling)

After fully expanding, the cooled-but-still-hotter-than-ambient gas is expelled from the engine, and a new load of air and gasoline vapor enters the chamber. Technically the gas doesn't "cool isochorically," but this amounts to the same thing, as the chamber is soon filled with new gas at a lower temperature and at the same volume.

This example shows how we can make use of what we have learned about thermodynamic processes to analyze real-world situations, even though our understanding is based on ideal situations that don't exist in the real world. We simply look at the features of the real-world process, and match it as closely as possible to a quasi-static process. During this "matching" process, we take care that the endpoints match correctly (because those are equilibrium states), and that the heat/work transferred during the process makes sense. In the above example, this consisted of asking if the process occurred fast (no time for heat to flow) or if the volume didn't change (no work done). We'll see another form of this matching again shortly.

Let's look at the efficiency of this cycle. Keep in mind that our idealized version will be more efficient than what we are able to achieve in the real world, but this gives us an upper-limit on what we can hope for. To get the efficiency, we need the heat supplied by the hot reservoir and the heat taken by the cold reservoir. In this cycle, heat exchanges only occur during processes B-C and D-A, which are both isochoric, so the heat exchanges are proportional to the temperature changes. The efficiency is therefore given by:

$$e = 1 - \frac{|Q_C|}{|Q_H|} = 1 - \frac{nC_V\,(T_D - T_A)}{nC_V\,(T_C - T_B)} = 1 - \frac{(T_D - T_A)}{(T_C - T_B)} \tag{6.2.7}$$

It should be clear from this result that the engine runs more efficiently when the temperature difference between the two thermal reservoirs is greater. In this case, that is the difference between the temperature of the injected gas and the ignited gas. It should be clear from the diagram that this difference can be measured in terms of the difference between (or more correctly, the ratio of) the two volumes the gas occupies. In practical terms, the gas cannot be squeezed to as small of a volume as one wants before igniting it, because the rise in temperature due to the compression can itself spontaneously ignite the gas. Higher octane fuels allow for greater compression without this unwanted spontaneous ignition, improving efficiency.

As we can infer from above, it is possible to rewrite the efficiency of this engine in terms of a variable that we can measure more easily than the temperatures – namely a property of the engine itself. Two of the four processes are isochoric, which means that the volume

only changes twice during the whole cycle, which means we have only two volumes to worry about – the maximum and the minimum. The maximum occurs when the piston is completely expanded, and the minimum when it is completely compressed. We don't even really care what these values are when it comes to efficiency, but rather all we care about is the *ratio* of these volumes, which is called the *compression ratio*:

$$r = \frac{V_{max}}{V_{min}} = \frac{V_{A\ or\ D}}{V_{B\ or\ C}} \tag{6.2.8}$$

We have a relation between the temperature and the volume along the path of an adiabat, which gives:

$$\left.\begin{array}{l} TV^{\gamma-1} = const \\ V_A = V_D\ ,\quad V_B = V_C \end{array}\right\} \quad \Rightarrow \quad \left\{\begin{array}{lll} T_A V_A^{\gamma-1} = T_B V_B^{\gamma-1} & \Rightarrow & T_B = \left(\frac{V_A}{V_B}\right)^{\gamma-1} T_A = r^{\gamma-1} T_A \\ T_C V_C^{\gamma-1} = T_D V_D^{\gamma-1} & \Rightarrow & T_D = \left(\frac{V_A}{V_B}\right)^{1-\gamma} T_C = r^{1-\gamma} T_C \end{array}\right. \tag{6.2.9}$$

Plugging the Equation 6.29 into Equation 6.27 results in the following equation for the efficiency of this cycle for a given compression ratio:

$$e = 1 - r^{1-\gamma} \tag{6.2.10}$$

Diesel Cycle

With a small alteration to the Otto cycle, the efficiency can be improved somewhat. This alteration consists of controlling the ignition process so that it occurs at a constant pressure rather than a constant volume. This engine design uses what is called the *diesel cycle*. This of course means that the ignition has to occur less "explosively," which reduces the rate at which the cycle can occur, and we know from Physics 9A that the rate at which work is output is the power of the cycle, so while this cycle comes out to be more efficient, it provides less power.

To determine the efficiency difference, one needs only to change the denominator of Equation 6.2.7, which accounts for the ignition process (from B to C). Instead of occurring at constant volume, it occurs at constant pressure, which simply changes the C_V to C_P, giving:

$$e = 1 - \frac{C_V\,(T_D - T_A)}{C_P\,(T_C - T_B)} = 1 - \frac{1}{\gamma}\,\frac{T_D - T_A}{T_C - T_B} \tag{6.2.11}$$

The term that subtracts from the efficiency is lowered by a factor of gamma, resulting in higher efficiency. In addition, higher compression ratios are possible, because the air is compressed without the fuel present (the fuel is added gradually using *fuel injectors* during the ignition process, keeping the pressure constant), removing the problem of the fuel igniting during compression. Of course, while this process was exclusive to diesel engines many years ago, nowadays fuel injection and its accompanying higher compression ratios are standard in gasoline-burning automobiles.

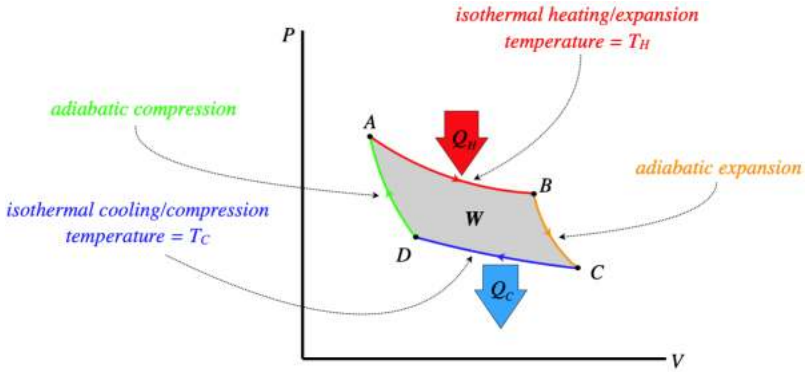
Carnot Cycle

We were able to cleverly describe the Otto and diesel cycles in terms of 4 quasi-static processes, by treating the ignition of a gas as if heat is added rather than coming from a chemical reaction, and by treating replacement of gas as if heat is expelled. Without these tricks, keeping these processes quasi-static would make them very slow, and could not happen between reservoirs of two fixed temperatures, as shown in Figure 6.2.3, because for a process involving heat transfer and temperature change (which is the case for both isochoric and isobaric processes) to occur quasi-statically, the reservoir must change temperature to remain only infinitesimally different from the temperature of the gas. We can't get something for nothing, and in fact both the ignition and gas replacement processes are irreversible, making these processes only approximately the quasi-static cycles we declared them to be.

From this analysis, we see that the problem with including isochoric and isobaric processes in the "real world" case of an engine constrained to function between two reservoirs with fixed temperatures is that we can't make these processes reversible. But even under this fixed-temperature constraint on the reservoirs, there are two processes we can (in principle) perform quasi-statically. The adiabatic process involves not heat transfer at all, so the relative temperature of the engine and the reservoir is not relevant. The isothermal process leaves the temperature of the engine fixed, so if it happens to equal the temperature of the reservoir, there is no problem.

In the discussion that followed Equation 5.8.20, we noted that at any given point on a PV diagram for a gas, the adiabat that passes through that point is steeper than the isotherm that also passes through it. Because of this, we can create a cyclic process that utilizes two isothermal processes (one on the top, one on the bottom of the PV diagram) and two adiabatic processes (one on each side of the PV diagram), and this cycle can be driven by two fixed-temperature reservoirs. This is known as the *Carnot cycle*.

Figure 6.2.5 – The Carnot Cycle



We can compute the efficiency of this engine as we did with the Otto and diesel cycles. Noting that there is no heat transferred during the two adiabatic processes, and using Equation 5.8.16 for the heat transferred during the two isothermal processes, we have:

$$\left.\begin{aligned} |Q_H| &= W_{out} = nRT_H \ln\left[\frac{V_B}{V_A}\right] \\ |Q_C| &= -W_{in} = -nRT_C \ln\left[\frac{V_D}{V_C}\right] = nRT_C \ln\left[\frac{V_C}{V_D}\right] \\ e &= 1 - \frac{|Q_C|}{|Q_H|} \end{aligned}\right\} \quad \Rightarrow \quad e = 1 - \frac{T_C}{T_H}\frac{\ln\left[\frac{V_C}{V_D}\right]}{\ln\left[\frac{V_B}{V_A}\right]} \tag{6.2.12}$$

There is more we can do here, however. The volume of state *B* is related to the volume of state *C* by virtue of being on the same adiabat (the same is true of states *D* and *A*). Thus:

$$TV^{\gamma-1} = const \quad \Rightarrow \quad \left\{\begin{aligned} T_H V_B^{\gamma-1} &= T_C V_C^{\gamma-1} \\ T_C V_D^{\gamma-1} &= T_H V_A^{\gamma-1} \end{aligned}\right\} \quad \Rightarrow \quad \frac{V_B}{V_A} = \frac{V_C}{V_D} \tag{6.2.13}$$

Plugging these in above has the logarithms in the numerator and denominator cancelling, making the efficiency of a Carnot cycle a simple function of the temperatures of the two reservoirs:

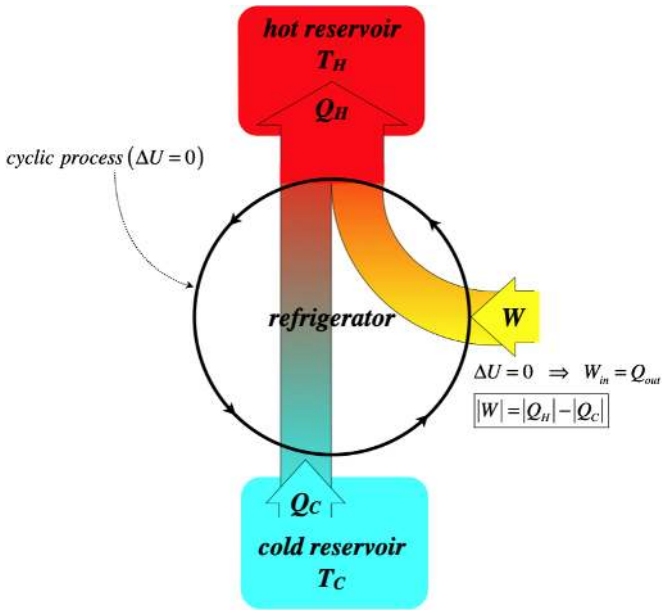
$$e = 1 - \frac{T_C}{T_H} \tag{6.2.14}$$

The greater the temperature difference is between the two reservoirs, the greater the efficiency of the Carnot engine.

Refrigerators

One thing we have seen consistently in our discussion of engines is that the cycles are clockwise on the *PV* diagram. This ensures that after a full cycle work comes *out* of the system as heat goes *in*. What happens if we run the cycle in reverse? Then work goes in and heat comes out. This is the basis of a *refrigerator*. Naturally this doesn't mean that we can take an internal combustion engine, put it in "reverse," and it turns into an air-conditioner. For one thing, we can't "un-ignite" the gas. But we can perform the processes in the opposite direction by other means. First, let's look at a schematic for a refrigerator, as we did for a heat engine:

Figure 6.2.6 – Real-World Schematic of a Refrigerator



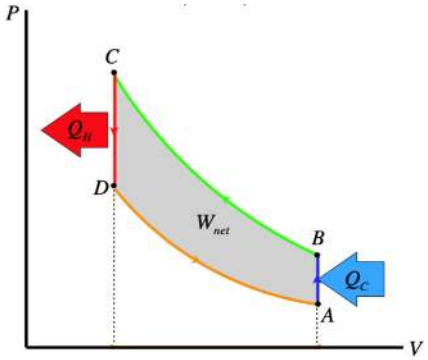
The efficiency of a refrigerator is not defined in the same way as an engine, since the goal here is to remove as much heat as possible from the cold reservoir while putting in as little work as possible. We therefore define the *coefficient of performance* as the ratio of the heat removed to the work required:

$$K = \frac{|Q_C|}{W} = \frac{|Q_C|}{|Q_H| - |Q_C|} \tag{6.2.15}$$

An extremely-simplified way to think about how a refrigerator works is this: We know that if we very suddenly compress a gas, it gets much hotter (see the example at the very end of Section 5.8). Unsurprisingly, the reverse is also true: Allowing a gas to suddenly expand a piston results in the gas cooling greatly. Suppose we want to make the interior of a refrigerator cooler than the exterior (duh, that's the definition of a refrigerator!). Start with a gas in a piston outside the refrigerator, and compress it to a small volume, and wait with it compressed until it comes to the outside temperature. Then release the piston suddenly and quickly carry it into the refrigerator. If we compressed it enough, the temperature change of the gas in the piston will bring its temperature below that of the interior of the refrigerator. We wait a little while, as the interior of the refrigerator surrenders heat to the cold air in the piston, thereby cooling the air within the refrigerator. When these reach equilibrium, we carry the piston back outside, and repeat the process. This transports thermal energy out of the refrigerator.

The work done on the gas during compression exceeds the work done by the gas during expansion (i.e. net work needs to be put in). The compression and expansion processes are adiabatic, while the "waiting" processes are isochoric, which gives a PV diagram that looks something like this:

Figure 6.2.7 – PV Diagram of a Simple Refrigerator



Obviously we sacrificed a lot of reality for this easy-to-understand "refrigerator." We obviously don't need to transport the piston in and out of the refrigerated chamber, and can instead pipe the gas into and out of it, compressing it as it leaves, and expanding it as it enters. But there is still a fairly big problem with this design. In order for heat to be transferred in the proper directions at the proper times, we

need the temperature of the gas after it cools from expansion to be lower than the ambient temperature in the refrigerator. On the *PV* diagram, the temperatures inside and outside the refrigerators correspond more-or-less to the temperatures of states *B* and *D*, respectively. This means that if we draw isotherms through points *B* and *D*, that the gap between those isotherms represents the maximum temperature gap we can maintain between the hot and cold regions. Obviously this is a function of the pressure difference we can create between the compressed gas and the expanded gas, but in practical terms, this is a substantial obstacle.

The way this limitation is overcome is to carry much of the thermal energy in the *phase* of the refrigerant. We know that we can change phases by combinations of compressing/expanding and heating/cooling the fluid, and the latent heat of vaporization is substantial compared to the specific heat capacity for a small temperature change. This leads to this basic process:

- the *compressor* changes the phase of the refrigerant into liquid, which warms it above the outside temperature
- the fluid then enters a *condenser coil*, which has the purpose of increasing the contact area with the outside air, speeding up the process of dumping heat
- by the time the fluid has passed through the condenser coil, it is at a high pressure, but has come to thermal equilibrium with the outside air, and it then passes into an *expansion valve*, where it expands adiabatically, changing phase back to gas and dropping significantly in temperature, below the temperature of the inside air
- the gas then passes through an *evaporator coil*, which increases the rate as which the heat can enter the refrigerant from the inside air, and at the end of the evaporator coil, it reenters the compressor to start the cycle again.

6.3: Entropy

Extensive and Intensive State Variables

All state variables can be classified into one of two categories, which we call *extensive* and *intensive*. Extensive state variables are those that are additive when two systems are combined into one, while intensive state variables are not additive. For example, suppose we have two identical boxes of the same gas in the same thermodynamic state that are separated by a barrier. Removing the barrier and treating the new bigger box as a single system, we find that we have doubled the volume, particle number, and internal energy of the system. These are therefore extensive state variables. But if we measure the pressure or temperature of the new state, we find that these quantities are unchanged, making them intensive variables. All state variables come in one variety or the other.

The way to tell mathematically whether a state variable is extensive or intensive is to look at what happens to it as a function of other state variables when they are scaled according to whether they are extensive or not. So for example, suppose we know that the number of moles n is extensive and the total energy U in an ideal gas system are extensive, but we are not sure about temperature T. Writing temperature in terms of n and U gives:

$$T(n,U) = \frac{U}{nC_V} \tag{6.3.1}$$

Scaling the system by a factor of k (i.e. putting together k identical systems) changes n and U by that factor but has no effect on T:

$$T(k\cdot n, k\cdot U) = \frac{k\cdot U}{k\cdot nC_V} = \frac{U}{nC_V} = T(n,U) \tag{6.3.2}$$

Example 6.3.1

Given that n and V are extensive, and T is intensive, use the ideal gas law to show that P is intensive.

Solution

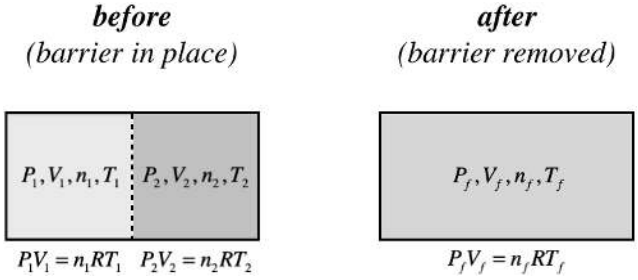
Following the same process as above where we scale the extensive variables ($n \to k\cdot n$, $V \to k\cdot V$), and do nothing to the intensive variables ($T \to T$) we have:

$$P(k\cdot n, k\cdot V, T) = \frac{k\cdot nRT}{k\cdot V} = \frac{nRT}{V} = P(n,V,T)$$

Since P doesn't change upon scaling the system, it is an intensive state variable.

Whether a variable is intensive or extensive also affects the final result when we combine two systems that are not identical. If we again consider two systems separated by a barrier which this time have different state variables (but are the same type of gas), then we get the following differences in the results for each type of variable:

Figure 6.3.1 – State Variables in Combined Systems



The extensive variables will simply add. This is obvious for particle number and volume, but the sum of the internal energies of the two systems must equal the internal energy of the combined system as well (energy isn't created or destroyed by the combination):

$$V_f = V_1 + V_2\ , \quad n_f = n_1 + n_2\ , \quad U_f = U_1 + U_2 \tag{6.3.3}$$

We can use these to determine what happens to the intensive variables. First use the internal energy relation to determine the fate of the temperature:

$$n_f C_V T_f = n_1 C_V T_1 + n_2 C_V T_2 \quad \Rightarrow \quad T_f = \frac{n_1 T_1 + n_2 T_2}{n_1 + n_2} \tag{6.3.4}$$

And combining this with the ideal gas law gives a similar result for pressure:

$$P_f V_f = n_f R T_f \quad \Rightarrow \quad P_f \left(V_1 + V_2\right) = \left(n_1 + n_2\right) R \left[\frac{n_1 T_1 + n_2 T_2}{n_1 + n_2}\right] \quad \Rightarrow \quad P_f = \frac{V_1 P_1 + V_2 P_2}{V_1 + V_2} \tag{6.3.5}$$

So you see that the temperature and pressure of the combined system are "weighted averages" of the temperatures and pressures of the individual systems.

A New State Variable

We know that work and heat are similar in that they are the two quantities in thermodynamics that are not state variables. We know specifically how work arises from a single infinitesimal step in a quasi-static process: $dW = PdV$, but as yet we have no equivalent relation for heat. If we assume, because of its similarity with work, that heat does have such a relation with two state functions, then it should look something like: $dQ = XdY$. Comparing this with the case of work, and keeping in mind what properties “drive” work and heat, which of the following are likely candidates for the state functions X and Y?

Well, we know that a pressure difference on the two sides of a piston will result in work being done, and while work is "driven" by a pressure difference, heat is driven by a temperature difference. We therefore infer that the state variable X is temperature, but Y is a mystery. One thing we can perhaps conclude that since work is defined from a combination of an intensive function (P) and the change of an extensive function (dV), then dY should be extensive (note that the driving variable temperature is intensive, as pressure is in the case of work).

None of the state variables we have seen so far fit the bill, so we'll just postulate the existence of another state variable we call *entropy*, and give it the symbol S. We therefore have, in analogy with work the relation for a small quantity of heat:

$$dQ = TdS \tag{6.3.6}$$

The units for entropy are the same as those for heat capacity ($\frac{J}{K}$), but heat capacity and entropy are not the same. As with the case of work, we can add up all the small contributions to heat transferred during a quasi-static process:

$$Q = \int TdS \tag{6.3.7}$$

This is good place to point out that these relationships for work and heat can also be turned around to result in a change of a state function for a quasi-static process:

$$\begin{aligned} dV = \frac{dW}{P} \quad &\Rightarrow \quad \Delta V = V_B - V_A \quad = \quad \int_A^B \frac{dW}{P} \\ dS = \frac{dQ}{T} \quad &\Rightarrow \quad \Delta S = S_B - S_A \quad = \quad \int_A^B \frac{dQ}{T} \end{aligned} \tag{6.3.8}$$

Of course, the first of these relations is virtually never used, because we can generally just look at what the piston does to determine the volume change – we don’t have to calculate it. The second relation is another matter – the change in entropy is not immediately apparent, and as we will see, knowing the change of entropy can be quite important to understanding a process.

Question: A gas is sealed in an insulated cylinder with a piston. The piston is then compressed slowly (a quasi-static process), and the temperature of the gas rises. In what direction does the entropy function change for that gas?

Answer: The piston is *insulated*, which means that the gas does not exchange heat with the outside environment. This is a quasi-static process, so $dQ = 0$ throughout the process means that $dS = 0$ throughout the process, which means that $\Delta S = 0$.

Here we can see why the ΔS equation is so much more commonly-used than the ΔV equation – we know that an adiabatic process occurs when we insulate the container, and that this prevents heat from being transferred, so the entropy change is zero. The equivalent process for the case of work is the isochoric process, which we know immediately involves zero change in volume – we

don’t need to reason that we have rigged the apparatus so that no work is exchanged and therefore conclude that there must not have been a volume change.

In the case of a quasi-static process where zero work is done, we were able to characterize it in terms of a changing state variable – an isochoric process. Until now, we had no such characterization for a quasi-static process where no heat is exchanged. But so long as we are talking about a quasi-static process, we can rename the adiabatic process in terms of the unchanging state variable: an *isentropic process*.

Ideal Gases

When we first started discussing state functions, we said that there are four independent state variables needed to define a state, but when we know more about how the particles interact (such as an ideal gas, where they don't interact), this number drops to three. We said that this allows us to express one state variable in terms of three others, as we did in Equation 5.5.2. Well, now we have a new state variable, and it is useful to express it as a function of three others. Derivation of this expression is a bit involved, but here it is:

$$S(N,U,V) = Nk_B \ln\left[\left(\frac{V}{N}\right)\left(\frac{U}{N}\right)^{\frac{1}{\gamma-1}}\right] \tag{6.3.9}$$

One thing to note here is that while the ideal gas law state equation is the same for *all* ideal gases, the state equation for entropy depends upon the *type* (monatomic, diatomic, etc) of ideal gas, as evidenced by the presence of the constant γ.

This beastly-looking expression actually carries with it quite a lot of power, though it can take a bit of mathematics to extract it. We can get more by rewriting the entropy in terms of three other variables. For example, we can replace internal energy with pressure by noting that:

$$U = nC_V T = \frac{C_V}{R} nRT = \frac{C_V}{R} PV = \frac{1}{\gamma-1} PV \tag{6.3.10}$$

Using the properties of the logarithm and doing the algebra to simplify the entropy function gives:

$$S(N,P,V) = \frac{Nk_B}{\gamma-1} \ln[PV^\gamma] + f(N) \ , \tag{6.3.11}$$

where $f(N)$ is a function of N that does not need to be written out here. Holding N constant as we always do in our limited treatment of thermodynamics, we can write this equation as:

$$S(N,P,V) = (constant) \ln[PV^\gamma] + (constant) \tag{6.3.12}$$

If we now ask, “On what curve on a PV graph will the entropy of the state remain constant ($\Delta S = 0$)?”, the answer is obvious:

$$\Delta S = 0 \text{ when } PV^\gamma = constant \tag{6.3.13}$$

This is the equation for an adiabat, which is a curve for which there is no heat transfer when the process is quasi-static. This makes sense, given the relationship between entropy and heat exchange.

Entropy Changes in Special Processes

Back when we studied the four “special” processes, we derived all the changes in state functions for each process. We can do the same for entropy using the function above, but it is simpler to do with the integral. We have already stated (and shown) that for a quasi-static adiabatic process the entropy change is zero. Let’s see how we get the change for the other processes. In every case, we are using Equation 6.3.7, and plugging in what we know about heat for each case from Section 5.8 (note that the final equality in each result uses the ideal gas law and the constant variable in that case):

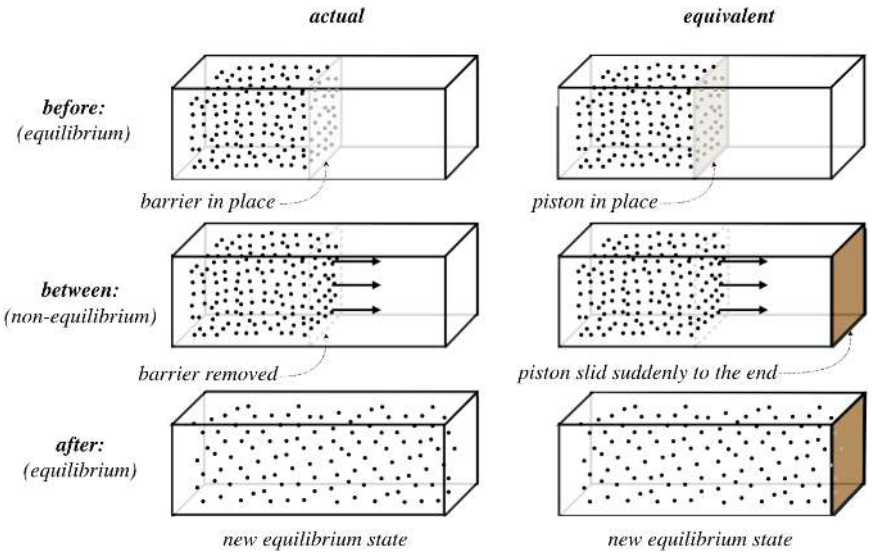
$$\Delta S = \int_A^B \frac{dQ}{T} \quad \Rightarrow \quad \begin{cases} \text{isochoric process:} & \Delta S = \int_A^B \frac{nC_V dT}{T} = nC_V \ln\left[\frac{T_B}{T_A}\right] = nC_V \ln\left[\frac{P_B}{P_A}\right] \\ \text{isobaric process:} & \Delta S = \int_A^B \frac{nC_P dT}{T} = nC_P \ln\left[\frac{T_B}{T_A}\right] = nC_P \ln\left[\frac{V_B}{V_A}\right] \\ \text{isothermal process:} & \Delta S = \frac{1}{T}\int_A^B dQ = \frac{Q}{T} = nR \ln\left[\frac{V_B}{V_A}\right] = nR \ln\left[\frac{P_A}{P_B}\right] \end{cases} \tag{6.3.14}$$

Free Expansion

We ended Section 5.8 with an important discussion about how we can deal with non-quasi-static processes even though all of our models and results regarding processes are centered around solving quasi-static processes. The subtle and important point is that if the *endpoints* of such a process are equilibrium states, then the values of the state variables at those points are well-defined, and we can sometimes use what we know about the process to relate those points to each other, and then use what we know about special processes to solve for something useful. What follows is undoubtedly the best example of this method.

Consider an insulated container of a monatomic ideal gas that has all of the gas confined at equilibrium in *one half* of the container, thanks to a membrane (barrier) between the two halves, with the other half completely evacuated. Suddenly, the membrane ruptures, and gas quickly fills the container and eventually returns to equilibrium. We call this process *free expansion*. It helps to put this process into our usual context of a gas confined with a piston, so here is an appropriate picture of what is going on:

Figure 6.3.2 – Free Expansion as a Sudden Shove of a Piston



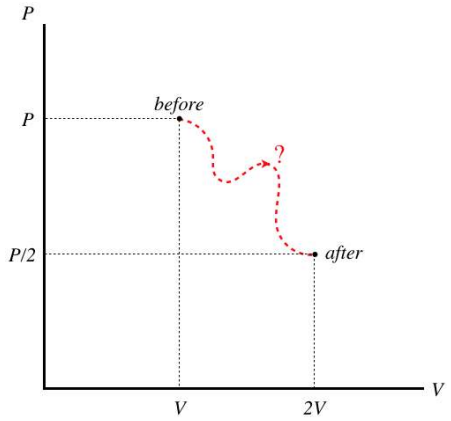
Clearly this is not a quasi-static process. But that doesn't mean we can't say anything about how the beginning and ending states are related. We are defining our system as the whole container, and it is insulated, so we know for sure that no heat enters the system while it evolves. Also, the piston is shoved so quickly that the gas particles do not rebound against it as it moves. Consequently, the gas exerts no force on the piston, and therefore doesn't do any work. [*That is not to say that no work is done at all, of course – **we** do work on the piston from outside the system. But the point is that the work done on the piston is not reflected in the state of the gas.*]

With no work done on or by the gas, and no heat added to or taken away from the gas, the first law ensures that the internal energy of the gas is unchanged. This actually makes sense, since the internal energy is the sum of the kinetic energies of all the particles, and why would suddenly removing the barrier cause any of the particles to change the kinetic energies they had just before the barrier moved? But when one considers that an unchanging internal energy means that the temperature is also unchanged, then it starts to seem a bit weird – we are used to quick expansions cooling the gas. But here we have to be careful. A gas cools when it expands *adiabatically*, which assumes that work is done on the piston. We only treated "quick" expansion as adiabatic because it was a practical means of having an expansion without significant heat loss. But *free* expansion is not the same as adiabatic expansion – there is no work done on a piston at all, and therefore no way for the internal energy to exit the system.

So with no work done, no heat transferred, and no change in temperature, what *has* changed? Well, clearly the volume has doubled. With the temperature and number of particles unchanged, the ideal gas law tells us that the pressure is cut in half. That accounts for all of the state variables except for one... What about the entropy? It is tempting to say that because there is no heat exchanged that $Q = \int TdS$ tells us that the entropy change is also equal to zero, but this is not correct! The reason is that this relation, like $W = \int PdV$, only links two equilibrium states when a quasi-static process is followed, and this is not such a process.

We therefore turn to our trick – we use what we know about the actual process and the equilibrium endpoints to *invent any quasi-static process whatsoever* between the two endpoints to calculate the entropy change. To see how we might do this, let's start by plotting the two points on a PV diagram:

Figure 6.3.3 – PV diagram of Free Expansion



There is no "correct" path here – the system does not pass through any equilibrium states during its journey. But as we are looking for a change in a state function, only the endpoints matter, and if a particular path is useful, we can go ahead and use it. So let's just run through our options one-by-one. Both the pressure and volume are changing, so neither an isochoric nor an isobaric process will connect these dots. Let's try an adiabat. in this case, the pressure and volume are related in a specific way, so we can check to see if the endpoints lie along the same adiabat:

$$P_1 V_1^\gamma = P_2 V_2^\gamma \quad \Rightarrow \quad PV^\gamma = \left(\frac{P}{2}\right)(2V)^\gamma \quad \Rightarrow \quad PV^\gamma \neq 2^{\gamma-1} PV^\gamma \tag{6.3.15}$$

The two points do not lie along an adiabat. Wait a minute – we already said that the temperature doesn't change, so we know these points lie along an isotherm. We already have this solution in terms of the volume (or pressure) change from Equation 6.3.14, so we get the answer immediately:

$$\Delta S = nR\ln\left[\frac{V_B}{V_A}\right] = nR\ln\left[\frac{2V}{V}\right] = nR\ln 2 \tag{6.3.16}$$

Example 6.3.2

For the free-expansion case above, show that you can get the same entropy change, even if you choose the less-convenient path between the endpoints that is first isochoric to the proper pressure, and then isobaric to the proper volume.

Solution

The first process is isochoric from P to $\frac{P}{2}$, and we have the entropy change for this process in Equation 6.3.14:

$$\Delta S_1 = nC_V \ln\left[\frac{P_B}{P_A}\right] = nC_V \ln\left[\frac{\frac{P}{2}}{P}\right] = nC_V \ln\left[\frac{1}{2}\right] = -nC_V \ln 2$$

The second process is isobaric from V to $2V$, and once again we have already computed this result in Equation 6.3.14:

$$\Delta S_2 = nC_P \ln\left[\frac{V_A}{V_B}\right] = nC_P \ln\left[\frac{2V}{V}\right] = nC_P \ln 2 = nC_P \ln 2$$

The total change in entropy is the sum of these changes:

$$\Delta S = -nC_V \ln 2 + nC_P \ln 2 = n\left(C_P - C_V\right)\ln 2 = nR\ln 2$$

Example 6.3.3

For the free-expansion case above, show that you can get the same entropy change using Equation 6.3.11.

Solution

Noting that the particle number doesn't change and computing the entropy change directly using properties of the logarithm:

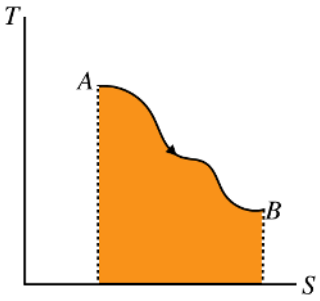
$$\Delta S = S_2 - S_1 = \frac{Nk_B}{\gamma - 1}\left\{\ln[P_2 V_2^{\gamma}] - \ln[P_1 V_1^{\gamma}]\right\} = \frac{nR}{\gamma - 1}\left\{\ln\left[\left(\frac{1}{2}P\right)(2V)^{\gamma}\right] - \ln[PV^{\gamma}]\right\} = \frac{nR}{\gamma - 1}\ln\left[2^{\gamma - 1}\right] = nR\ln 2$$

6.4: The Second Law of Thermodynamics

TS Process Diagrams

We have spent a lot of time with PV diagrams because they so clearly relate to the work done in a process through the area under the curve. We have also looked at a number of process diagrams that involve two other state variables. Now that we have added entropy to our arsenal, it's time to get it in on the action. Returning to our original motivation for introducing the entropy – as a way of creating an integral for heat analogous to that for work, it comes to mind that a TS process diagram can do for heat what PV diagrams do for work.

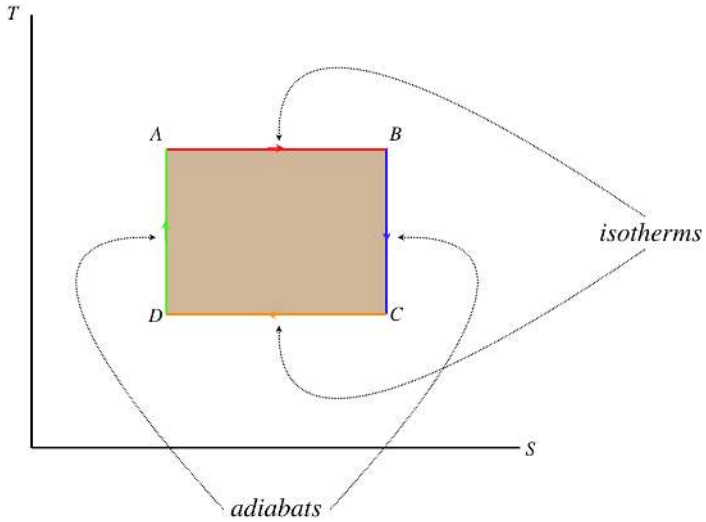
Figure 6.4.1 – Heat Transferred as the Area Under the T vs. S Curve



$$Q\,(A \to B) = \int_A^B dQ = \int_A^B T dS \tag{6.4.1}$$

Interestingly, for a cyclic process, since $\Delta U = 0$ (so $Q = W$), the area inside the closed loop (which for a TS diagram is the total heat transferred), *still* equals the total work done over the cycle. Another interesting feature is the Carnot cycle that was such an ugly mess on the *PV* diagram is a nice, easy rectangle on the *TS* diagram.

Figure 6.4.2 – Carnot Cycle on a TS Diagram



Of course, the nice rectangular (isochoric & isobaric) cycle on the *PV* diagram becomes quite a mess on the *TS* diagram, as the Carnot cycle is on the PV diagram.

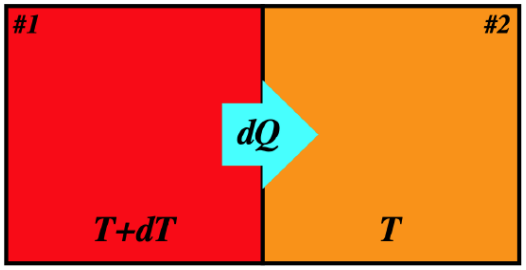
Multiple Systems

The topic of entropy becomes particularly important when you consider the effect that an exchange of heat or work has on *all* of the systems involved in the exchange. When we think about these kinds of situations, we need to keep two things in mind:

- The energy lost by one system is gained by the other(s). That is, we are considering only *isolated* combinations of systems – there is no interplay between the systems in question and the "outside."
- The entropy function is extensive, which means it is additive when two systems are considered as one.

Consider first two systems that exchange heat reversibly, meaning that their temperatures are infinitesimally close. The entropy change of *the two systems combined* remains unchanged:

Figure 6.4.3 – Entropy Change for Reversible Heat Exchange



Let's write the heat lost by system #1 and the heat gained by system #2 in terms of the temperatures and entropy changes:

$$\begin{aligned} dQ_1 &= T_1 dS_1 = (T+dT)\, dS_1 \\ dQ_2 &= T_2 dS_2 = T dS_2 \end{aligned} \tag{6.4.2}$$

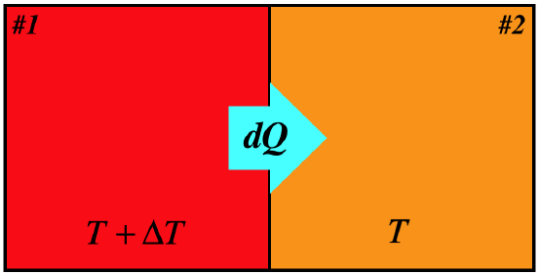
Now we note that the heat gained by system #2 equals the heat lost by system #1 (remember, by assumption these systems do not interact with their surroundings). Also, when comparing a single differential to a product of differentials, the latter vanishes, which means that one of the terms above goes away, leaving:

$$dQ_1 = -dQ_2 \quad \Rightarrow \quad T dS_1 + dT dS_1 = -T dS_2 \quad \Rightarrow \quad dS_1 + dS_2 = d\,(S_1 + S_2) = dS_{tot} = 0 \tag{6.4.3}$$

So we see that for a reversible transfer of heat, the total entropy of the two systems involved remains unchanged. Recall that we said that the key feature of reversible processes is that the system doesn't carry any "momentum" from one state to the next – every state in the process is in equilibrium, and the process can stop instantly in an equilibrium state. We can now see that the state variable of entropy gives us a way of characterizing thermal equilibrium. Whenever a function satisfies $df = 0$, is means that the function has hit an extremum (a maximum or a minimum). We formally state it this way: *A closed system is in a state of thermal equilibrium whenever the entropy function of the system hits an extremum, that is, when* $dS = 0$.

The question now becomes, "Is the extremum defined by $dS = 0$ a maximum or a minimum? To answer this, let's look at a case of two systems undergoing an irreversible heat transfer because their temperatures differ by more than an infinitesimal amount.

Figure 6.4.4 – Entropy Change for Irreversible Heat Exchange



This time with the temperature difference ΔT finite, we don't get the "product of infinitesimals vanishes" situation that we got with the reversible case:

$$\left.\begin{aligned} dQ_1 = T_1 dS_1 \quad &\Rightarrow \quad dS_1 = \frac{dQ_1}{T+\Delta T} \\ dQ_2 = T_2 dS_2 \quad &\Rightarrow \quad dS_2 = \frac{dQ_2}{T} \\ |dQ| = -dQ_1 = dQ_2 \quad & \end{aligned}\right\} \quad \Rightarrow \quad dS_{tot} = dS_1 + dS_2 = \frac{-|dQ|}{T+\Delta T} + \frac{|dQ|}{T} > 0 \tag{6.4.4}$$

We see that the change of entropy for the closed system is positive here, which means that for equilibrium the zero change in entropy within the closed system corresponds to a maximum. We therefore make our former statement more specific, and elevate it with its very own pink lettering:

A closed system is in a state of thermal equilibrium if and only if the entropy function of the system is a maximum: $dS = 0$.

The Second Law

The relationship between the entropy state function and thermal equilibrium now established, we can extend it to processes that take place within a closed system. The subsystems within a closed system can exchange work and/or heat to produce processes, and if these are reversible, the entropy of the full system is unchanged, though the entropy of each subsystem can change (one goes up while the other goes down the same amount).

But if the subsystems differ such that the processes they produce are irreversible (due to finite differences in temperature or pressure), then the entropy of the full system will not be a maximum. If a process occurs due to the subsystem imbalance that brings the whole system closer to equilibrium, then the entropy of the whole system must get closer to its maximum – *it must go up*. Note that the entropy for a single subsystem can go down, but in that case the entropy of the other subsystem must go up more.

Accounting for both the reversible and irreversible cases, we end up with the *second law of thermodynamics*:

For any process within a closed system, the values of the entropy function at the endpoints of the process satisfy $\Delta S \geq 0$.

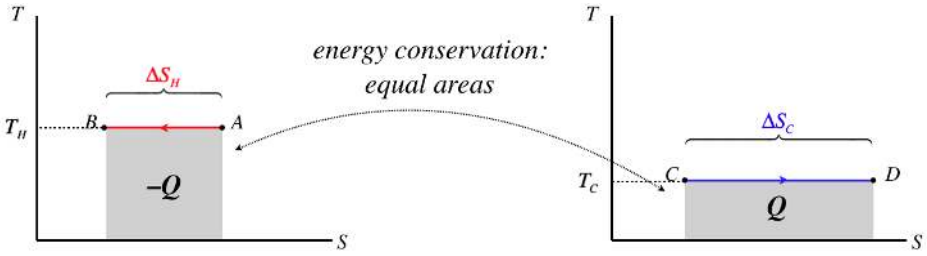
The equality occurs when the process is reversible, and irreversible processes result in an increase of this state variable.

Alert

The most common error made by students when considering the second law is that they focus entirely upon the fact that entropy increases, and forget about the "closed system" requirement. This leads to confusion when a calculation shows that a system's entropy goes down – oftentimes the calculation is correct, but the system under discussion is not actually isolated. This fact can sometimes be quite subtle and hard to see.

A nice way to illustrate the second law is with TS diagrams. Let's consider two thermal reservoirs which directly exchange heat with each other, and are at temperatures that differ by more than an infinitesimal amount and exchange heat. We'll draw TS diagrams for both of these systems side-by side, with the hotter subsystem on the left and the temperature axes on the same scales:

Figure 6.4.5 – TS Diagrams for Thermal Reservoirs Exchanging Heat



As these are thermal reservoirs, their temperatures don't change during this heat exchange. We know that the horizontal line on diagram for the hotter reservoir is higher than that of the cooler reservoir. We also know that the magnitude of the areas under the curves have to be equal, since all the heat that leaves the hotter reservoir enters the cooler one. Therefore the *length* of the graphs cannot be equal – the length of the segment on the diagram for the cooler reservoir must be longer.

Our sign convention for heat requires that the heat that leaves the hot reservoir is negative (and the integral goes right-to-left), while the heat entering the cooler reservoir is positive. Therefore we have:

$$\left.\begin{array}{l} |\Delta S_C| > |\Delta S_H| \\ \Delta S_C > 0\,, \quad \Delta S_H < 0 \end{array}\right\} \quad \Rightarrow \quad \Delta S_{tot} = \Delta S_C + \Delta S_H > 0 \tag{6.4.5}$$

The combined entropy of the two reservoirs goes up, because the entropy gain of the colder reservoir is greater than the entropy loss of the hotter reservoir.

Forbidden Cyclic Processes

It was stated without proof in Section 6.2 that engines cannot run in a cycle with only a single thermal reservoir. That is, they must always give up heat to a colder reservoir, forcing them to have less than 100% efficiency. We can now prove this is true with the second law. The engine runs in a cycle, which means that its thermodynamic state returns to the state at which it starts. Since entropy is a state variable, it comes back to where it started, which means that the engine's entropy does not change during the course of a full cycle. The reservoir that provides the heat converted into work by the engine is *losing* heat, which means that its entropy *goes down*. If this was the only reservoir, then the closed system of the engine and reservoir would experience a decrease in total entropy, violating the second law. With the addition of a cold reservoir, the heat that goes into it causes its entropy to rise, and the closed system of the engine and both reservoirs avoids having its total entropy go down.

Another forbidden device is a refrigerator that draws heat from a colder reservoir and puts it into a warmer one without work being put in. A refrigerator works in a cycle, so its internal energy (which is a state function) doesn't change from beginning to end. This means that all of the heat that leaves the colder reservoir enters the warmer one. Therefore with no work put in, the amount of heat that leaves equals the heat that enters. But with a lower temperature and $dS = \frac{dQ}{T}$, this means that the colder reservoir loses more entropy than the warmer one gains. The refrigerator's cycle leaves its entropy unchanged, so this means that the entropy of the closed system consisting of the two reservoirs and the refrigerator has gone down, in contradiction with the second law. But why does the work make a difference, if a work process has no effect on

the entropy? Because the energy that comes in as work gets *added to the heat that goes into the hot reservoir*. This increases the entropy that enters the hot reservoir, compensating for the entropy lost by the cooler reservoir.

Entropy and Efficiency

It is also enlightening to have another look at engine efficiency from the perspective of entropy. The definition of the efficiency of an engine between two thermal reservoirs is given by Equation 6.26. We know that the state of the engine returns to where it started, and entropy is a state function, so for a full cycle it has no change in entropy. Also we assume that the reservoirs do not change temperature, so we can immediately write the heat they gain or lose in terms of their entropy changes:

$$|Q_C| = T_C\,|\Delta S_C| \ , \qquad |Q_H| = T_H\,|\Delta S_H| \tag{6.4.6}$$

Plugging these into the engine efficiency formula gives:

$$e = 1 - \frac{T_C\,|\Delta S_C|}{T_H\,|\Delta S_H|} \tag{6.4.7}$$

If all of the processes for the engine are reversible, then the entropy change of the full system (engine and both reservoirs) is zero, which means that the entropy change of the reservoirs are equal in magnitude, which gives:

$$e = 1 - \frac{T_C}{T_H} \tag{6.4.8}$$

This was the same result we got for the Carnot cycle, which makes sense, because that cycle allows for constant-temperature reservoirs and reversible processes to coexist (other engines require heat be supplied-to and/or accepted-by a region that keeps changing temperature to stay infinitesimally close to the temperature of the engine).

Now consider what happens to the efficiency if there are irreversible heat exchanges with one or both of the thermal reservoirs (i.e. they have temperatures differing from that of the engine by a finite amount). The entropy of the engine itself still doesn't change for a cycle, but the entropy of the closed system must get *larger*. The entropy change of the cold reservoir is positive (heat is going in), and for the hot reservoir it is negative, so to end up with a net positive, the former must exceed the latter. When this happens, the magnitude of the negative second term in the efficiency grows, causing the efficiency to drop. The upshot is that the *efficiency given for the Carnot cycle is the maximum possible attainable efficiency for the two reservoirs provided*, and as soon as irreversible heat exchanges are introduced, that efficiency declines.

Why Do Systems Head Toward Equilibrium at All?

There is a danger at this point of falling into the following circular reasoning:

- Closed systems head toward equilibrium because their entropy must increase toward its maximum.
- Entropy increases because heat flows from hotter regions to cooler ones.
- Heat flows from hotter regions to cooler ones because they always head toward equilibrium.

The question we need to answer to get out of this ugly infinite loop is why systems head toward equilibrium at all. Why can't the hotter subsystem get even hotter while the colder subsystem gets even colder? Energy is still conserved in this case, so it doesn't seem to violate any fundamental laws of physics, other than the second law of thermodynamics, which we only derived by assuming that systems do head toward equilibrium.

It's actually a very vexing question, and we will not be able to answer it rigorously in this class, but the general idea is not hard to grasp, and is quite elegant. The crux of the matter lies in the idea of bridging the microscopic goings-on of many particles to the macroscopic quantities we can measure. We first saw this in the kinetic theory of gases (Section 5.5), when we related pressure and temperature to the random motions of particles in a gas.

With a mole of gas containing more than 10^{23} particles, it's safe to say that we can't account for all of them at once. We therefore need to account for them with some sort of "overview" that can't distinguish between exact situations. An accounting of every gas particle's position and momentum tells us everything there is to know about the physical state of the system, and we call this the *microstate* of that system. Clearly there are many different microstates that will result in the same average kinetic energy of a particle (which is proportional to the temperature), and these average quantities define what we have called the thermodynamic state, which is also called the *macrostate* of the system. The following is therefore clear: *Every macrostate is associated with many possible microstates which we cannot distinguish from each other because we can't account for the motions of every one of* ~10^{23} *particles.* Let's do an exercise to see how this view of thermodynamic states leads to the fact that closed systems must evolve toward equilibrium...

We enclose a mole of a gas in a container which includes an airtight barrier that divides it in two. At the beginning, the barrier is in place, and all of the gas is confined to one side of the container. Then the barrier is removed, and the gas is allowed to expand into both chambers for a long period of time, after which the barrier is replaced.

1. **Clearly waiting a long time allows the gas in the chamber to come to equilibrium before the barrier is replaced. What does this tell us about the amount of gas in each chamber? Is this answer exact?**

Most people would claim that there is an equal amount of gas in each chamber, but would then admit that this is an approximation, because it would be folly to assume that a bunch of particles bouncing around randomly are divided exactly in half. We certainly couldn't tell if, in a collection of 10^{23} particles, there is an imbalance of 2 or 3... Or for that matter, 2 or 3 billion (10^9), which would still only account for a vanishingly-small percentage discrepancy – the pressure difference, for example, would not be measurable if one side got a billion particles more than the other.

Clearly the key to defining the macrostate is defining the tolerances to which we can measure, and these tolerances will be based on *percentages*, not absolute numbers. We can see how this works more clearly by talking about situations with numbers we can deal with more easily. From now on, we will need to keep in mind that we assume that individual particles bounce around in a random fashion.

2. **If we were to attach a tiny flag to one of the gas particles and wait for awhile as it moves around in the chamber, what would be the probability that it would be found on the right side of the chamber?**

Given the assumption of "random motion," we are forced to conclude that the probability of finding a specific particle in a specific half of the chamber is $\frac{1}{2}$.

3. **Suppose there is a total of 10 particles in the entire gas. If we let this gas move through both sides of the chamber for a long time and then replace the barrier, what is the probability that the gas will be divided exactly in half?**

Now our discussion takes a mathematical turn. To compute the probability, we need to know how many ways there are to arrange the particles such that they split evenly, and divide that number by all the ways that the particles can split. The numerator is found by computing a *combination*, which is often stated as "n-choose-r." In this case, this counts the number of unique ways we can select 5 particles out of 10 to place in one side of the container. We will not go into the details of this calculation, but will rather state the result:

$$n\text{-choose-}r = \binom{n}{r} = \frac{n!}{(n-r)!\ r!} \tag{6.4.9}$$

Plugging in our specific numbers here gives:

$$10\text{-choose-}5 = \binom{10}{5} = \frac{10!}{5!\ 5!} = 252 \tag{6.4.10}$$

The total number of ways to arrange the 10 particles is easy to compute. Every particle can go into one of two sides, so for one particle the number is 2. For two particles, the second particle has two choices for each of the first particle's two choices, giving a total of 4. Then the third particle gets two choices for each of the first four, giving 8, and so on. When we reach n particles, there is a total of 2^n ways to arrange them. For this example, we therefore get that the probability of seeing the 10 particles split exactly in half is:

$$P\,(5:5) = \frac{252}{2^{10}} = \frac{252}{1024} = 0.246 \tag{6.4.11}$$

There is less than a one-fourth probability that we will see this gas split exactly in half.

4. **The problem with 10 particles is that we can tell at a glance if they are divided evenly between the two chambers (it doesn't require any detailed examination to determine this). This is not a fair model for equilibrium, given what we discussed earlier. So let's include a margin for error. Let's say that we'll only recognize that the gas is not in equilibrium if the discrepancy from equilibrium is more than 10%. So for example, our blurry vision allows us to only distinguish distributions that are skewed by more than 60%-40%. So if there are 6 particles on one side of the barrier and 4 on the other, we don't immediately notice the difference, and we declare the gas "evenly-distributed" in accordance with equilibrium. That is, we define our macrostate to be defined by this 60-40 tolerance. What is the probability we will find the gas in equilibrium under this definition?**

Now there are more states that fall under our "roughly half-and-half" requirement than was true in the exact case. We have to account for the 10-choose-4 and 10-choose-6 results now, and add them into the probability calculation:

$$\begin{aligned} probability\,(10\ particles,\ 10\%\ margin) &= P\,(5:5) + P\,(6:4) + P\,(4:6) \\ &= \frac{\binom{10}{5}}{1024} + \frac{\binom{10}{6}}{1024} + \frac{\binom{10}{4}}{1024} \\ &= 0.646 \end{aligned} \tag{6.4.12}$$

Not surprisingly, the likelihood of seeing an "equilibrium" goes up greatly as our ability to distinguish detail gets worse. But we won't let it get any worse from here. Let's see what happens if we keep this 10% margin for distinguishing microstates but increase the number of particles.

5. **Using the same 10% margin-of-error definition of equilibrium as we used above, what is the probability of finding a 20-particle gas in equilibrium? Repeat the calculation one more time after doubling the particle number again to 40.**

For 20 particles, the 10% margin encompasses distributions of (8:12), (9:11), (10:10), (11:9), and (12:8), giving:

$$\begin{aligned} probability\,(20\ particles,\ 10\%\ margin) &= P\,(8:12)+P\,(9:11)+P\,(10:10)+P\,(11:9)+P\,(12:8) \\ &= \frac{\binom{20}{8}+\binom{20}{9}+\binom{20}{10}+\binom{20}{11}+\binom{20}{12}}{2^{20}} \\ &= 0.737 \end{aligned} \tag{6.4.13}$$

For 40 particles, the 10% margin encompasses distributions ranging from (16:24) to (24:16). Sparing the reader the math, the result is:

$$probability\,(40\ particles,\ 10\%\ margin) = 0.846 \tag{6.4.14}$$

Notice that keeping the same percentage margin, the likelihood that we will not be able to distinguish the macrostate from the exact 50-50 distribution gets larger as the number of particles grows. From this example with a very small particle number, it's not difficult to imagine that this probability becomes indistinguishable from 1 when the particle number approaches the enormous number of 10^{23}, even if we make the percentage margin significantly smaller.

So what have we learned here? We found that a macrostate is simply a result of not being able to distinguish microstates, and that the macrostate associated with the equilibrium state (the one where the gas is divided 50-50) includes an absurdly large percentage of all the possible microstates. We state it this way:

The macrostate associated with equilibrium is the one that includes the largest number of microstates.

At last we are in a position to explain why systems head toward equilibrium. Naturally the particles "don't know anything" about what they need to do to get the state to equilibrium, but they don't have to. All they need to do is move randomly, and as they pass through all the possible microstates, eventually the system evolves through the weird, rare microstates and then will spend virtually all of its time in one of the vast majority of microstates associated with equilibrium. From this point on, every time we look at it, it looks the same as before (even though the microscopic state has changed), so we declare it to be at equilibrium.

An analogy might help here. Suppose we have a cookie sheet with 100 pennies on it, all with heads facing up. Clearly this is a highly-ordered state, and the macrostate of "all heads" has only one microstate. Now we rap on the bottom of the cookie sheet, and some pennies randomly flip over. We did not hit it hard, so we can still tell that the pennies are not evenly-distributed between heads and tails, but it is clearly more disordered than before. If we keep hitting the sheet, eventually enough pennies flip between heads and tails that we estimate half of them to be heads and half to be tails. No matter how many more times we hit the sheet from now on, our estimate doesn't change, because the probability that a large number of coins will randomly flip to one variety is very small. In other words, the distribution of heads and tails reaches equilibrium.

We noticed that as we hit the sheet, the state of the pennies became more "disordered," and the equilibrium state was reached when this disorder reached its maximum. In this way, entropy is thought of as a measure of disorder of a system, and the second law of thermodynamics is merely a statement of probability – microstates are far, far more likely to randomly evolve into new microstates that are associated with an equilibrium macrostate than any others. Without getting too technical about how we count the number of microstates using positions and momenta of particles, the entropy of a system's macrostate is defined in terms of the logarithm of the number of microstates available to that macrostate (represented by Ω below, called the system's *multiplicity*):

$$S = k_B \ln \Omega \tag{6.4.15}$$

To see that this makes sense, consider the case of two systems at equlibrium being combined into one. If system A has Ω_A possible microstates, and system B has Ω_B possible microstates, then how many microstates are possible for the combined system? Well, every individual microstate of system A can be combined with every individual microstate of system B to create a new unique microstate for the combined system, so the combined system's total number of available microstates is the *product* of the multiplicities of the two individual systems:

$$\Omega_{combined} = \Omega_A \cdot \Omega_B \tag{6.4.16}$$

Plugging this into the entropy definition above and using the property of the logarithm gives:

$$S_{combined} = k_B \ln[\Omega_A \cdot \Omega_B] = k_B \ln[\Omega_A] + k_B \ln[\Omega_B] = S_A + S_B \tag{6.4.17}$$

This confirms that entropy is an extensive state function with this definition.

CHAPTER OVERVIEW

7: Fluid Mechanics

7.1: Static Fluids

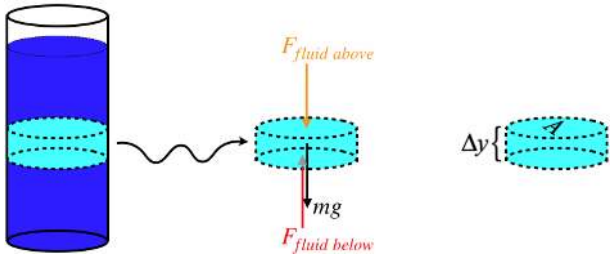
Pressure in Static Fluids

A *fluid* is a collection of matter that is not a solid. That is, the particles that comprise the matter are not strongly bound to each other in a fixed lattice. This includes the liquid, gaseous, and even plasma states of matter. We will actually need to restrict the properties of the fluids we examine to some extent, so that we can get some approximate results, but the usefulness of the applications of these models cannot be overstated.

A key property in all fluids is pressure. We have discussed this at length in the context of gases in thermodynamic equilibrium, but in the context of fluids, we need to loosen the restrictions we previously placed on this quantity. The main difference is that unlike gases at equilibrium, where we assumed that the pressure was a state variable that was the same in every cubic centimeter of a volume of gas at equilibrium, now we will allow the pressure to vary from one position to the next within a fluid. [*In truth, having a pressure that varies with position within a gas also lies in the purview of thermodynamics, though we did not delve sufficiently deeply into the subject to discuss it.*]

Consider a container of fluid in a gravitational field. If we draw a free-body diagram for a small section of that fluid (and exclude the horizontal forces), then three forces come to mind: The force of gravity on that section, the amount that the fluid above that section pushes down on it, and the amount that the fluid below that section pushes up on it. This diagram is really no different from one we would draw for a book in the middle of a stack of books.

Figure 7.1.1 – Free-Body Diagram for a Section of Fluid



The question is, how exactly does one region of fluid exert a force on another? Truthfully, it does not. Particles in one region are changing places with particles in the other region all the time. Also, as we know, if this fluid happens to be an ideal gas, then the particles aren't even interacting with each other. However, if we were to place a membrane between two regions, then particle reflections off the membrane *would* result in a force, and the resulting physics would be the same – the segment of fluid does not accelerate up or down. We therefore take the view that although over time it is not actually the same particles enclosed within the small volume, it behaves in precisely the same way as if we could somehow confine them to that volume, and our force analysis is valid.

As we first discussed in Section 5.5, we can express this particle-reflects-off-surface force in terms of a property of the fluid that we call pressure. The forces from the fluid above and below is therefore equal to the pressures of the fluid above and below multiplied by the cross-sectional areas, and the zero net force from the FBD gives:

$$\left.\begin{aligned} F_{fluid\ above} &= P_{above}A \\ F_{fluid\ below} &= P_{below}A \end{aligned}\right\} \quad \Rightarrow \quad P_{below}A = P_{above}A + mg \tag{7.1.1}$$

As we can readily see, the pressure varies with position in a fluid (namely, at differing heights) in the presence of gravity, even though the fluid is in equilibrium. Perhaps the reader is wondering how we got away with ignoring this effect in our study of thermodynamics. A sample calculation will answer this question.

Example 7.1.1

Compute the pressure difference between the top and bottom of a volume of 1 mole of diatomic nitrogen at $300K$ at atmospheric pressure ($101kPa$) confined to a cubical container.

Solution

Start by treating the gas as ideal and compute the volume:

$$V = \frac{nRT}{P} = \frac{(1mol)\left(8.31\frac{J}{mol\ K}\right)(300K)}{1.01\times10^5 Pa} = 0.0247m^3$$

The container is cubical, so from the volume we can determine the length of each side, which gives us the cross-sectional area:

$$L = V^{\frac{1}{3}} = 0.291m \quad \Rightarrow \quad A = 0.0848m^2$$

The molecular mass of N_2 is 28.0, so the mass of the enclosed gas is $28.0g$ This means that the weight of the gas enclosed is:

$$W_{gas} = mg = (0.028kg)\left(9.8\frac{m}{s^2}\right) = 0.274N$$

This weight, divided by the horizontal cross-sectional area equals the difference in pressure between the top and bottom:

$$\Delta P = \frac{0.274N}{0.0848m^2} = 3.23Pa$$

When measured against the average pressure of $\approx 10^5 Pa$, this difference is easily negligible.

Given that we need very large height differences to start seeing significant pressure differences in gases, we will primarily be dealing with liquid fluids from this point forward. More importantly, considering only liquids takes away another troublesome complication that exists for gases – liquids are largely *incompressible*. This means that we don't have to worry about the effect that increased pressure has on the mass density of the liquid – an important quantity, as we will see next.

Pascal's Law

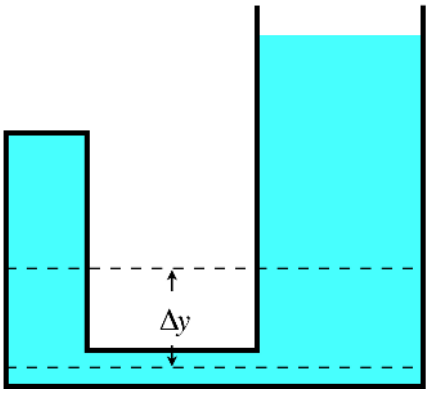
We can rearrange things in Equation 7.1.1 so that the pressure difference between two heights in the fluid are related to the *mass density* (ρ, mass per unit volume) of the fluid:

$$P_{below} = P_{above} + \frac{mg\Delta y}{A\Delta y} = P_{above} + \frac{mg\Delta y}{V} = P_{above} + \rho g \Delta y \tag{7.1.2}$$

The nice thing about this result is that it directly compares two pressures, and doesn't rely upon a choice of a particular disk of fluid for our free-body diagram. This means that if we choose a disk that has less cross-sectional area (but with the same thickness), we find the same difference in pressure. In other words, it turns out that the *pressure difference at two depths only depends upon the density of the fluid and the difference in height between the two depths*.

Just how robust is this conclusion? For example, suppose we have a fluid confined to a container like the one shown in the figure below. Is the pressure difference between the two heights indicated the same on both sides of the container?

Figure 7.1.2 – Comparing Pressures in Different Parts of a Continuous Fluid



Well, naturally the same analysis applies to both sides of this container, so the pressure *difference* is the same on both sides, but are they the *same two pressures* on both sides? Here we need to invoke the principal that pressure does not have a direction. That is, whatever the pressure is at the bottom of Δy, that results in a force pushing up on the section of fluid, it also results in a force pushing on a segment of fluid *to the side* of the position in question. That lateral segment is static just like the remainder of the fluid, so the pressure on both sides must be equal, which means that the pressure is the same everywhere along the lower dashed line.

We already concluded that the pressure difference between the two dashed lines is the same in both columns of fluid, so since the pressures are the same in both columns at the bottom line, they are also the same at the top line. We therefore conclude *Pascal's law*:

The pressures measured at all points at the same height in a continuous static fluid are equal.

Alert

It is important to make a note of all the qualifiers in Pascal's law. Besides being at equal heights, the fluid must be static – later we will get a slightly more complicated result when the fluid is moving. Also, the fluid must be continuous – if a solid barrier cuts one section off completely from the other, then the requirement that pressure is the same horizontally no longer holds. Another way for the fluid to not be "continuous" is for the fluid's density to not be the same everywhere. So two unmixed fluids in contact will result in the density changing across the boundary of the two fluids, and we cannot conclude that pressures at the same height within the different fluids are equal.

The Hydraulic Lift

The result we obtained above compares the pressure at any two points in the fluid, but by far the most common comparison is between the pressure at the top surface of the fluid and the pressure at some depth below that surface. Calling the pressure at the surface P_o and the depth below that surface d, Equation 7.1.2 becomes:

$$P(d) = P_o + \rho g d \tag{7.1.3}$$

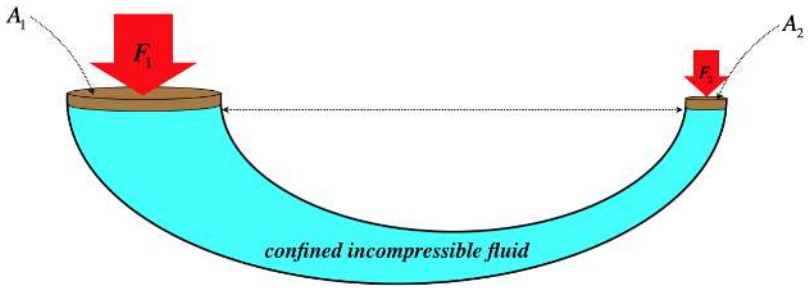
Digression: Rule of Thumb for Water

With the density of liquid water equal to $1000\frac{kg}{m^3}$, atmospheric pressure at sea level equal to about $10^5\frac{N}{m^2}$, and the constant g approximately equal to $10\frac{m}{s^2}$, it's easy to compute approximately how fast the pressure rises with depth under water – the pressure increases by an amount of one atmosphere with every $10m$ of added depth.

The salt water of the oceans is slightly denser than fresh water, so the pressure goes up slightly faster than this, and all certified scuba divers that use the English system of units are aware that one atmosphere of pressure is gained for every 33 feet of added depth. At the deepest point in the ocean, this amounts to a pressure of almost 1100 atmospheres!

The top surface of the fluid may just be exposed to the atmosphere, in which case P_o is simply $101kPa$ (at sea level). But the fluid could also be confined by a piston, which leads to an interesting application called a *hydraulic lift*. The idea is to confine the liquid with pistons of different areas on opposite ends of a continuous fluid, as shown in the following figure.

Figure 7.1.3 – Balancing Forces on a Hydraulic Lift



With the bottom surfaces of the two pistons in contact with the same continuous fluid at the same height, the pressure of the fluid is the same at both surfaces. The *forces* exerted by the fluid on the two pistons are not the same, however, because the areas are not equal. With pressures equal at both ends, the fluid exerts more force on the piston with a larger area, which means a heavier weight on the larger piston is balanced by a lighter weight on the other piston. This allows for a mechanical advantage in lifting a heavy weight with far less force than the weight, determined by the ratio of the piston areas:

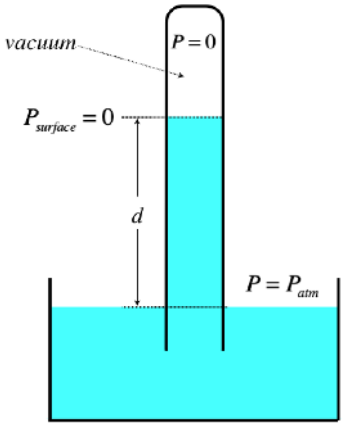
$$P_1 = P_2 \quad \Rightarrow \quad \frac{F_1}{A_1} = \frac{F_2}{A_2} \quad \Rightarrow \quad \frac{F_1}{F_2} = \frac{A_1}{A_2} \tag{7.1.4}$$

Manometers

Pascal's law also gives us a way to measure unknown pressures by comparing them to known values. A generic device of this kind is called a *manometer*. There are many specialized devices designed to measure specific pressures, such as a *barometer* (which measures the pressure of the atmosphere), and a *sphygmomanometer* (which measures blood pressure).

The simplest designs for manometers involve exposing one end of a continuous confined liquid to the region for which the pressure is to be measured, and the other end to a vacuum (i.e. a region where the pressure is effectively zero), then measure the difference in height of the two columns of the liquid and use the liquid's known density to compute the pressure difference. For example, here is a simple barometer:

Figure 7.1.4 – A Simple Barometer



In the figure above, the fluid is exposed to the atmosphere in the large tub, and the top of the vertical tube is evacuated. This is most easily achieved by filling the tube with the fluid, covering the open end, inverting it in the tube, then uncovering the open end so that the fluid runs out until equilibrium is

reached. [*Technically this does not result in an actual vacuum, as the particles in the liquid will leave the surface to go into vapor to fill this region, but this vapor pressure is very small compared to atmospheric pressure (less than 4% of atmospheric pressure at* $300K$*).*]

Applying Pascal's law, the pressure of the atmosphere at the exposed surface is the same as within the column at the same height, and applying our equation for pressure at depth, we have:

$$P_{atm} = P_o + \rho g d = \rho g d \tag{7.1.5}$$

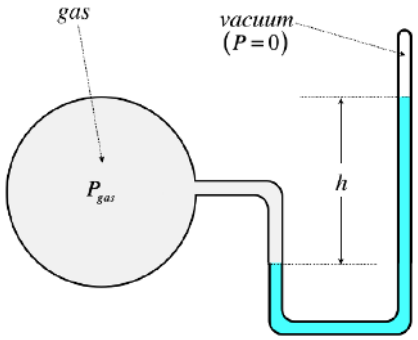
Thus we can measure the pressure of the atmosphere by measuring the height of the column of fluid and knowing its density. Note that this means that there is a maximum height to which we can raise this column by evacuating the space above it, and it points out an important common misconception. It is commonly believed that "suction" is an attractive force – that the more suction a pump can apply, the more it can "draw" a fluid. Here we see that all "suction" does is *remove* the pushing force due to one source of pressure, allowing the force exerted due to a pressure elsewhere to act unbalanced.

Plugging in the density of water and atmospheric pressure at sea level, we find that the farthest a column of water can be held up with a vacuum is a bit over $10m$. If we wish to pump water out of a well (or to a higher floor of a building), we can't "draw it up" by creating a vacuum above it from any depth greater than that. The approach that is therefore used instead is to pump the *bottom* of the column of water to increase its pressure. The only limit to depth from which the water can be extracted in this case is in the horsepower of the motor driving the pump.

Assuming we are working with pressures in the vicinity of an atmosphere, using water makes for a pretty tall manometer! The way to bring down this column size is to use a fluid that is more dense. The substance that is a liquid at room temperature and has the greatest density is mercury, and so ubiquitous was its use many years ago, that a unit of measurement for pressure was based on it. It turns out that 1 atmosphere will only support a column of Hg that is $760mm$ high. One atmosphere of pressure is therefore expressed in alternative units as "$760mmHg$" or "$760Torr$."

Here is a simple design for a simple gas pressure manometer (i.e. for enclosed gases of any variety, not just the atmosphere):

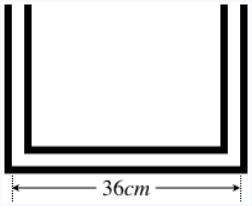
Figure 7.1.5 – A Simple Manometer



When the pressure is measured in this manner, we say that it is the *absolute pressure* that is measured. It seems silly to add the unnecessary extra word "absolute," but there is a practical reason for this. When it comes to gauges that measure pressure of gases in industrial applications, what the operator of a machine wants to know is if the pressure is close to exceeding the capacity of the device confining the gas. Well, any pressure acting from outside the device is helping to keep it from exploding, so what is needed is the pressure *difference* between the trapped gas and what is outside. We can measure this directly using the device above by opening the top of the tube so that there is no longer a vacuum there. Then the column height measures the difference between the trapped gas and the ambient pressure. This measurement of pressure is called *gauge pressure*. There is an analogy here with temperature – absolute temperature is measured on the kelvin scale, and zero is as low as it can go. The celsius scale has the same grading as kelvins, but puts the zero point at the freezing point of water, thus allowing for negative values. Similarly, gauge pressure places the zero point at the ambient (typically atmospheric) pressure, also allowing for a negative value of the pressure measurement.

Example 7.1.2

A U-shaped glass tube of constant cross-sectional area and two right angles has a base that is $36cm$ *long, and is open to the air at both ends, as shown in the figure below. Some water is poured into this tube, and it just exactly fills the bottom, horizontal part of the tube. An equal volume of oil with a density of* $0.80\frac{g}{cm^3}$ *is then slowly added to one side of the tube, such that the oil and water remain unmixed, in contact at a single point in the tube.*



a. *Find the distance from the contact point of the oil and water to the center of the glass tube.*
b. *The end of the tube with the water is connected to a confined volume of gas at an unknown pressure, and the contact point of the oil and water moves to the center of the tube. Find the gauge pressure of the gas.*

Solution

a. The contact point of the oil and water is a point where the pressure of the oil equals the pressure of the water (otherwise one would displace the other). Assuming this point of contact is somewhere in the bottom section of the tube (and it has to be, since the oil can't simply come in contact with the water without pushing it at least a little bit up the other side), the pressure for the oil an water are each related to how high their columns go up their respective sides:

$$\left.\begin{aligned} P_{water} &= P_{atm} + \rho_{water} g h_{water} \\ P_{oil} &= P_{atm} + \rho_{oil} g h_{oil} \end{aligned}\right\} \quad P_{water} = P_{oil} \quad \Rightarrow \quad \cancel{P_{atm}} + \rho_{water} g h_{water} = \cancel{P_{atm}} + \rho_{oil} g h_{oil} \quad \Rightarrow \quad h_{oil} = \frac{\rho_{water}}{\rho_{oil}} h_{water}$$

The density of water is $1.0 \frac{g}{cm^3}$, so the ratio of the height of the oil column to the water column is $\frac{5}{4}$. The liquids are incompressible, so since there are equal volumes of oil and water, the total length of the two columns must be $36cm$, which gives us a second equation, allowing us to solve for the heights of the two columns:

$$\left.\begin{aligned} h_{oil} &= \tfrac{5}{4} h_{water} \\ h_{oil} + h_{water} &= 36cm \end{aligned}\right\} \quad \Rightarrow \quad h_{oil} = 20cm \ , \quad h_{water} = 16cm$$

With the top of the column of oil $4cm$ higher than the top of the column of water, the contact point of the oil and water must be offset from the center by $2cm$.

b. With the contact point centered in the bottom section of the tube, the columns of water and oil are equal heights. Once again, the pressures of the fluids at the contact point are equal, but the pressures at the tops of the columns are not. The gauge pressure is the difference between the pressure of the gas and the atmosphere, so:

$$\left.\begin{aligned} P_{water} &= P_{gas} + \rho_{water} g h \\ P_{oil} &= P_{atm} + \rho_{oil} g h \end{aligned}\right\} \quad P_{water} = P_{oil} \quad \Rightarrow \quad P_{gauge} = P_{gas} - P_{atm} = (\rho_{oil} - \rho_{water}) g h = \left(-200 \frac{kg}{m^3}\right) \left(9.8 \frac{m}{s^2}\right) (0.18m) = -353 Pa$$

The gauge pressure is negative, which means the gas pressure is lower than atmospheric pressure.

7.2: Buoyancy

Archimedes's Principle

In the previous chapter, we discussed the pressure difference between the top and bottom of a section of the fluid. What happens if we replace that same section with a solid object? As we have seen, the presence of the solid object doesn't affect the pressure difference at the two heights, since the fluid is continuous and static. But in the case of a solid object, the higher pressure at the bottom and the lower pressure at the top result in actual forces on the bottom and top surfaces of the object. The result of these two forces is a total force by the fluid upward, which is called the *buoyancy force*.

We concluded in the previous section that the pressures on top and bottom differed just enough to balance the weight of the section of fluid, so we know precisely what the resulting force of this unbalanced pressure is – it is the weight of the fluid that would be in that section if the solid object was not there. Put concisely, we have *Archimedes's principle*:

The buoyancy force on an object in a fluid equals the weight of the fluid displaced by that object.

As simple as this seems, it is very easy to get confused about this force. The main source of confusion tends to be distinguishing the buoyancy force from the net force on the object (which also experiences gravity). Here are some secrets to winding one's way through the daunting mazes commonly encountered regarding buoyancy:

- Keep in mind that buoyancy is *just* the force from the fluid on the object – it is completely independent from the gravity force on the object, and it is not the net force. Draw force diagrams whenever possible, with separate force vectors for gravity and buoyancy, and apply Archimedes's principle *only* to the buoyancy force vector.
- Apply Archimedes's principle *very strictly* – the volume of fluid displaced does not always equal the volume of the object (it has to be completely submerged for that).
- Be careful about drawing conclusions based on the density of the object only – whether an object sinks or floats, it experiences the buoyancy force described by Archimedes's principle.

Assuming the object in the fluid is completely submerged, then its full volume displaces fluid. This means it feels a buoyancy force equal to the weight of fluid that occupies that same volume. The net force on such an object is the buoyancy force up minus the gravity force down, so if the object weighs more than the displaced fluid, it sinks. Given that the object and the displaced fluid have equal volumes, we can just as easily compare their mass-to-volume ratios to determine which is heavier. This ratio for the fluid is simply its uniform density. For the object, it is its *average* density. The distinction is that the object may, for example, be hollow. This explains how an aircraft carrier, made of materials significantly denser than water, can float – the hollow parts of the vessel that contain only air reduce its average density greatly.

Examples

The full flower of the tricky topic of buoyancy only becomes apparent with examples, so here are a few...

Example 7.2.1

Two identical hollow metal cubes are in the Earth's atmosphere. One cube contains helium (at one atmosphere of pressure), and the other a vacuum. Which of these cubes experiences the greatest buoyancy force, and which registers the lowest weight on a scale?

Solution

The cubes are identical, so they displace the same volume of air, and the buoyancy force is the same on both. Because the buoyancy forces are equal, the cube that registers the higher weight on the scale will be the one that possesses the greater mass. Helium has more mass than a vacuum, so the cube filled with helium will register a larger weight on the scale. [Note: Helium does not "naturally rise" and pull things upward. It has mass, like all matter, and is therefore subject to gravity like everything else!]

Example 7.2.2

A tub of water sits on a spring scale. A toy boat is floated on the surface of the water, and the scale is read. Later, the boat is submerged until it takes on enough water to sink and settle on the bottom. Still later, the boat is removed from the water and placed on the scale beside the tub. Order the weighings from least to greatest.

Solution

The simple answer is that all of the forces between the water, the tub, and the boat are internal to the water + tub + boat system, and therefore all have Newton's third law pairs that cancel each other. The force external to the system, from the spring scale, is the same in all three cases.

*While this makes sense, it may be a bit unsatisfying, so let's look very briefly at the details of the internal forces. Strange as it seems, buoyancy forces come in third-law pairs, just like every other force, which means that while the water pushes up on a floating boat, the floating boat pushes down equally on the body of water. The scale beneath the tub of water must be great enough to balance all the other forces on it, which includes the weight of the tub + water **plus** the buoyancy force down on the water by the boat. Since the boat isn't accelerating up or down, the buoyancy force up on the boat must equal its weight, which means that the buoyancy force down on the water (and balanced by the scale) is exactly the weight of the boat.*

One might try to argue that the force down on the tub by the water must equal the pressure of the water at the bottom multiplied by the area of the bottom of the tub, and aren't these two things the same whether the boat is in the water or not? No! When the boat is floating, it is displacing water. Where does this displaced water go? Nowhere - it just gets deeper! The increased depth changes the pressure at the bottom just enough to contribute an additional force equal to the buoyancy force on/by the boat, and with the boat floating, this equals the weight of the boat.

Example 7.2.3

A small balloon containing air (which we can treat as an ideal gas) has negligible mass, and is attached to a rock with a string. This combination is thrown into the middle of Lake Tahoe, and the hanging rock pulls down on the balloon enough to submerge one-eighth of its volume. A scuba diver grabs the balloon and submerges it, swimming downward. Find the depth to which the balloon must be submerged such that it will sink when it is released. Assume the temperature of the water doesn't change appreciably during the dive.

Solution

The gravity force on the rock equals the weight of water occupying one-eighth of the balloon's volume in the atmosphere. Therefore, when the balloon no longer displaces at least one eighth of its starting volume, the buoyancy force will be insufficient to keep it from sinking. But what would cause the balloon to stop displacing so much volume? The pressure on the outside of the balloon approximately balances the pressure inside the balloon (not counting the force from the balloon's elasticity), and as the balloon goes deeper, the outside pressure gets greater. With the temperature and the number of moles not changing, the increased pressure must be balanced by decreased volume.

The volume of the balloon must decrease by a factor of 8 to keep sinking, so the pressure of the water must equal 8 times atmospheric pressure. We found in the digression of the previous section that 1 atmosphere of pressure is added for approximately every 10 meters of depth in fresh water. The balloon starts at the surface with 1 atmosphere, so it needs 7 more atmospheres to reach the goal of one eighth of its original volume – the diver must push the balloon $70m$ below the surface in order to pollute the bottom of Lake Tahoe with it.

While this answer is in keeping with information we have been given to this point, it is not actually correct, and is off by a fairly large margin! The problem lies not in the logic of the solution, but in the assumption that the pressure increases by the amount of the atmosphere with every 10 meters of added depth. This is incorrect because that rule-of-thumb only applies to atmospheric pressure at sea level. The pressure of the atmosphere at the elevation of Lake Tahoe is only about 79% of what we refer to as "atmospheric pressure." This means that at the surface of the lake the pressure of the air in the balloon is 0.79 atmospheres, and in order to increase the pressure by that much, the balloon must be submerged only $7.9m$. Following the logic of the original solution, in order to increase the pressure to eight times the pressure at the lake surface, the balloon must be submerged $7 \times 7.9m \approx 55m$.

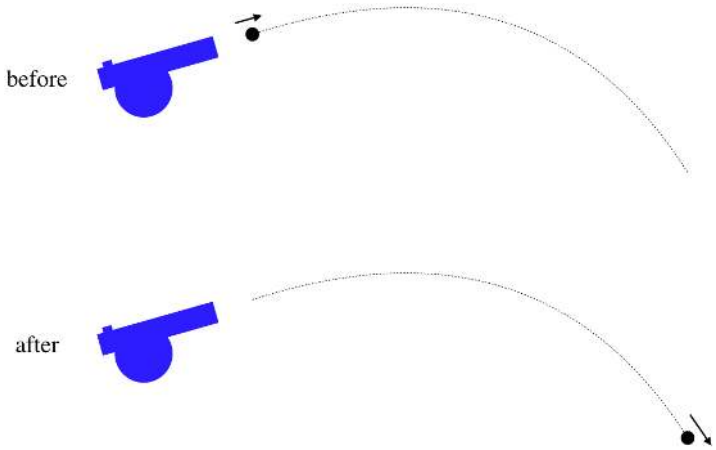
7.3: Fluid Dynamics

Steady-State

As with thermodynamics, fluids involve trillions of trillions of particles with freedom of movement, and also like thermodynamics, we would rather not treat them one particle at a time. One useful concept (which we have used previously but never discussed explicitly) is the concept of “steady-state.” We will find this idea to be extremely useful in our study of fluids, so we will take a closer look at what it means.

Consider the energy of a cannonball projectile.

Figure 7.2.1 – Before and After for a Single Cannonball



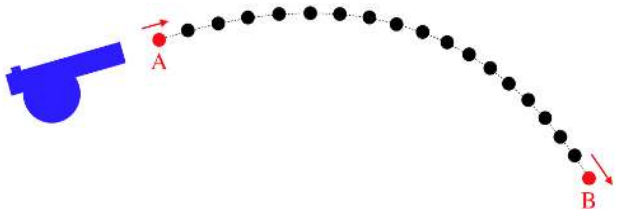
We know that energy is conserved for this cannonball as it flies through the air (we will ignore air resistance here), and we express this mathematically in terms of a zero change from "before" to "after."

$$\Delta KE + \Delta PE = 0 \tag{7.3.1}$$

Note especially that the Δ here means "after minus before," that is, it expresses a difference of quantities that are measured at different times.

Suppose now that cannonballs can be launched in rapid succession.

Figure 7.2.2 – Common Trajectory for Stream of Cannonballs



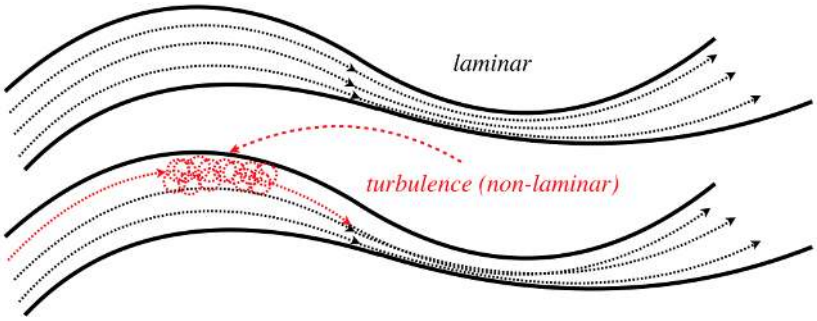
Of course they all follow the same trajectory, and this fact allows us to write down a *new* equation that looks identical to the one above, but has a different interpretation. In this case, we are not comparing the energies of a single cannonball at different times, but rather the energies of two *different cannonballs at the same time*. This leads us to a different type of model from what we used before. Previously we dealt with energy conservation from a before/after perspective, but with a steady flow, we are now able to use the concept of *steady-state* to make our comparisons at different positions rather than different times. Steady-state is a condition for flowing systems whereby the conditions at a fixed position remain constant in time. In other words, if you take a snap shot at two different times of the system as a whole, you would not be able to distinguish one snap shot from the other.

Assumptions: Incompressibility and Laminar Flow

As you might imagine, with particles in a fluid free to move around, it turns out that their motions can be very complicated. As with everything we study in physics, we need to make some simplifying assumptions so that we can reach some testable conclusions, and it is no different with fluid flow. We have already made one simplifying assumption in the previous section – we will treat fluids as though they are incompressible. Again, this means that we have to be very careful about applying these results to gases (though we can sometimes do so, at least approximately), but mostly the model we should keep in mind is that of a liquid.

The second simplifying assumption has to do with how the particles move relative to one another as they flow in steady-state. We will require that the flow of the fluid be *laminar* (also called *streamline*). Essentially what this means is that adjacent particles are, to a large degree, moving along parallel paths. When the flow gets too fast, or the shape of whatever is confining the motion of the flow changes too dramatically (actually, it is a combination of these things), then swirls and eddy currents arise, and *turbulence* ensues.

Figure 7.2.3 – Laminar vs. Turbulent Flow



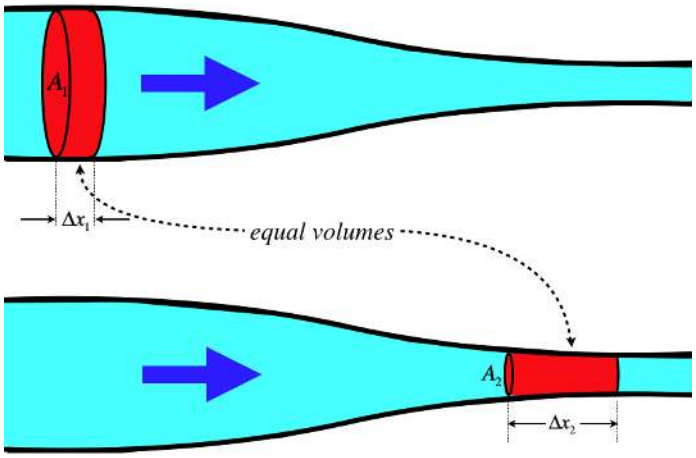
The phenomenon of turbulence arises from effects that are *non-linear*, which makes it incredibly difficult to model and study. A non-linear phenomenon can loosely be described as one which feeds back on itself, and is therefore extremely sensitive to the conditions placed upon it. Most people are familiar with the unpredictability of weather patterns, and have perhaps heard of the famous "butterfly effect." This is an example of a non-linear system whose evolution is extremely sensitive to initial conditions. As important of a physics topic as turbulence is, we simply are not equipped to discuss it in any detail in this class ("not equipped" meaning we don't have a supercomputer to work with).

It should be noted that laminar flow does *not* assume that the inner walls of the vessel carrying the fluid are free of friction. For the flow to be streamline, the neighboring particles only have to move parallel to each other, they don't have to move at the same speed. For a tube with friction against its inner surface, the particles will be moving slower than the particles near the center of the tube, but the flow is still streamline if they are moving parallel to each other.

Current, Volumetric Flow Rate, and Continuity

In fluid statics, the assumption of incompressibility was helpful in deriving a simple relationship between pressure and depth, because the density ρ remained constant throughout the fluid. In fluid dynamics, it has another interesting effect, when the vessel through which the fluid is flowing changes its cross-sectional area. Following a section of fluid through such a flow reveals:

Figure 7.2.4 – Incompressibility in a Bottleneck



With the fluid unable to compress, a given collection of particles in the fluid must morph their shape in the conduit such that their volume remains the same. This means that as the cross-sectional area of the conduit gets smaller, the volume of the fluid section we are tracking must get longer. That is:

$$V_1 = V_2 \quad \Rightarrow \quad A_1 \Delta x_1 = A_2 \Delta x_2 \tag{7.3.2}$$

Now we invoke our steady-state condition on this as follows: Let's wait a period of time equal to Δt, which is the amount of time it takes all the fluid that is initially within volume V_1 (the wide, short, red cylinder in the diagram above) to exit that volume. It should be clear

from the fact that this is operating at steady-state that in this same period of time, all the fluid in volume V_2 (the thin, long, red cylinder in the diagram above) has just enough time to exit that volume. The speed of the flow through each volume is the distance the fluid at the rear of the cylinder must travel to get to the front of the cylinder, divided by the time required, so we can relate the velocities of the fluid at the two points in question:

$$A_1 \frac{\Delta x_1}{\Delta t} = A_2 \frac{\Delta x_2}{\Delta t} \quad \Rightarrow \quad A_1 v_1 = A_2 v_2 \tag{7.3.3}$$

The value of each side of this equation is the rate at which volume of fluid is passing a fixed point. This is commonly referred to as *volumetric flow rate*, or more generically as *current* (which technically can be any kind of flow rate, not just volume), and can be written as:

$$I = \frac{dV}{dt} = Av \tag{7.3.4}$$

What Equation 7.3.3 tells us is that for an incompressible fluid, the volumetric flow rate is the same everywhere in the fluid. This relation is referred to as the *equation of continuity*, which can be alternatively expressed as:

$$I = \frac{dV}{dt} = Av = constant\ throughout\ the\ fluid \tag{7.3.5}$$

This expresses a conservation principle of sorts – all the fluid that comes into a given region in a given period of time also passes out of that region in the same period of time – there is no build-up or loss of fluid in that region.

Energy Density

In studying the dynamics of fluids, we seek to describe mathematically how the various physical properties of a moving fluid are related to each other. Judging from our study of fluid statics, there are already a number of properties that we know must play a role:

- the density of the fluid, ρ
- the height at some position in the fluid, y
- the pressure at some position in the fluid, P

Now that we are dealing with *dynamics*, we need to add this to the list:

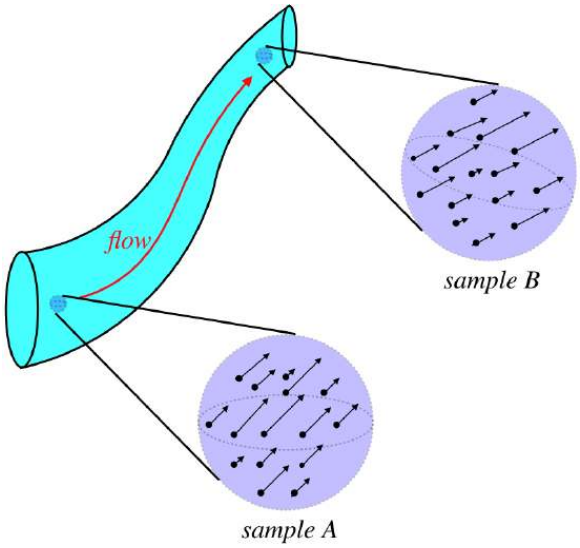
- the speed of the fluid at some position, v

Whatever relationship we find between these quantities must reduce to what we found previously in our discussion of statics, when we set the speed equal to zero.

We are not worried about directions here, so a complicated vector analysis is (thankfully) unnecessary. We therefore turn to energy conservation. We also will make use of our steady-state condition, which means that our energy conservation equation will involve a comparison of two different positions in a continuous fluid, rather than the fate of a single sample of fluid at two different times. This will certainly fit nicely with our study of fluid statics, where we also compared properties of a continuous fluid at different places.

There is one problem we need to overcome in our application of energy conservation, however. What energies are we comparing? Each particle in the fluid has some amount of energy, but there is a degree of randomness to their motions, so two particles that are side-by-side in the fluid do not necessarily have the same energy. Really what we want is to compare the *average* energies of particles. To do this, we need to compare energies of a collection of particles. Suppose we take a tiny sampling of the energies of particles at two different positions in a fluid:

Figure 7.2.5 – Sampling Small Volumes of Particles in a Fluid



The average energy per particle in each of these samplings is found by adding up their individual energies and dividing by the number of particles:

$$\langle E\rangle = \frac{(KE_1 + PE_1) + (KE_2 + PE_2) + \ldots}{N} \tag{7.3.6}$$

Assuming no work is done on the particles from position A to position B by forces other than those accounted-for in the potential energy, these averages should be equal for each sample. The number of particles is difficult to work with, so we note that we are assuming that this fluid is incompressible, which tells us that the particle density is the same everywhere in the fluid. So rather than use the average energy per particle as a basis for comparison, we can use the average energy per volume:

$$\frac{(KE_1 + PE_1) + (KE_2 + PE_2) + \ldots}{V} = \langle E\rangle \frac{N}{V} \tag{7.3.7}$$

We call this the *average energy density* of the fluid for the sample selected. We want to study the fluid at a *specific position*, so we need to make our sample volume smaller and smaller. As we do so, the energy contained therein also gets smaller and smaller, but the ratio of energy and volume converges to the *energy density* at that position:

$$\mathcal{E} = \lim_{V\to 0} \frac{E}{V} \tag{7.3.8}$$

We can break this energy density into the kinetic and potential parts. We'll start with kinetic...

When the focus is narrowed down to a single point (i.e. the volume goes to the zero limit), the average speed of the fluid doesn't change, but with the constant density, the mass gets ever smaller. We have, therefore, that the kinetic energy density is written in terms of the mass density and speed:

$$\frac{d(KE)}{dV} = \frac{d}{dV}\left(\frac{1}{2}mv^2\right) = \frac{1}{2}\frac{dm}{dV}v^2 = \frac{1}{2}\rho v^2 \tag{7.3.9}$$

The only potential energy we will be dealing with here is gravitational. Once again, we see that only the mass gets small in the limit as the volume vanishes, that is, the height doesn't change. so we have:

$$\frac{d(PE)}{dV} = \frac{d}{dV}(mgy) = \frac{dm}{dV}gy = \rho gy \tag{7.3.10}$$

So the full mechanical energy density is:

$$\mathcal{E} = \frac{1}{2}\rho v^2 + \rho gy \tag{7.3.11}$$

Energy Conservation

So now we have a way of comparing energy at different positions in a fluid – we compare the energy densities at those points. Before we just launch into invoking energy conservation, there is one thing we need to address. Above we made the qualification, "*Assuming no work is done on the particles from position A to position B by forces other than those accounted-for in the potential energy...*" Are there any

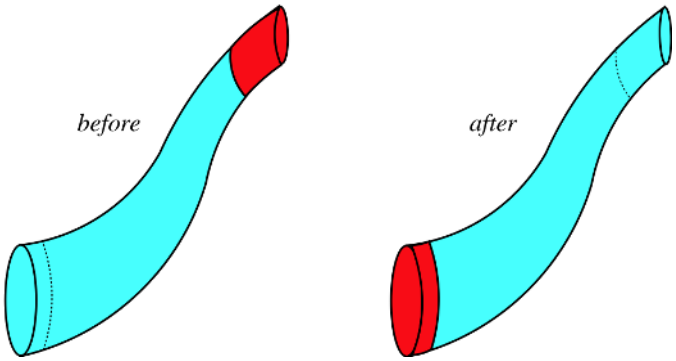
forces present that could be responsible for work like this? Well, there could be friction along the walls of the vessel, but in true Physics 9 fashion, we will treat this as negligible. There is, however, one source of work that we can't ignore.

We saw in the case of static fluids that the pressure can vary from one place to the other. Now that the fluid can move, if the pressure in front of a moving segment of fluid is different from the pressure behind it, the net force created by the difference in the pressures can do work. We therefore need to include the work done by pressure differences in our energy conservation calculations. Fortunately, from our work in thermodynamics, we are already familiar with how to compute work done by pressure:

$$dW = PdV \tag{7.3.12}$$

The difference in this case is that the work done is not acting to compress or expand a gas – the volume change in this case is just the displacement of an incompressible segment of fluid. The picture we have then is of a fluid passing through a conduit which in general changes heights and cross-sectional areas. Because we are more comfortable dealing with energy conservation from a before/after perspective, we'll start with that, then invoke steady state to extend our conclusions to compare different positions in the fluid.

Figure 7.2.6a – Fluid Flow Through a Pipe



The fluid we are watching here is blue. The red sections also contain fluid, but we are confining ourselves to watching what happens to the blue segment of fluid. In the time between "before" and "after," the fluid flows up through the pipe, and the same amount of fluid that vacates the bottom portion of the pipe (turning that section red) fills in the top portion of the pipe (turning it from red to blue).

This blue segment of fluid is experiencing a force due to the adjacent fluid, so the bottom face of the blue segment of fluid experiences a force in the direction of motion of the fluid segment, while the top face of the blue segment of fluid experiences a force opposing its motion. The total work done on the blue segment of fluid is is therefore:

$$dW = P_{bottom}dV_{bottom} - P_{top}dV_{top} \tag{7.3.13}$$

The fluid is incompressible, which means that the volumes for top and bottom (represented by the red region) are equal to each other, which tells us that the work done on this segment of fluid is:

$$dW = \left(P_{bottom} - P_{top}\right)dV \tag{7.3.14}$$

Now we are ready to invoke energy conservation. The work done on the system equals the change in its mechanical energy, which is a change from one small quantity to another (both top and bottom involve a tiny volume), so:

$$dW = d\left(KE_{top}\right) - d\left(KE_{bottom}\right) + d\left(PE_{top}\right) - d\left(PE_{bottom}\right) \tag{7.3.15}$$

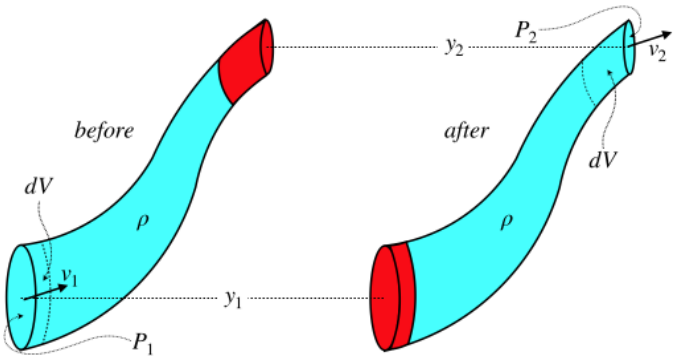
Dividing every term in this equation by dV and plugging in Equations 7.3.9, 7.3.10, and 7.3.14 gives:

$$P_{bottom} - P_{top} = \frac{1}{2}\rho\left(v_{top}^2 - v_{bottom}^2\right) + \rho g\left(y_{top} - y_{bottom}\right) \tag{7.3.16}$$

We can write this result (called *Bernoulli's equation*) two ways, as with our usual energy conservation principle:

$$\begin{aligned} &\Delta P + \tfrac{1}{2}\rho\Delta\left(v^2\right) + \rho g\Delta y = 0 \\ &P_1 + \tfrac{1}{2}\rho v_1^2 + \rho g y_1 = P_2 + \tfrac{1}{2}\rho v_2^2 + \rho g y_2 \end{aligned} \tag{7.3.17}$$

Figure 7.2.6b – Fluid Flow Through a Pipe

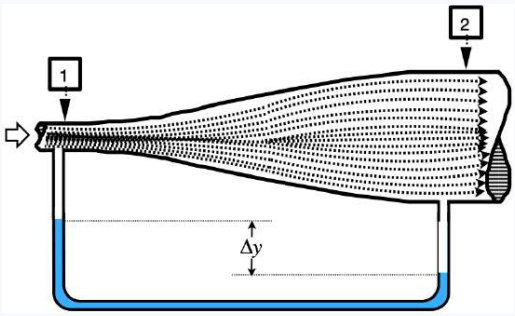


Note that although we derived this using a "before/after" approach, the Δ refers to a difference in positions, in keeping with the steady-state assumption.

Recall we wanted to check the result we got in fluid statics against this result for $v = 0$. Sure enough, if we put in $v_1 = v_2 = 0$, we get back to Equation 7.1.2.

Example 7.3.1

Air flows from left to right through a horizontal pipe that has different diameters in two different sections, as shown in the diagram. The narrow section has a radius of $5.0\ cm$, and the wider section has a radius of $20\ cm$. At point 1, the air is flowing with a velocity of $10\ m/s$, and has an absolute pressure of $150,000\ Pa$. A narrow U-shaped tube has been attached to the air tube, and this U-shaped tube is filled with water. Note that no air actually flows into or out of the U-shaped tube, but instead flows right across its openings. The pressures at points 1 and 2 are different, so the water in the U-shaped tube is not level. Assume that the transitions between the sections have a negligible effect on the frictionless flow of air, and that the density of air is a constant $1.2\frac{kg}{m^3}$ throughout the pipe. Also, despite how the diagram looks, assume that the difference in height of the connection points of both ends of the U-shaped tube is negligible.



a. *Find the speed of the air and its pressure at point 2.*
b. *Find the difference in heights (Δy) of the water in the U-shaped tube.*

Solution

a. We can apply Bernoulli's equation to the moving air in the pipe and in the U-shaped tube, but we must do so separately, since they are two separate fluid systems. In order to work out Δy, we need to know the pressure difference at the two ends of the U-shaped tube, which means we have to work out the air flow part first. Start by finding the speed of the air at point 2, using the continuity equation (Equation 7.3.3):

$$A_1 v_1 = A_2 v_2 \quad \Rightarrow \quad v_2 = \frac{A_1}{A_2} v_1 = \frac{\pi r_1^2}{\pi r_2^2} v_1 = \frac{(20cm)^2}{(5.0cm)^2}\left(10\frac{m}{s}\right) = 6.25\frac{m}{s}$$

Applying Bernoulli's equation with no change in height gives us the pressure at point 2:

$$\Delta P + \frac{1}{2}\rho\Delta\left(v^2\right) + 0 \quad \Rightarrow \quad P_2 = P_1 + \frac{1}{2}\rho\left(v_1^2 - v_2^2\right) = 150,000Pa + \frac{1}{2}\left(1.2\frac{kg}{m^3}\right)\left[\left(10\frac{m}{s}\right)^2 - \left(0.625\frac{m}{s}\right)^2\right] = 150,060Pa$$

b. The difference in pressures at the two ends of the U-shaped tube is $60\ Pa$, and this determines the difference in heights of the columns of water. The water in these tubes is not moving, so we can plug in zero for the velocities (or equivalently, just use the fluid statics formula):

$$\Delta P + 0 + \rho g \Delta y = 0 \quad \Rightarrow \quad |\Delta y| = \left| \frac{\Delta P}{\rho g} \right| = \frac{60 Pa}{\left(1000 \frac{kg}{m^3}\right)\left(9.8 \frac{m}{s^2}\right)} = 0.61 cm$$

Index

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: UCD: Physics 9B – Waves, Sound, Optics, Thermodynamics, and Fluids

Webpages: 54

All licenses found:

- CC BY-SA 4.0: 83.3% (45 pages)
- Undeclared: 16.7% (9 pages)

By Page

- UCD: Physics 9B – Waves, Sound, Optics, Thermodynamics, and Fluids - *CC BY-SA 4.0*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Waves - *CC BY-SA 4.0*
    - 1.1: Wave Mathematics - *CC BY-SA 4.0*
    - 1.2: Wave Properties - *CC BY-SA 4.0*
    - 1.3: Energy Transmission - *CC BY-SA 4.0*
    - 1.4: Superposition and Interference - *CC BY-SA 4.0*
    - 1.5: Standing Waves - *CC BY-SA 4.0*
  - 2: Sound - *CC BY-SA 4.0*
    - 2.1: Fundamentals of Sound - *CC BY-SA 4.0*
    - 2.2: Doppler Effect - *CC BY-SA 4.0*
    - 2.3: Interference Effects - *CC BY-SA 4.0*
  - 3: Physical Optics - *CC BY-SA 4.0*
    - 3.1: Light as a Wave - *CC BY-SA 4.0*
    - 3.2: Double-Slit Interference - *CC BY-SA 4.0*
    - 3.3: Diffraction Gratings - *CC BY-SA 4.0*
    - 3.4: Single-Slit Diffraction - *CC BY-SA 4.0*
    - 3.5: Thin Film Interference - *CC BY-SA 4.0*
    - 3.6: Reflection, Refraction, and Dispersion - *CC BY-SA 4.0*
    - 3.7: Polarization - *CC BY-SA 4.0*
  - 4: Geometrical Optics - *CC BY-SA 4.0*
    - 4.1: Images - *CC BY-SA 4.0*
    - 4.2: Magnification - *CC BY-SA 4.0*
    - 4.3: Spherical Reflectors - *CC BY-SA 4.0*
    - 4.4: Spherical Refractors - *CC BY-SA 4.0*
    - 4.5: Thin Lenses - *CC BY-SA 4.0*
    - 4.6: Multiple Optical Devices - *CC BY-SA 4.0*
    - 4.7: Wrap-Up - *CC BY-SA 4.0*
  - 5: Fundamentals of Thermodynamics - *CC BY-SA 4.0*
    - 5.1: Temperature - *CC BY-SA 4.0*
    - 5.2: Thermal Expansion - *CC BY-SA 4.0*
    - 5.3: Heat Capacity and Phase Transitions - *CC BY-SA 4.0*
    - 5.4: Modes of Heat Transfer - *CC BY-SA 4.0*
    - 5.5: Thermodynamic States of Ideal Gases - *CC BY-SA 4.0*
    - 5.6: Equipartition of Energy - *CC BY-SA 4.0*
    - 5.7: Thermodynamic Processes - *CC BY-SA 4.0*
    - 5.8: Special Processes - *CC BY-SA 4.0*
  - 6: Applications of Thermodynamics - *CC BY-SA 4.0*
    - 6.1: More Processes - *CC BY-SA 4.0*
    - 6.2: Engines and Thermal Efficiency - *CC BY-SA 4.0*
    - 6.3: Entropy - *CC BY-SA 4.0*
    - 6.4: The Second Law of Thermodynamics - *CC BY-SA 4.0*
  - 7: Fluid Mechanics - *CC BY-SA 4.0*
    - 7.1: Static Fluids - *CC BY-SA 4.0*
    - 7.2: Buoyancy - *CC BY-SA 4.0*
    - 7.3: Fluid Dynamics - *CC BY-SA 4.0*
  - Back Matter - *Undeclared*
    - Index - *Undeclared*
    - Glossary - *Undeclared*
    - Detailed Licensing - *Undeclared*