Notes follow and parts taken from *College Physics* (Wilson & Buffa) and *Physics* (6th Edition, Cutnell & Johnson)

Thermodynamics

Thermodynamics is the study of heat transfer. We will need to define a few terms to discuss this subject. First, a **system** will mean an amount of matter inside real or imaginary boundaries. The volume enclosed by the boundaries can also change. The system may do work, or have work done on it. It may be **thermally isolated** (no heat transfer in or out – like an ideal thermos bottle) or it may be in contact with a **heat reservoir** which will supply any amount of heat to the system (without changing the reservoir’s temperature).

The systems will be described by **equations of state** which will relate things like pressure, volume, temperature, and mass. The ideal gas law is therefore an equation of state. We will sometimes follow the evolution of a gas under changing situations by following its progress on a **p-V diagram**. If we know the pressure & volume of the gas, that’s a single point on the diagram. If one of those variables changes, the system moves to a new point. As long as the amount of mass doesn’t change, the knowledge of $p$ and $V$ will give us the temperature $T$ as well (for most gases). Depending on the circumstances, we may plot any two of these three variables.

**Processes** involve changes in the three variables describing our system. A certain process might change our variables from $(p_1, V_1, T_1)$ to $(p_2, V_2, T_2)$. We can have **irreversible** processes and **reversible** processes. An irreversible process is basically one which happens too quickly for the intermediate states to be in equilibrium. If we make our changes very slowly (for example, changing the temperature over a very long time), we should be able to track the progress of the system and it should remain in equilibrium as we move from our initial condition to the final one. This process would be called reversible. In the real world, all processes are at least partially irreversible.

The First Law of Thermodynamics

Using the conservation of energy, we can make some general statements about how systems must behave when heat is added or removed. One of those is called the **first law of thermodynamics**, which says that the energy added as heat will either end up as internal energy of the system or as work done on it. Mathematically, we write this as

$$Q = \Delta U + W$$

where $Q$ is the heat added to or removed from the system, $\Delta U$ is the change in its internal energy, and $W$ is the work done on or by the system. Positive values for $Q$ indicate heat added to the system, while negative values would represent heat removed. Similarly, positive work is done by the system, while negative work is done on the system. Keep in mind that $Q$ and $W$ are not properties of the system, like its temperature (and therefore $\Delta U$) are. They represent changes undergone by the system. We can use this to show that the change in internal energy is independent of the path taken to change it, just as we saw that the change in an object’s gravitational potential energy is independent of the path taken. Only the difference between the initial and final points matters. Several special cases of these circumstances are explained below.
**Isobaric processes** are processes in which the system’s pressure does not change. On a $p$-$V$ diagram, an isobaric process would be represented by a line perpendicular to the $p$ axis (obviously, since pressure isn’t changing). If pressure remains constant (and we don’t add or subtract matter), the only things left are volume and temperature. These are related by $V/T = Nk_B/p$. Everything on the right hand side is a constant in our experiment, so any change in $V$ must be accompanied by a change in $T$ of the same size. If the temperature (in Kelvin, of course) is doubled, the volume will also double so that the ratio remains constant. This change in volume will do work (because there is a pressure). We know that $W = F \Delta x$ (the typical design of this experiment has the gas expanding against a piston, so the force is parallel to the displacement and the cosine factor is just one), and that pressure $= F/A$. We use this to rewrite the work formula as $W = pA \Delta x$, but we then recognize that the area of the piston multiplied by its displacement gives us the change in volume, so we can just as easily say that $W = p \Delta V = p(V_2 - V_0)$. This value will be the same as the area under the line on the $p$-$V$ diagram described above. If we don’t have an isobaric process, the work is still the area under the curve, but it’s obviously more complicated to find, and now dependent upon the path taken on the $p$-$V$ graph. For an expanding gas, with a final volume greater than its initial volume, the work will be positive, which means the system does work rather than has work done on it. We can now write (if pressure doesn’t change) the first law as $Q = \Delta U + p \Delta V$. This also tells us that compressing a gas at constant pressure requires you to do work on the gas, and means heat flows out of the gas as its temperature decreases.

**Isometric processes** are those in which the volume of the system is kept constant. These would be represented by lines perpendicular to the $V$ axis on our $p$-$V$ graph. Without a change in volume, no work will be done. In this case, all heat goes into increasing the temperature (or internal energy) of the system and we can write $Q = \Delta U$.

In an **isothermal process**, the temperature is kept constant. This path will look like a hyperbola, since we can rewrite the ideal gas law (with all the constants on the right) as $pV = Nk_B T$. If the temperature can’t change, $\Delta U$ will be constant so that $Q = W$. All the added heat is used to do work.

**Adiabatic processes** are ones where $Q = 0$, meaning no heat is transferred into or out of the system. The system has to be thermally isolated for the process to be adiabatic. The alternative to total thermal isolation would be changing conditions so quickly that heat doesn’t have time to move into or out of the system. We can write $W = -\Delta U$, which explains why compressing a gas quickly (meaning $W$ is negative) gives a positive change in internal energy, which means a temperature rise. Similarly, suddenly reducing the pressure on a system will cause a temperature drop. If your refrigerator is set at just the right temperature, you can see ice crystals form in a soft drink as soon as you unscrew the top (reducing the pressure).

**The Second Law of Thermodynamics**

The second law of thermodynamics provides what is sometimes called the “arrow of time”. It’s one of the ways we can tell that a movie is running backwards. Essentially, many processes are allowed by the first law (energy conservation) but are never observed to happen. We never see ice cubes get larger as the drink they sit in gets hotter. Heat always flows from the warm drink into the ice cubes, bringing the two into thermal equilibrium. Energy could still be conserved if the drink got hotter while the cubes grew, so something else must prevent it. We can write this law by saying that heat **does not flow from a cold body to a warmer one**. Equivalently, we can state that heat is never
completely transformed into work. This law is a favorite target of crackpots, because it also means that a perpetual motion machine can’t exist. A perpetual motion machine is one which can run forever with no loss of energy. These machines are so certainly impossible that the patent office is supposed to discard all applications for patents for them.

We can describe the difference between allowed and disallowed processes by using a concept called entropy. This boils down to disorder. As entropy increases, disorder increases. A solid (a regular array of atoms) is more ordered than a liquid, which is more ordered than a gas. The change in a system’s entropy $\Delta S$ can be written as

$$\Delta S = \frac{Q}{T}$$

When heat is added to a system, $Q$ will be positive, so the change in entropy will be positive. When a system is composed of a hot object and a cool object, the two will come to equilibrium. One will have a positive value of $Q$, and one will have a negative $Q$. One will have a positive change in entropy while the other will have a negative change. In every situation, though, the total entropy will increase (as long as the system is isolated). If we look at all parts of a system, and include everything that could matter, entropy will always increase. The universe’s entropy increases during every natural process.

Entropy has something in common with internal energy – both are independent of the path taken. The change in entropy only depends on its initial and final values. We can make a plot of $T$ vs. $S$ if we want. The area under the curve showing the path taken by a system will give us the heat flow. If we return to our adiabatic process, which had $Q = 0$, we should have no change in entropy (and we can call the process isentropic). Including this possibility of isentropic processes, we can write the second law most simply as

**Heat Engines**

Devices which turn heat into work are known as heat engines. They may be complicated or very simple (a balloon can be a heat engine – if you add heat, it will expand against the pressure of the atmosphere, doing work). More commonly, heat engines operate cyclically so that we can get work out of them over long periods of time. This cycle involves moving around on a $p-V$ diagram in such a way that an area is enclosed. This area is just the work done. In the general case, when the lines which form the boundary of this area aren’t straight, we need calculus to figure out the work. If we look at the area enclosed by isobars & isomets, we just get a rectangle, and it’s easy to find the area of it.
Looking at the figure above, the closed cycle goes from 1-2-3-4-1. The line from 1 to 2 is going from a small volume to a larger volume at a fixed pressure. This indicates expansion, which means the system is doing work. The line from 2 to 3 shows a drop in pressure as volume remains constant, which means cooling is happening. From 3 to 4, volume is dropping as pressure stays constant, so the system is being compressed (work is being done on the system). This blows combustion products out of the cylinder in a car engine. From 4 to 1, pressure is rising while volume stays constant, so temperature is going up (corresponding to the spark plug firing in an engine).

We can use this approach to talk about an engine’s thermal efficiency. This tells you how much net work the engine does for a given heat input. This is written as

$$\varepsilon_{th} = \frac{W_{net}}{Q_{in}}$$

Since we’re returning the system to its original state, there’s no net change in its internal energy (meaning its temperature is the same at the end of the cycle as it was at the beginning of the cycle), so $\Delta U = 0$. This is equivalent to saying the heat flow out plus the heat flow in is the net work.

$$Q_{hot} - Q_{cold} = W_{net} \Rightarrow \varepsilon_{th} = \frac{Q_{hot} - Q_{cold}}{Q_{hot}} = 1 - \frac{Q_{cold}}{Q_{hot}}$$

(we’re using $-Q_{cold}$ to represent the flow out). This tells us that we could only have a 100% efficient engine if no heat were lost and all of the input heat were turned into work, which can’t happen by the second law. This gives yet another statement of the second law, which is no heat engine operating in a cycle can convert its heat input completely to work.

Thermal Pumps
A thermal pump is a device that moves heat from a cold place to a warmer one (the input of work is obviously required, or this would violate the 2nd law). In a heat pump or refrigerator, something with a low boiling point (like Freon, which boils at –30˚ C) is used to absorb heat from the hot place. The low boiling point means that the latent heat of vaporization can be used to give the material the ability to carry much more heat energy away from the hot side of the pump. The gas is drawn into a chamber by a descending piston and then compressed when the piston comes back up. When compressed, the vapor will be heated and then exhausted to a condenser. As the heat is radiated out into the air, the vapor condenses (in the condenser). When the liquid returns to the hot side, the pressure is lowered enough to cause it to boil, which draws a great deal of heat out of the hot side, bringing us back to the beginning of the cycle.

If the refrigerator is keeping the system at a constant temperature, $\Delta U$ is zero and we can write $Q_{\text{cold}} + W_{\text{in}} = Q_{\text{hot}}$. Efficiency for a refrigerator is given as its coefficient of performance, or $\text{COP}_{\text{ref}}$, defined as

$$\text{Typical values for the COP}_{\text{ref}} \text{ are between 3 – 5. A heat pump works similarly to a refrigerator, except that it can work in either direction. When the heat pump acts as an air conditioner, removing heat from the house and exhausting it to the outdoors, we can use the COP}_{\text{ref}} \text{ just as we did for a refrigerator. When it acts to heat the house, we instead write}$$

Values between 2 – 4 are common. It’s frequently said that electrical resistance heat (like the small space heaters which plug into the wall) is 100% efficient, which is true – all of the energy they consume eventually becomes heat. Heat pumps, though, don’t manufacture heat – they move it from one place to another. For that reason, heat pumps are more economical than resistance heat, which is really what you’re generally concerned about.

**Carnot Cycle and Ideal Heat Engines**

Because of the second law, there must always be some wasted heat in the operation of a heat engine. Sadi Carnot looked for the most efficient cycle which even an ideal heat engine could use. He found that the cycle’s path on a $p-V$ diagram consisted of the isothermal absorption of heat from a reservoir kept at constant $T_{\text{hot}}$, followed by an adiabatic expansion, the release of heat (again isothermally) to a cold reservoir kept at $T_{\text{cold}}$, followed by an adiabatic compression to return the system to its original state. This closed loop on the $p-V$ diagram is called the Carnot cycle. It’s just a rectangle of area $Q$ on a $T$-$S$ diagram, though. Since the shape is so simple, we know that the area $Q$ is also just $(T_{\text{hot}} - T_{\text{cold}}) \Delta S$ (because it’s a $T$-$S$ diagram – a rectangle on an x-y graph would just have an area of $\Delta x \Delta y$, so we shouldn’t be surprised. For the top side of the rectangle, $Q_{\text{hot}} = T_{\text{hot}} \Delta S$, but for the bottom side, $Q_{\text{cold}} = T_{\text{cold}} \Delta S$. Because the change in entropy ($\Delta S$) is the same for both, we can equate them by writing
So we could write the ideal **Carnot efficiency** of the heat engine as

\[
\varepsilon_C = 1 - \frac{T_{\text{cold}}}{T_{\text{hot}}} = \frac{T_{\text{hot}} - T_{\text{cold}}}{T_{\text{hot}}}
\]

This is the absolute maximum that any heat engine operating with these input and output temperatures can achieve. No real engine will ever be able to equal this efficiency. The formula does show that, as the difference between the two temperatures increases, the efficiency will also increase. Sometimes, efficiencies of real engines are compared to this theoretical maximum to find the **relative efficiency**, or

\[
\varepsilon_{\text{rel}} = \frac{\varepsilon_{\text{th}}}{\varepsilon_C}
\]

where the subscript “th” stands for the efficiency of the actual device, and the relative efficiency is generally multiplied by 100% to state it as a percentage.

Finally, this brings us to the **third law of thermodynamics**. If we wanted a 100% efficient heat engine, we’d need for \( T_{\text{cold}} \) to be 0 K, or absolute zero. The second law says that we can never have a 100% efficient heat engine, so these two facts together tell us that it is impossible to reach **absolute zero**, which is the third law. We can get arbitrarily close to it, but we can never actually reach it.

**Standing Waves & Resonance**

If we have a string attached to a wall at a fixed point (so that we get reflected waves) we can shake the string at certain frequencies and see a steady waveform which is called a **standing wave**. A standing wave will look sort of like a sine wave sitting on top of a negative of the same sine wave:
The notable feature of these waves is that the points marked with green dots don’t ever move. If this were a regular wave, the hills would be traveling down the rope and all points would move at different times. Here, the wall creates a wave moving back down the rope carrying the same energy as the wave moving towards the wall, and the result is a wave with stationary points (green dots) called nodes. The points in between the nodes (called antinodes) do the most moving as they go from the tallest hills to the lowest valleys. The distance between two nodes (or two antinodes) is one half of a wavelength, or $\lambda/2$. Various frequencies will cause this behavior – as the frequency increases, the number of hills & valleys that fit between you & the wall increases. These special frequencies are called resonant frequencies or natural frequencies.

These resonant frequencies are found by examining the length of the rope. If an integral number (whole number) of half-wavelengths can fit in the length of the string, the resonant pattern will appear. The connection between wavelength and string length is then

$$L = n \left( \frac{\lambda_n}{2} \right) \text{ where } n = 1, 2, 3, \ldots \text{ so } \lambda_n = \frac{2L}{n}$$

The relation between frequency, wave speed, and wavelength means that the resonant frequencies are then

$$f_n = \frac{v}{\lambda_n} = n \left( \frac{v}{2L} \right) = n f_1 \text{ where } n = 1, 2, 3, \ldots$$

The lowest frequency, where $n = 1$, is called the fundamental frequency, and all other frequencies are integer multiples of it (known as harmonics). The velocity of a wave on a string depends on its mass per unit length (because a km of dental floss may have the same mass as a meter of heavy rope – we need to know the mass of a given length) and the tension in the string. Since tension is measured in kg m/s², and mass per unit length is measured in kg/m, the form of the velocity is
\[ v = \sqrt{\frac{T}{\mu}} \]

meaning that the resonant frequencies are

\[ f_n = \frac{n}{2L} \sqrt{\frac{T}{\mu}} \text{ where } n = 1, 2, 3, \ldots \]

Resonance is also found when you push someone on a swing at the “right” time over & over again to send them very high. You can find the resonant frequency of a wine glass by wetting your finger & running it around the rim of the glass over & over again. In the past, the wind has hit at least one bridge at just the right frequency (resonance) to cause it to swing back & forth so violently that it collapsed.

**Sound**

Sound waves in air are essentially all longitudinal. The transverse components can’t propagate because gases (and liquids) don’t support shear (so when a mass of water is dragged past another mass of water, there’s essentially no interaction between the two masses). These changes in pressure are detected by the eardrum and the tiny bones in your inner ear. Humans can generally hear sounds which have frequencies from about 20 Hz to about 20,000 Hz. Whales, elephants, etc. can hear some of the sounds below 20 Hz (known as the **infrasound** region). Dogs and other animals can hear above 20 kHz (**ultrasonic** region). Bats can hear up to about 100 kHz, and they use this ability to find the insects they eat. They send out a very high-pitched noise and listen for reflections, which means they’re really using **sonar**. The frequency has to be very high, because the bat needs a small wavelength to detect small insects. Waves with a frequency of 30 kHz would have a wavelength of about a centimeter. The same technique is used by doctors to examine people without exposing them to X-ray radiation. There is an ongoing debate about the effects of small doses of radiation on people. We know that large doses can cause sickness or even death, but determining the effects of small doses is much more difficult. There are two general theories – one is that, below a certain threshold, radiation has no effect on people. You can think of this as working like a car wreck – in a collision with a brick wall at 80 mph, perhaps 99% of people involved would be killed. If we lower the speed, the crash becomes more survivable, but there is still a good chance for death even at 35 mph. However, if we drop the speed to 1 mph, there is really no damage done to anyone. If a million people were in collisions like this, we would not expect it to cause any deaths.

Another opinion says that every little bit of radiation does damage, and that, while a large dose might kill 99% of the people exposed, a very small dose may still kill 10 people out of 1,000,000 by causing a mutation that causes cancer. As uncertain as this debate is in people, it’s even cloudier when applied to a fetus, since the fetus’ cells are already dividing very rapidly and cells are thought to be more vulnerable during division. For this reason, no one wants to X-ray a pregnant woman. That’s why pictures of a fetus are always made using ultrasound, because there seems to be no way
for low levels of sound to cause damage (although they can cause heating or a skin burn if the technician is not careful).

**Speeds**

Sound moves faster in solids than in air, in general, because the molecules in a solid are connected to each other with stronger forces and therefore snap back into place more quickly when disturbed. The velocity depends on Young’s modulus for a solid, or the Bulk modulus for a liquid. Also, the density of the material is important, as that describes how close the molecules are to one another. The velocity can then be written:

\[
\begin{align*}
v_{\text{solid}} &= \sqrt{\frac{Y}{\rho}} \quad \text{or} \quad v_{\text{liquid}} = \sqrt{\frac{B}{\rho}}
\end{align*}
\]

In air, the speed of sound depends on the temperature (and other things we’ll ignore, like the humidity), and we get

\[
v_{\text{air}} = (331 + 0.6 T_C) \text{ m/ s}
\]

where the 0.6 must clearly have units of m/(°C s). This gives us a speed of about 343 m/s at room temperature in air. For comparison, sound moves at about 1500 m/s in water and over 5000 m/s in aluminum.

**Sound Intensity**

Most people are probably familiar with the **decibel** as a measure of the intensity of sound. The metric unit of intensity is actually power/area, or W/m². For many different phenomena (including light & gravity), we can find intensity by using a simple geometric argument. Imagine a point source of sound and a detector (microphone). Waves from the point source go out in all directions equally, forming spheres. Assume the microphone has a square face with an area of 1 cm². Look below at the power hitting the microphone at different points:
From the picture above, the point source is the black dot, the expanding sound wave is the tan circle, and the red triangle represents the amount of sound that actually strikes the microphone (blue). If we move the microphone out to a greater distance, all of the sound from the red triangle keeps going with the expanding sound wave (now yellow) and is fainter when it hits the microphone, because it has to cover a larger area. This two-dimensional drawing is misleading, because when we move the microphone twice as far away, the base of the triangle is only twice as large as it was. In reality, since these are spheres rather than circles and cones rather than triangles, the sound gets spread over \(2^2\) or four times the area. Sound (and light) are just like paint in this respect – if you have to spread it over four times the area, you can only spread it to \(\frac{1}{4}\) the thickness (or intensity). If we moved the microphone 3 times as far away, the sound would be spread over an area which is \(3^2\) or nine times as great. This reduces the intensity to \(\frac{1}{9}\) of what it was. We can guess now that the general form of the intensity formula should include a factor which shows how the area will change. We’ve said intensity is power divided by area, and we’ve said the sound expands in spheres, so we can find the intensity by dividing the total power by the area of a sphere which would stretch out to our measuring device:

\[
I(\text{at a dist. } r \text{ away from source}) = \frac{P}{A} = \frac{P}{4\pi r^2}
\]

This tells us that the intensity of sound falls off as the inverse square of the distance to the source. This general result will return when you examine light later. Basically, intensity goes as \(1/r^2\) because area increases as \(r^2\). The range of intensities humans can hear (and stand) is enormous. The threshold of hearing (quietest sound you can detect) has an intensity of about \(10^{-12}\) W/m\(^2\), while the loudest sound you can stand (threshold of pain) has an intensity of around 1 W/m\(^2\). Our ears don’t respond linearly to sound. In other words, an intensity of \(10^6\) W/m\(^2\) does not seem 10 times louder to us than an intensity of \(10^7\) W/m\(^2\). It would actually only seem about twice as loud. For this reason, it is common to use a logarithmic scale to describe sound intensity (if you study astronomy, you’ll see a similar logarithmic scale is used because the eye’s response to light is not linear, either. In astronomy, this is called the magnitude system).

Logarithmic scales are a way to compress a wide range of information into a smaller range (at the expense of detail). For example, if you want to make someone a map of how to get from your house to Miami, you will probably want to have a great deal of detail near your house so that they can find the streets that will take them to Highway 17. Once on 17, you will need to omit most of the details except the big ones, like the connection to I-95. Once they are on 95, you might draw the entire length of I-95 in Florida as just a straight line a few inches long. Your scale is a compromise – you want to give the person a very detailed map to get him/her out of town, but there is no way (or need) for you to give a very detailed map of the whole trip from here to Miami. The logarithmic scale is a similar compromise. A linear scale including the threshold of pain would have everyday sounds being too tiny to distinguish from silence.

This logarithmic scale is the source of the decibel level we’re familiar with. To calculate decibels, we first establish a reference level of sound (called \(I_0\)), which is the threshold of hearing or \(10^{-12}\) W/m\(^2\). Intensities of sound are divided by this intensity, and log of the result is then taken. For example, if we want to know the decibel level of a sound with an intensity of \(10^4\) W/m\(^2\), we would do this:
The $B$ is the abbreviation for bel. One tenth of a bel is a decibel, so this sound would be 80 decibels (pretty loud). This scale fits the range from near-silence to painfully loud into 12 bels or 120 decibels.

**Sound Phenomena**

As with other wave disturbances, we can have interference, reflection, refraction, and diffraction with sound waves. Reflection of sound can be found in echoes. It’s easy for you to experiment with this if you can find a large, solid wall or a set of concrete bleachers. If you clap your hands loudly, you can hear the claps coming back to you as they bounce off of different objects. Fireworks shows in cities are a great way to hear this happen if you can get to the roof of a building. Refraction in sound is less common, but it can happen if there is a layer of cool air underneath a layer of warm air. As sound moves out and encounters the warm air (where it speeds up), it will bend at the interface and can come back down to ground, enabling you to hear sounds from a greater distance than normal. Diffraction is heard every time you hear a sound from around a corner – the wall is not sufficient to block all of the sound, and the sound will bend around the corner.

Interference works the same way we mentioned before. At some points, two peaks will coincide and produce a much higher peak, while at other points, a peak and a trough will coincide and the resulting intensity will be greatly reduced. If there are two sources of sound (and a wall producing an echo can be one of the sources), the important thing to know is the phase of each wave at a point. If the two waves are in phase, we will get constructive interference and increased intensity. If the two waves are out of phase, we’ll get destructive interference and reduced intensity. For the case of two speakers sending out the same sound at the same time, all we need to know is the distance from our measuring device to speaker A and the distance from our device to speaker B. In fact, we really just need to know the difference in these distances. If the difference corresponds to a whole number of wavelengths, we will get constructive interference because the waves will be peaking at the same time (and the valleys will also occur at the same time). If the difference is equal to an odd number of half wavelengths, we get two waves 180° out of phase and we should get silence. It has to be an odd number of half wavelengths, because an even number of them would give a whole number of full wavelengths, which gives constructive interference. We can then write our condition for perfect constructive or destructive interference as

$$\Delta L = \frac{n \lambda}{2} \quad \text{where } n = 0, 2, 4, \ldots \text{(const.) or } n = 1, 3, 5, \ldots \text{(dest.)}$$

where $\Delta L$ is the path length difference (source 1 to measuring device minus source 2 to measuring device) and $\lambda$ is the wavelength. Ask yourself why you might not have noticed this effect with your home stereo, even if you listened to a pure note (like a recording of a tuning fork). Hint: why do you have two different outputs for your speakers? Are they playing exactly the same sounds?

Having two notes which are almost the same frequency causes the ear to hear beat frequencies. This is usually heard as a warble (kind of a wah-wah-wah noise) where the frequency of the beats
depends on the difference between the two frequencies heard. This beat frequency is therefore just \( f_b = |f_1 - f_2| \).

**Doppler Effect**

An effect familiar to everyone is the change in pitch of a police or ambulance siren as the siren moves towards or away from you. This is called the **Doppler effect**, and it is caused by the “bunching up” or spreading out of sound waves as the source moves towards or away from us. The same effect happens with light, but is not noticed in ordinary situations because the size of the effect depends on the speed of the source compared to the speed of sound or light. Since light is so much faster than sound, we don’t notice a color shift as a police car approaches us, but if we had incredibly accurate measuring devices, we could see it. We get a similar effect if we move and the source is stationary. The situations is not identical because there is a preferred reference frame provided by the air the sound travels through. This is not the case for light, which can travel without a medium, and therefore has only one formula for Doppler shift, where the **relative velocity** is the important thing. For sound in air, with a moving source and a stationary observer, we get the formula below for the frequency heard by the observer:

\[
f_{obs} = \left( \frac{v}{v \pm v_s} \right) f_s
\]

where \( f_s \) is the frequency sent by the source, \( v \) is the velocity of sound, and \( v_s \) is the velocity of the source where a “+” sign indicates that the source is moving away from the observer, and a “-” sign indicates a source moving towards the observer. For example, if an ambulance siren has a frequency of 1000 Hz and the ambulance is moving at 30 m/s towards you, you should hear a frequency of about \((343/(343-30))*1000 = 1096\) Hz. After it passes, you’ll hear a frequency of about \((343/(343+30))*1000 = 920\) Hz.

If you’re doing the moving, the formula is slightly different. The frequency you observe is given by

\[
f_{obs} = \left( \frac{v \pm v_{obs}}{v} \right) f_s
\]

where the “+” now represents motion toward the source and “-” means motion away from it. Now, if you’re moving towards the ambulance at 30 m/s, you’ll hear its 1000 Hz siren as \(((343+30)/343)*1000 = 1087\) Hz and \(((343-30)/343)*1000 = 913\) Hz as you move away.

If something is moving through the air at a speed greater than the speed of sound, we will hear a **sonic boom**. This is caused by a shock wave produced by constructive interference between sound waves piling up at the front of the object. The boom happens continually and trails the object as long as it moves faster than the speed of sound – there’s not just a single “boom” when the sound barrier is broken. This is the reason the Concorde doesn’t fly from NY to LA – the plane would cause a sonic boom to be heard on the ground all along its path, and these can be very loud, annoying, and destructive to things on the ground. For this reason, planes are generally forbidden from breaking the sound barrier while over the heavily-populated continental US. Speeds of supersonic planes are sometimes quoted in **Mach number** instead of meters per second. A plane
flying at Mach 2.5 is moving two and a half times faster than sound. Below, you can illustrations of the Doppler effect for a moving source.

In this picture, the black square is the source of sound. You can see from the circles that the square was at the origin when it emitted the red, outermost circle which has now spread out quite a bit. By the time it emitted the sound corresponding to the green circle, it had already moved to the right by a reasonable amount, so the green circle is centered around the point to the right which the square occupied at that time. Currently, the square has just emitted the sound represented by the purple circle, and you can see that it is very far from the origin. The blue observer will be hit by very many sound waves in a short time, which he will perceive as a high frequency sound, while the red observer will be receiving waves which are more spread out, and therefore sound like a lower frequency note. Supersonic speeds would have the circles appearing outside previous circles, as below:
The black lines represent a conical shock wave that is heard as a “boom”. The opening angle of the cone is found by using the formula

\[ \sin \theta = \frac{v_s}{v} \]
Wave Motion

Waves can generally be thought of as disturbances of a medium which carry energy. (Light and other forms of electromagnetic radiation are waves which don’t require a medium. We won’t talk about these waves until next semester). We’ll typically be looking at motions which repeat and are therefore periodic. The particles which make up the wave (water for water waves, air for sound waves, rock for earthquake waves, etc.) move back and forth in space as time goes on. That means the motion is periodic in both space and time. Periodic in time means that, if we keep watching the same spot, it will complete one full cycle of motion in one period. Periodic in space means that if we take a snapshot of the wave, we’ll see the same pattern repeated over and over after a fixed distance, known as a wavelength. If the wavelength is 10 cm, then each peak of the wave will be separated by 10 cm.

Other properties of wave motion include the velocity of the wave. If the frequency is the number of cycles completed in a second, and the wavelength is the length of a cycle, multiplying them together should give us the wave’s speed in meters per second:

\[ v = \lambda f \]

Also, a wave carries an energy which depends on the square of its amplitude \( A \).

Waves can be of two different types – transverse waves, and longitudinal waves. A longitudinal wave is a wave in which the oscillations are in the same direction as the wave’s motion. Imagine lying a slinky (or other spring) on its side and slapping one side of it. A compressional wave will travel down the slinky, and we’ll see it as the rings of the slinky being closer in one spot (which travels) than everywhere else. Sound waves in air are longitudinal, as they represent pressure differences. As a sound wave travels, it compresses the air very slightly and then, as it passes a point, it creates a region of lower pressure (kind of like a vacuum, only a very weak one). That’s why a speaker works by shaking a large cone back & forth to alternately compress & rarify the air.

A transverse wave is more like what you would see if you stretched a rope horizontally & tied it to something (or handed one end to someone like you were both going to swing a jump rope) and then moved your hand up & down rapidly. The rope itself moves up and down, but the wave will travel from you to the other end. If you’re washing a car and the hose gets stuck under a tire, or you’re cleaning house and the vacuum cleaner cord has gotten hung on something, you’ll probably try to
free it by sending a transverse wave down the hose/wire. Sending a longitudinal wave down the wire wouldn’t do you any good. You need to be able to move it perpendicularly to its length, and that requires a transverse wave. Some earthquake waves are transverse, and these waves are also called shear waves, because the wave is trying to shear the medium. As we’ve seen before, liquids and gases don’t support shear (you’ll never see a nail made out of water). They do, of course, support compression, so this is how sound travels in a liquid or a gas.

**Wave Properties**

Waves behave in ways vastly different from particles. One of the most basic ideas about waves is that they obey the **principle of superposition**. This says that the net result of several different waves passing through the same point at the same time is just the sum of the intensities they would have independently. Below, you can see that the red, green, and blue waves all have different frequencies, amplitudes, and phase shifts. At each point along the $x$ axis, their values are added to make the black wave. When the three waves are all near a peak, the total wave is very large. If one of the waves is near a peak and the other is near a valley, the total wave is much smaller.

The proof that many different waves add to form a total wave is found in your CD player. A person’s voice is typically composed of many different sine waves with frequencies and amplitudes that change as they speak. Even a song can have many different sine waves in it. Still, if you have a CD or cassette made at a party where dozens of people are talking, those hundreds and thousands of sine waves all add to produce one total. That total wave is what tells the magnet in the speaker to move back & forth in a particular way at any given instant. There is only one signal being sent to the speaker, and you can think of it as the height of the black wave above at each instant.

When waves hit a peak at the same place, we call that **constructive interference**, because they add to make a higher peak or deeper valley than they would have produced alone.
If the waves are out of phase and the peaks of one wave match valleys of another, the result is a flat spot and we call it **destructive interference**.

These pictures are examples of **total constructive/destructive interference**. The waves are the same amplitude and either exactly in phase or exactly out of phase. Waves are also **reflected** (at least partially) when they arrive at interfaces between two different media. If you have a wave moving through a light rope, one end of which is tied to a heavy rope, part of the wave will be reflected when it hits the heavy rope. When light waves hit water or glass or anything else, some of them get reflected (that’s how we see things). For waves in strings, we could have a 180° phase shift after reflection (meaning a hill traveling along the string becomes a valley after reflection) or not, depending on the attachment point of the string. If the string is tied to the wall, there will be a reflection because the hill will try to pull the wall “up” into its shape. The wall will resist and reverse the wave. If the string is tied to a movable piece that can slide up and down the wall, it will do just that and the wave will return in the same shape as it came.

Part of the wave which isn’t transmitted is **refracted**, or bent as it enters the new medium. This effect (which you’ll see more about in 102) is the reason why magnifying glasses bend sunlight down into a point. When a wave’s speed changes at the interface between two different media (which it usually does), its direction will also change unless it enters the new medium perpendicular to the interface. You can see this below as light travels from air to water (where it moves slower).
The angle of the light ray (always measured relative to the normal) is smaller in water than in air because the wave is bent towards the normal when it enters the water. You can see why it happens by looking at the example on the right, which shows the motion of a wheelchair going from cement (where it’s fast) to grass (where it slows down). The wheelchair will turn at the interface.

If wave speed depends on frequency, we see something called dispersion. This happens to light in glass or water – this is the basis for prisms and rainbows. When the different colors which make up white light pass through a dispersive medium, they are separated slightly. Diffraction is another wave effect which describes the way waves can bend around obstacles. This is the reason we can hear around corners. We can’t see around corners because the bending depends on the size of the obstacle compared to the wavelength of the wave. Sound waves are not very small when compared to walls (for example, since the speed of sound in air is about 330 m/s, and one of the typical frequencies making up a voice might be 1000 Hz, we get a wavelength of 0.33 m (about a foot)). Light waves are much, much smaller, typically having wavelengths of around 500 nanometers! Unless we are looking at holes or slits which are not very large compared to this size, we won’t see diffraction. You can actually do this with your fingers if you bring two of them very close together in front of a light source. You’ll start to see lines parallel to the fingers, and they are produced by diffraction.