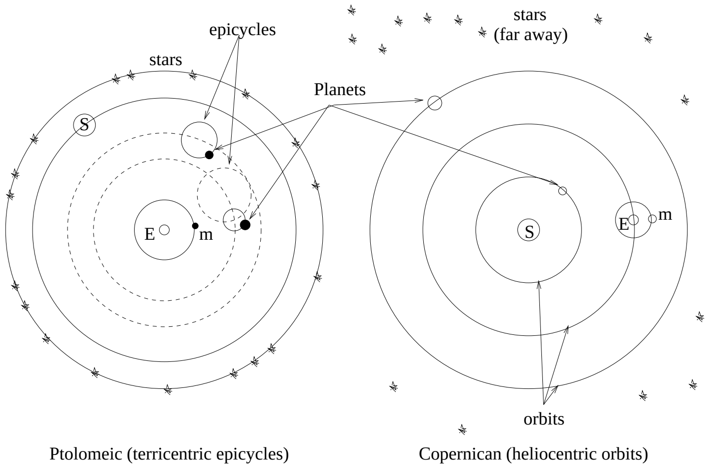
Physics 51 Review/Lecture Notes

Robert G. Brown
Duke University Physics Department
Durham, NC 27708-0305
rgb@phy.duke.edu

April 12, 2004

Contents

1 Gravity

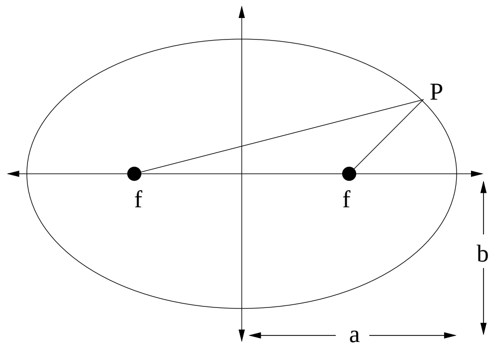


1.1 Cosmology and Kepler's Laws

- Early western cosmology earth-centered.
- Ptolemy (140 A.D.) "explained" planets (which failed this model) with "epicycles". Church embraced this model as consistent with Genesis.
- Copernicus (1543 A.D.) solar-centered model.
- Tycho Brahe accumulated data and Johannes Kepler fit that data to specific orbits and deduced laws:

Kepler's Laws

1. All planets move in elliptical orbits with the sun at one focus (see next section).
2. A line joining any planet to the sun sweeps out equal areas in equal times (dA/dt = constant).
3. The square of the period of any planet is proportional to the cube of the planet's mean distance from the sun ($T^2 = CR^3$). Note that the semimajor or semiminor axis of the ellipse will serve as well as the mean, with different contants of proportionality.



1.2 Ellipses and Conic Sections

- An ellipse is one of the conic sections (intersections of a right circular cone with a crossing plane). The others are hyperbolas and parabolas (circles are special cases of ellipses). All conic sections are actually possible orbits, not just ellipses.

- One possible equation for an ellipse is:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{1}$$

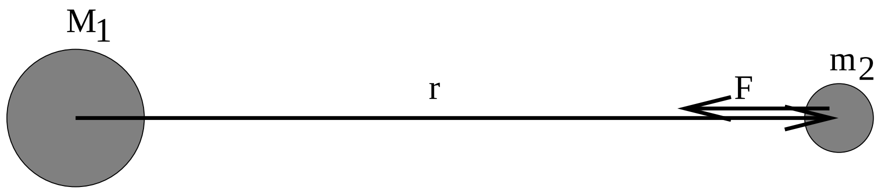
  The larger parameter (a above) is called the semimajor axis; the smaller (b) the semiminor; they lie on the similarly defined major and minor axes, where the foci of the ellipse lie on the major axis.

- Not all ellipses have major/minor axes that can be easily chosen to be x and y coordinates. Another general parameterization of an ellipse that is useful to us is a parametric cartesian representation:

$$x(t) = x_0 + a\cos(\omega t + \phi_x) \tag{2}$$
$$y(t) = y_0 + b\cos(\omega t + \phi_y) \tag{3}$$

  This equation will describe *any* ellipse centered on (x_0, y_0) by varying ωt from 0 to 2π. Adjusting the phase angles ϕ_x and ϕ_y and amplitudes a and b vary the orientation and eccentricity of the ellipse from a straight line at arbitrary angle to a circle.

- The foci of an ellipse are defined by the property that the sum of the distances from the foci to every point on an ellipse is a constant (so an ellipse can be drawn with a loop of string and two thumbtacks at the foci). If f is the distance of the foci from the origin, then the sum of the distances must be $2d = (f + a) + (a - f) = 2a$ (from the point $x = a$, $y = 0$. Also, $a^2 = f^2 + b^2$ (from the point $x = 0$, $y = b$). So $f = \sqrt{a^2 - b^2}$ where by convention $a \geq b$.



1.3 Newton's Law of Gravity

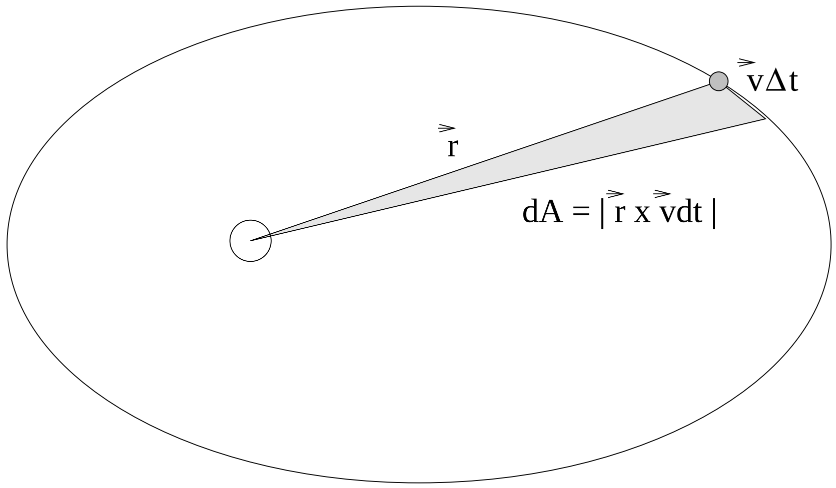
1. Force of gravity acts on a line joining centers of masses.
2. Force of gravity is attractive.
3. Force of gravity is proportional to each mass.
4. Force of gravity is **inversely proportional to the distance between the centers of the masses**.

or,

$$\vec{\mathbf{F}}_{12} = -\frac{Gm_1m_2}{r^2}\hat{\mathbf{r}} \tag{4}$$

where $G = 6.67x10^{-11}$ N-m^2/kg^2 is the **universal gravitational constant**.

Kepler's first law follow from **solving** Newton's laws and the equations of motion for this particular force law. This is a bit difficult and beyond the scope of this course, although we *will* show that circular orbits are one special solution that easily satisfy Kepler's Laws.



Kepler's Second Law is proven by observing that this force is radial, and hence exerts no torque. Thus the angular momentum of a planet is constant! That is,

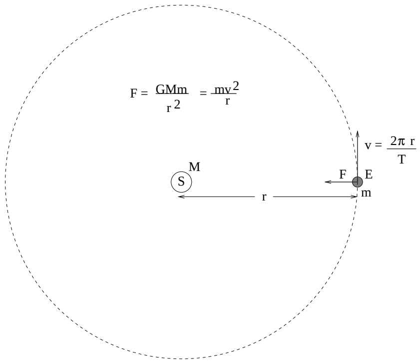
$$dA = \frac{1}{2}|\vec{\mathbf{r}} \times \vec{\mathbf{v}}dt| = \frac{1}{2}|\vec{\mathbf{r}}||\vec{\mathbf{v}}dt| \sin\theta \tag{5}$$

$$= \frac{1}{2m}|\vec{\mathbf{r}} \times m\vec{\mathbf{v}}dt| \tag{6}$$

or

$$\frac{dA}{dt} = \frac{1}{2m}|\vec{L}| = \text{constant} \tag{7}$$

(and Kepler's second law is proved for this force).



For a circular orbit, we can also prove Kepler's Third Law. The orbit is circular, so we have a relation between v and F_r.

$$\frac{GM_s m_p}{r^2} = m_p a_r = m_p \frac{v^2}{r} \tag{8}$$

so that

$$v^2 = \frac{GM_s}{r} \tag{9}$$

But, v is related to r and the period T by:

$$v = \frac{2\pi r}{T} \tag{10}$$

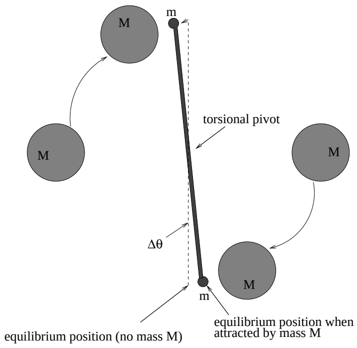
so that

$$v^2 = \frac{4\pi^2 r^2}{T^2} = \frac{GM_s}{r} \tag{11}$$

Finally,

$$r^3 = \frac{GM_s}{4\pi^2} T^2 \tag{12}$$

and Kepler's third law is proved for circular orbits (and the constant C evaluated for the solar system!).



1.4 The Gravitational Field

We define the gravitational field to be the **cause** of the gravitational force.

We define it conveniently to be the force per unit mass

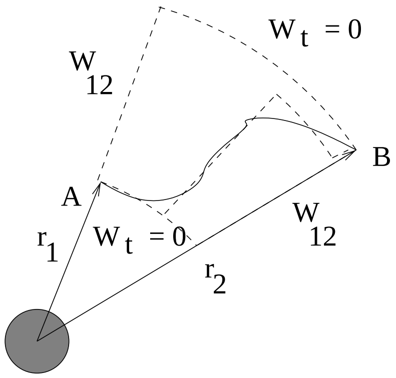
$$\vec{\mathbf{g}}(\vec{\mathbf{r}}) = -\frac{GM}{r^2}\hat{r} = \frac{\vec{\mathbf{F}}}{m} \tag{13}$$

The gravitational field at the surface of the earth is:

$$g(r) = \frac{F}{m} = \frac{GM_E}{R_E^2} \tag{14}$$

This equation can be used to find g, R_E, M_E, or G, from any of the other three, depending on which ones you think you know best. g is easy. R_E is actually also easy to measure independently and some classic methods were used to do so long before Columbus. M_E is hard! What about G?

Henry Cavendish made the first direct measurement of G using a torsional pendulum and some really massive balls. From this he was able to "weigh the earth" (find M_E). By measuring $\Delta\theta(r)$ (r measured between the centers) it was possible to directly measure G. He got 6.754 (vs 6.673 currently accepted) $\times 10^{-11}$ N-m^2/kg^2. Not bad!



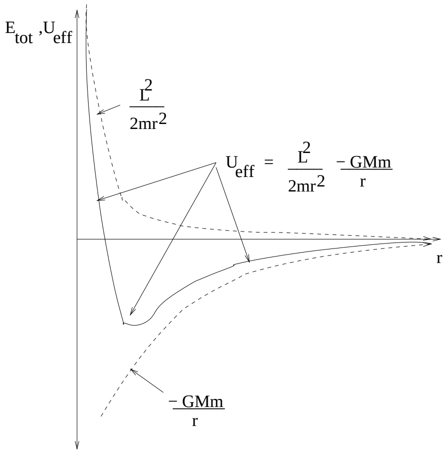
1.5 Gravitational Potential Energy

Gravitation is a conservative force, because the work done going "around" the attractor (perpendicular to the force) is zero, and the work done varying r is the same going out as in, so the work done is independent of path (see figure above). So:

$$\begin{aligned} U(r) &= -\int_{r_0}^{r} \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}} & (15) \\ &= -\int_{r_0}^{r} -\frac{GMm}{r^2} dr & (16) \\ &= -(\frac{GMm}{r} - \frac{GMm}{r_0}) & (17) \\ &= -\frac{GMm}{r} + \frac{GMm}{r_0} & (18) \end{aligned}$$

where r_0 is the radius of an arbitrary point where we define the potential energy to be zero. By convention, unless there is a good reason to choose otherwise, we require the zero of the potential energy to be at $r_0 = \infty$. Thus:

$$U(r) = -\frac{GMm}{r} \qquad (19)$$



1.6 Energy Diagrams and Orbits

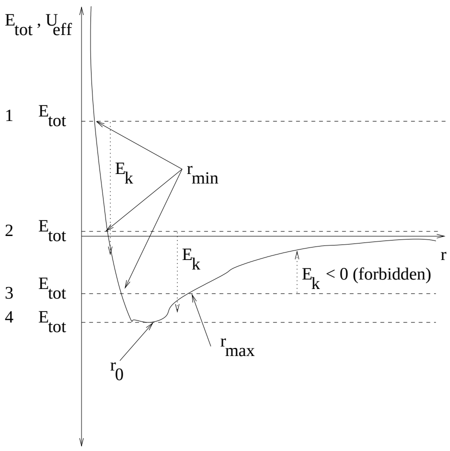
Let's write the total energy of a particle moving in a gravitational field in a clever way:

$$\begin{aligned}
E_{\rm tot} &= \frac{1}{2}mv^2 - \frac{GMm}{r} && (20)\\
&= \frac{1}{2}mv_r^2 + \frac{1}{2}mv_\perp^2 - \frac{GMm}{r} && (21)\\
&= \frac{1}{2}mv_r^2 + \frac{1}{2mr^2}(mv_\perp r)^2 - \frac{GMm}{r} && (22)\\
&= \frac{1}{2}mv_r^2 + \frac{L^2}{2mr^2} - \frac{GMm}{r} && (23)\\
&= \frac{1}{2}mv_r^2 + U_{\rm eff}(r) && (24)
\end{aligned}$$

Where

$$U_{\rm eff}(r) = \frac{L^2}{2mr^2} - \frac{GMm}{r} \qquad (25)$$

is the (radial) potential energy plus the transverse kinetic energy (related to the constant angular momentum of the particle). If we plot the effective potential (and its pieces) we get a one-dimensional radial energy plot.



By drawing a constant total energy on this plot, the difference between $E_{\rm tot}$ and $U_{\rm eff}(r)$ is the radial kinetic energy, which must be positive. We can determine lots of interesting things from this diagram.

In this figure, we show orbits with a given angular momentum $\vec{\mathbf{L}} \neq 0$ and four generic total energies $E_{\rm tot}$. These orbits have the following characteristics and names:

1. $E_{\rm tot} > 0$. This is a **hyperbolic** orbit.

2. $E_{\rm tot} = 0$. This is a **parabolic** orbit. This orbit defines **escape velocity** as we shall see later.

3. $E_{\rm tot} < 0$. This is generally an **elliptical** orbit (consistent with Kepler's First Law).

4. $E_{\rm tot} = U_{\rm eff,min}$. This is a circular orbit. This is a special case of an elliptical orbit, but deserves special mention.

Note well that all of the orbits are **conic sections**. This interesting geometric connection between $1/r^2$ forces and conic orbits was a tremendous motivation for important mathematical work two or three hundred years ago.

1.7 Escape Velocity, Escape Energy

As we noted in the previous section, a particle has "escape energy" if and only if its total energy is greater than or equal to zero. We define the **escape velocity** (a misnomer!) of the particle as the minimum **speed** (!) that it must have to escape from its current gravitational field – typically that of a moon, or planet, or sun. Thus:

$$E_{\rm tot} = 0 = \frac{1}{2} m v_{\rm escape}^2 - \frac{GMm}{r} \tag{26}$$

so that

$$v_{\rm escape} = \sqrt{\frac{2GM}{r}} = \sqrt{2gr} \tag{27}$$

where in the last form $g = \frac{GM}{r^2}$ (the magnitude of the gravitational field – see next item). For earth:

$$v_{\rm escape} = \sqrt{\frac{2GM_E}{R_E}} = \sqrt{2gR_E} = 11.2 \quad \text{km/sec} \tag{28}$$

(Note: Recall the form derived by equating Newton's Law of Gravitation and mv^2/r in an earlier section for the velocity of a mass m in a circular orbit around a larger mass M:

$$v_{\rm circ}^2 = \frac{GM}{r} \tag{29}$$

from which we see that $v_{\rm escape} = \sqrt{2} v_{\rm circ}$.)

It is often interesting to contemplate this reasoning in reverse. If we drop a rock onto the earth from a state of rest "far away" (much farther than the radius of the earth, far enough away to be considered "infinity"), it will REACH the earth with escape (kinetic) energy. Since the earth is likely to be much larger than the rock, it will undergo an *inelastic* collision and release nearly *all* its kinetic energy as heat. If the rock is small, this is not a problem. If it is large (say, 1 km and up) it releases a *lot* of energy.

$$M = \frac{4\pi\rho}{3} r^3 \tag{30}$$

is a reasonable equation for the mass of a spherical rock. ρ can be estimated at 10^4 kg/m^3, so for $r \approx 1000$ meters, this is roughly $M \approx 4 \times 10^{13}$ kg, or around 10 billion metric tons of rock, about the mass of a small mountain.

This mass will land on earth with *escape velocity*, 11.2 km/sec, if it falls in from far away. Or more, of course – it may have started with velocity and energy from some other source – this is pretty much a minimum. As an exercise, compute the number of Joules this collision would release to toast the dinosaurs – or us!

2 Oscillations

2.1 Oscillations

Oscillations occur whenever a force exists that pushes an object back towards a stable equilibrium position whenever it is displaced from it.

Such forces abound in nature – things are held together in structured form *because* they are in stable equilibrium positions and when they are disturbed in certain ways, they oscillate.

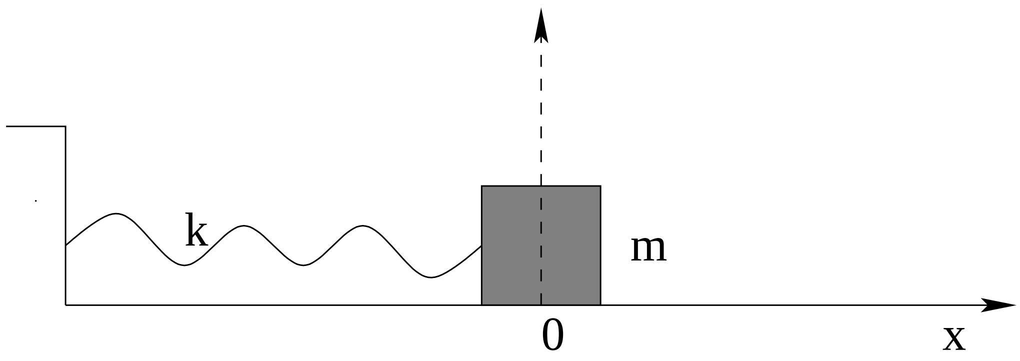
When the displacement from equilibrium is *small*, the restoring force is often *linearly* related to the displacement, at least to a good approximation. In that case the oscillations take on a special character – they are called *harmonic* oscillations as they are described by harmonic functions (sines and cosines) known from trigonometery.

In this course we will study **simple harmonic oscillators**, both with and without damping forces. The principle examples we will study will be masses on springs and various penduli.

Springs obey Hooke's Law: $\vec{F} = -k\vec{x}$ (where k is called the *spring constant*. A perfect spring produces perfect harmonic oscillation, so this will be our archetype.

A pendulum (as we shall see) has a restoring force or torque proportional to displacement for *small* displacements but is much too complicated to treat in this course for large displacements. It is a simple example of a problem that oscillates harmonically for small displacements but *not* harmonically for large ones.

An oscillator can be **damped** by dissipative forces such as friction and viscous drag. A damped oscillator can have exhibit a variety of behaviors depending on the relative strength and form of the damping force, but for one special form it can be easily described.



2.2 Springs

We will work in one dimension (call it x) and will for the time being place the spring equilibrium at the origin. Its equation of motion is thus:

$$F = -kx = ma = m\frac{d^2x}{dt^2} \tag{31}$$

Rearranging:

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0 \tag{32}$$

$$\frac{d^2x}{dt^2} + \omega^2 x = 0 \tag{33}$$

where $\omega^2 = k/m$ must have units of inverse time squared (why?).

This latter form is the standard *harmonic oscillator equation* (of motion). If we solve it *once and for all* now, whenever we can put an equation of motion into this form in the future we can just read off the solution by identifying similar quantities.

To solve it, we note that it basically tells us that $x(t)$ must be a function that has a second derivative proportional to the function itself. We know at least three functions whose second derivatives are proportional to themselves: cosine, sine and exponential. To learn something very important about the relationship between these functions, we'll assume the *exponential* form:

$$x(t) = x_0 e^{\alpha t} \tag{34}$$

(where α is an unknown parameter). Substituting this into the differential equation and differentiating, we get:

$$\begin{aligned} \frac{d^2 x_0 e^{\alpha t}}{dt^2} + \omega^2 x_0 e^{\alpha t} &= 0 && (35) \\ \left(\alpha^2 + \omega^2\right) x_0 e^{\alpha t} &= 0 && (36) \\ \left(\alpha^2 + \omega^2\right) &= 0 && (37) \end{aligned}$$

where the last relation is called the *characteristic* equation for the differential equation. If we can find an α such that this equation is satisfied, then our assumed answer will indeed solve the D.E.

Clearly:

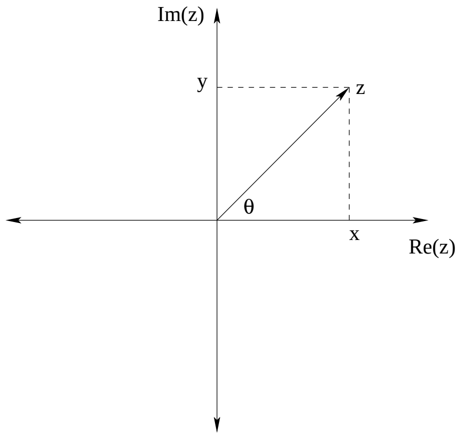
$$\alpha = \pm i\omega \tag{38}$$

and we get two solutions! We will always get n independent solutions for an nth order differential equation, so this is good:

$$\begin{aligned} x_+(t) &= x_{0+} e^{+i\omega t} && (39) \\ x_-(t) &= x_{0-} e^{-i\omega t} && (40) \end{aligned}$$

and an arbitrary solution can be made up of a sum of these terms:

$$x(t) = x_{0+} e^{+i\omega t} + x_{0-} e^{-i\omega t} \tag{41}$$



2.3 Math: Complex Numbers and Harmonic Trigonometric Functions

Some *extremely useful and important* True Facts:

2.3.1 Complex Numbers

This is a very terse review of their most important properties. An arbitrary complex number z can be written as:

$$\begin{aligned} z &= x + iy && (42) \\ &= |z|\cos(\theta) + i|z|\sin(\theta) && (43) \\ &= |z|e^{i\theta} && (44) \end{aligned}$$

where $x = |z|\cos(\theta)$, $y = |z|\sin(\theta)$, and $|z| = \sqrt{x^2 + y^2}$. *All complex numbers* can be written as a real amplitude $|z|$ times a complex exponential form involving a phase angle. Again, it is difficult to convey how incredibly useful this result is without further study, but I commend it to your attention.

2.3.2 Relations between cosine, sine and exponential functions

$$\begin{aligned} e^{\pm i\theta} &= \cos(\theta) \pm i\sin(\theta) && (45) \\ \cos(\theta) &= \frac{1}{2}\left(e^{+i\theta} + e^{-i\theta}\right) && (46) \\ \sin(\theta) &= \frac{1}{2i}\left(e^{+i\theta} - e^{-i\theta}\right) && (47) \end{aligned}$$

From these relations and the properties of exponential multiplication you can painlessly prove all sorts of trigonometric identities that were immensely painful to prove back in high school

2.3.3 Power Series Expansions

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \tag{48}$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots \tag{49}$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots \tag{50}$$

Depending on where you start, these can be used to prove the relations above. They are most useful for getting expansions for small values of their parameters. For small x (to leading order):

$$e^x \approx 1 + x \tag{51}$$

$$\cos(x) \approx 1 - \frac{x^2}{2!} \tag{52}$$

$$\sin(x) \approx x \tag{53}$$

$$\tan(x) \approx x \tag{54}$$

We will use these fairly often in this course, so learn them.

2.3.4 An Important Relation

A relation I will state without proof that is very important to this course is that the real part of the $x(t)$ derived above:

$$\Re(x(t)) = \Re(x_{0+}e^{+i\omega t} + x_{0-}e^{-i\omega t}) \tag{55}$$

$$= X_0 \cos(\omega t + \phi) \tag{56}$$

where ϕ is an arbitrary phase. You can prove this in a few minutes or relaxing, enjoyable algebra from the relations outlined above – remember that x_{0+} and x_{0-} are arbitrary *complex* numbers and so can be written in complex exponential form!

2.4 Simple Harmonic Oscillation

2.4.1 Solution

We generally are interested in *real* part of $x(t)$ when studying oscillating masses, so we'll stick to the following solution:

$$x(t) = X_0 \cos(\omega t + \phi) \tag{57}$$

where X_0 is called the *amplitude* of the oscillation and ϕ is called the *phase* of the oscillation. The amplitude tells you how big the oscillation is, the phase tells you when the oscillator was started relative to your clock (the one that reads t). Note that we could have used $\sin(\omega t + \phi)$ as well, or any of several other forms, since $\cos(\theta) = \sin(\theta + \pi/2)$. But you knew that.

X_0 and ϕ are two unknowns and have to be determined from the initial conditions, the givens of the problem. They are basically constants of integration *just like* x_0 and v_0 for the one-dimensional constant acceleration problem. From this we can easily see that:

$$v(t) = \frac{dx}{dt} = -\omega X_0 \sin(\omega t + \phi) \tag{58}$$

and

$$a(t) = \frac{d^2x}{dt^2} = -\omega^2 X_0 \cos(\omega t + \phi) = -\frac{k}{m} x(t) \tag{59}$$

(where the last relation proves the original differential equation).

2.4.2 Relations Involving ω

We remarked above that *omega* had to have units of t^{-1}. The following are some True Facts involving ω that You Should Know:

$$\begin{aligned} \omega &= \frac{2\pi}{T} && (60) \\ &= 2\pi f && (61) \end{aligned}$$

where T is the *period* of the oscillator the time required for it to return to an identical position *and* velocity) and f is called the *frequency* of the oscillator. Know these relations *instantly.* They are easy to figure out but will cost you valuable time on a quiz or exam if you don't just take the time to completely embrace them now.

Note a very interesting thing. If we build a perfect simple harmonic oscillator, it oscillates at the same frequency *independent of its amplitude*. If we know the period and can count, we have just invented the *clock*. In fact, clocks are nearly *always* made out of various oscillators (why?); some of the earliest clocks were made using a pendulum as an oscillator and mechanical gears to count the oscillations, although now we use the much more precise oscillations of a bit of stressed crystalline quartz (for example) and electronic counters. The idea, however, remains the same.

2.4.3 Energy

The spring is a *conservative force*. Thus:

$$U = -W(0 \to x) = -\int_0^x (-kx)dx = \frac{1}{2}kx^2 \tag{62}$$

$$= \frac{1}{2}kX_0^2 \cos^2(\omega t + \phi) \tag{63}$$

where we have arbitrarily set the zero of potential energy to be the equilibrium position (what would it look like if the zero were at x_0?).

The kinetic energy is:

$$K = \frac{1}{2}mv^2 \tag{64}$$

$$= \frac{1}{2}m(\omega^2)X_0^2 \sin^2(\omega t + \phi) \tag{65}$$

$$= \frac{1}{2}m(\frac{k}{m})X_0^2 \sin^2(\omega t + \phi) \tag{66}$$

$$= \frac{1}{2}kX_0^2 \sin^2(\omega t + \phi) \tag{67}$$

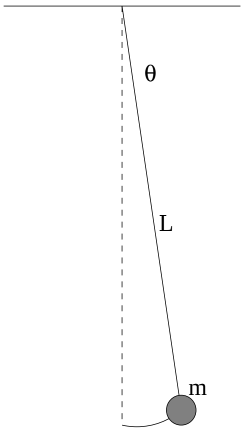
The total energy is thus:

$$E = \frac{1}{2}kX_0^2 \sin^2(\omega t + \phi) + \frac{1}{2}kX_0^2 \cos^2(\omega t + \phi) \tag{68}$$

$$= \frac{1}{2}kX_0^2 \tag{69}$$

and is *constant* in time! Kinda spooky how that works out...

Note that the energy oscillates between being all potential at the extreme ends of the swing (where the object comes to rest) and all kinetic at the equilibrium position (where the object experiences no force).



2.5 The Pendulum

The pendulum is another example of a simple harmonic oscillator, at least for small oscillations. Suppose we have a mass m attached to a string of length ℓ. We swing it up so that the stretched string makes a (small) angle θ_0 with the vertical and release it. What happens?

We write Newton's Second Law for the force component *tangent* to the arc of the circle of the swing as:

$$F_t = -mg\sin(\theta) = ma_t = m\ell\frac{d^2\theta}{dt^2} \tag{70}$$

where the latter follows from $a_t = \ell\alpha$ (the angular acceleration). Then we rearrange to get:

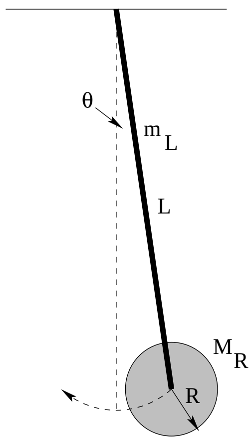
$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell}\sin(\theta) = 0 \tag{71}$$

This is *almost* a simple harmonic equation with $\omega^2 = \frac{g}{\ell}$. To make it one, we have to use the small angle approximation $\sin(\theta) \approx \theta$. Then

$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell}\theta = 0 \tag{72}$$

and we can just *read off the solution*:

$$\theta(t) = \theta_0\cos(\omega t + \phi) \tag{73}$$



If you compute the gravitational potential energy for the pendulum for *arbitrary* angle θ, you get:

$$U(\theta) = mg\ell\,(1 - \cos(\theta)) \tag{74}$$

Somehow, this doesn't look like the form we might expect from blindly substituting into our solution for the SHO above:

$$U(t) = \frac{1}{2} mg\ell\theta_0^2 \sin^2(\omega t + \phi) \tag{75}$$

As an interesting and fun exercise (that really isn't too difficult) see if you can prove that these two forms are really the same, *if* you make the small angle approximation for θ in the first form! This shows you pretty much where the approximation will break down as θ_0 is gradually increased. For large enough θ, the period of a pendulum clock *does* depend on the amplitude of the swing. This might explain grandfather clocks – clocks with very long penduli that can swing very slowly through very small angles – and why they were so accurate for their day.

2.5.1 The Physical Pendulum

In the treatment of the ordinary pendulum above, we just used Newton's Second Law directly to get the equation of motion. This was possible only because we could neglect the mass of the string and because we could treat the mass like a point mass at its end.

However, *real* grandfather clocks often have a large, massive pendulum like the one above – a long massive rod (of length L and mass m_L) with a large round disk (of radius R and mass M_R) at the end. The round weight rotates through an angle of $2\theta_0$ in each oscillation, so it has angular momemtum. Newton's Law for forces no longer suffices. We must use torque and the moment of inertia to obtain the frequency of the oscillator.

To do this we go through the *same steps* (more or less) that we did for the regular pendulum. First we compute the net gravitational torque on the system at an arbitrary (small) angle θ:

$$\tau = -\left(\frac{L}{2}m_L g + LM_R g\right)\sin(\theta) \tag{76}$$

(The - sign is there because the torque *opposes* the angular displacement from equilibrium.)

Next we set this equal to $I\alpha$, where I is the total moment of inertia for the *system* about the pivot of the pendulum and simplify:

$$\tau = -\left(\frac{L}{2}m_L g + LM_R g\right)\sin(\theta) = I\alpha = I\frac{d^2\theta}{dt^2} \tag{77}$$

$$I\frac{d^2\theta}{dt^2} + \left(\frac{L}{2}m_L g + LM_R g\right)\sin(\theta) = 0 \tag{78}$$

and make the small angle approximation to get:

$$\frac{d^2\theta}{dt^2} + \frac{\left(\frac{L}{2}m_L g + LM_R g\right)}{I}\,\theta = 0 \tag{79}$$

Note that for this problem:

$$I = \frac{1}{12}m_L L^2 + \frac{1}{2}M_R R^2 + M_R L^2 \tag{80}$$

(the moment of inertia of the rod plus the moment of inertial of the disk rotating about a parallel axis a distance L away from its center of mass). From this we can read off the angular frequency:

$$\omega^2 = \frac{4\pi^2}{T} = \frac{\left(\frac{L}{2}m_L g + LM_R g\right)}{I} \tag{81}$$

With ω in hand, we know everything. For example:

$$\theta(t) = \theta_0\cos(\omega t + \phi) \tag{82}$$

gives us the angular trajectory. We can easily solve for the period T, the frequency $f = 1/T$, the spatial or angular velocity, or whatever we like.

Note that the energy of this sort of pendulum can be tricky. Obviously its potential energy is easy enough – it depends on the elevation of the center of masses of the rod and the disk. The kinetic energy, however, is:

$$K = \frac{1}{2} I \left(\frac{d^2\theta}{dt^2} \right)^2 \tag{83}$$

where I do *not* write $\frac{1}{2} I \omega^2$ as usual because it confuses ω (the angular frequency of the oscillator, roughly contant) and $\omega(t)$ (the angular velocity of the pendulum bob, which varies between 0 and some maximum value in every cycle).

2.6 Damped Oscillation

So far, all the oscillators we've treated are *ideal.* There is no friction or damping. In the real world, of course, things *always* damp down. You have to keep pushing the kid on the swing or they slowly come to rest. Your car doesn't *keep* bouncing after going through a pothole in the road. Buildings and bridges, clocks and kids, real oscillators all have damping.

Damping forces can be very complicated. There is kinetic friction, which tends to be independent of speed. There are various fluid drag forces, which tend to depend on speed, but in a sometimes complicated way. There may be other forces that we haven't studied yet that contribute to damping. So in order to get beyond a very qualitative description of damping, we're going to have to specify a *form* for the damping force (ideally one we can work with, i.e. integrate).

We'll pick the simplest possible one:

$$F_d = -bv \tag{84}$$

(b is called the *damping constant* or *damping coefficient*) which is typical of an object being damped by a fluid at relatively low speeds. With this form we can get an exact solution to the differential equation easily (good), get a preview of a solution we'll need next semester to study LRC circuits (better), and get a very nice qualitative picture of damping besides (best).

We proceed with Newton's second law for a mass m on a spring with spring constant k and a damping force $-bv$:

$$F = -kx - bv = ma = m\frac{d^2x}{dt^2} \tag{85}$$

Again, simple manipulation leads to:

$$\frac{d^2x}{dt^2} + \frac{b}{m}\frac{dx}{dt} + \frac{k}{m}x = 0 \tag{86}$$

which is standard form.

Again, it looks like a function that is proportional to its own *first* derivative is called for (and in this case this excludes sine and cosine as possibilities). We guess $x(t) = x_0 e^{\alpha t}$ as before, substitute, cancel out the common $x(t) \neq 0$ and get the characteristic:

$$\alpha^2 + \frac{b}{m}\alpha + \frac{k}{m} = 0 \tag{87}$$

This is pretty easy, actually – each derivative just brings down an α.

To solve for α we have to use the dread *quadratic formula*:

$$\alpha = \frac{\frac{-b}{m} \pm \sqrt{\frac{b^2}{m^2} - \frac{4k}{m}}}{2} \tag{88}$$

This isn't quite where we want it. We simplify the first term, factor a $-4k/m$ out from under the radical (where it becomes $i\omega_0$, where $\omega_0 = \sqrt{k/m}$ is the frequency of the *undamped* oscillator with the same mass and spring constant) and get:

$$\alpha = \frac{-b}{2m} \pm i\omega_0\sqrt{1 - \frac{b^2}{4km}} \tag{89}$$

Again, there are two solutions, for example:

$$x_\pm(t) = X_{0\pm} e^{\frac{-b}{2m}t} e^{\pm i\omega' t} \tag{90}$$

where

$$\omega' = \omega_0\sqrt{1 - \frac{b^2}{4km}} \tag{91}$$

Again, we can take the real part of their sum and get:

$$x_\pm(t) = X_0 e^{\frac{-b}{2m}t}\cos(\omega' t + \phi) \tag{92}$$

where X_0 is the real initial amplitude and *phi* determines the relative phase of the oscillator.

2.7 Properties of the Damped Oscillator

There are several properties of the damped oscillator that are important to know.

- The amplitude damps *exponentially* as time advances. After a certain amount of time, the amplitude is halved. After the *same* amount of time, it is halved again.
- The frequency is shifted.
- The oscillator can be (under)damped, critically damped, or overdamped.

The oscillator is *underdamped* if ω' is real, which will be true if $4km > b^2$. It will undergo true oscillations, eventually approaching zero amplitude due to damping.

The oscillator is *critically* dampled if ω' is zero, when $4km = b^2$. The oscillator will *not oscillate* – it will go to zero *exponentially* in the shortest possible time.

The oscillator is *overdamped* if ω' is imaginary, which will be true if $4km < b^2$. In this case α is entirely real and has a component that damps very slowly. The amplitude goes to zero *exponentially* as before, but over a longer (possibly much longer) time and does not oscillate through zero at all.

A car's shock absorbers should be barely underdamped. If the car "bounces" once and then damps to zero when you push down on a fender and suddenly release it, the shocks are good. If it bounces three of four time the shocks are too underdamped and dangerous as you could lose control after a big bump. If it doesn't bounce up and back down at all at all and instead slowly oozes back up to level from below, it is overdamped and dangerous, as a succession of sharp bumps could leave your shocks still compressed and unable to absorb the impack.

Tall buildings also have dampers to keep them from swaying in a strong wind. Houses are build with lots of dampers in them to keep them quiet. Fully understanding damped (and eventually driven) oscillation is essential to many sciences as well as both mechanical and electrical engineering.

3 Waves

3.1 Wave Summary

- **Wave Equation**

$$\frac{d^2y}{dt^2} - v^2\frac{d^2y}{dx^2} = 0 \tag{93}$$

where for waves on a string:

$$v = \pm\sqrt{\frac{T}{\mu}} \tag{94}$$

- **Superposition Principle**

$$y(x,t) = Ay_1(x,t) + By_2(x,t) \tag{95}$$

(sum of solutions is solution). Leads to interference, standing waves.

- **Travelling Wave Pulse**

$$y(x,t) = f(x \pm vt) \tag{96}$$

where $f(u)$ is an arbitrary functional shape or pulse

- **Harmonic Travelling Waves**

$$y(x,t) = y_0 \sin(kx \pm \omega t). \tag{97}$$

where frequency f, wavelength λ, wave number $k = 2\pi/\lambda$ and angular frequency ω are related to v by:

$$v = f\lambda = \frac{\omega}{k} \tag{98}$$

- **Stationary Harmonic Waves**

$$y(x,t) = y_0 \cos(kx + \delta)\cos(\omega t + \phi) \tag{99}$$

where one can select k and ω so that waves on a string of length L satisfy fixed or free boundary conditions.

- **Energy (of wave on string)**

$$E_{\rm tot} = \frac{1}{2}\mu\omega^2 A^2 \lambda \tag{100}$$

is the total energy in a wavelength of a travelling harmonic wave. The wave transports the power

$$P = \frac{E}{T} = \frac{1}{2}\mu\omega^2 A^2 \lambda f = \frac{1}{2}\mu\omega^2 A^2 v \tag{101}$$

past any point on the string.

- **Reflection/Transmission of Waves**

  1. Light string (medium) to heavy string (medium): Transmitted pulse right side up, reflected pulse inverted. (A fixed string boundary is the limit of attaching to an "infinitely heavy string").
  2. Heavy string to light string: Transmitted pulse right side up, reflected pulse right side up. (a free string boundary is the limit of attaching to a "massless string").

3.2 Waves

We have seen how a particle on a spring that creates a restoring force proportional to its displacement from an equilibrium position oscillates harmonically in time about that equilibrium. What happens if there are *many* particles, *all* connected by tiny "springs" to one another in an extended way? This is a good metaphor for many, many physical systems. Particles in a solid, a liquid, or a gas both attract and repel one another with forces that maintain an average particle spacing. Extended objects under tension or pressure such as strings have components that can exert forces on one another. Even fields (as we shall learn next semester) can interact so that changes in one tiny element of space create changes in a neighboring element of space.

What we observe in all of these cases is that changes in any part of the medium "propagate" to other parts of the medium in a very systematic way. The motion observed in this propagation is called a *wave*. We have all observed waves in our daily lives in many contexts. We have watched water waves propagate away from boats and raindrops. We listen to sound waves (music) generated by waves created on stretched strings or from tubes driven by air and transmitted invisibly through space by means of radio waves. We read these words by means of light, an electromagnetic wave. In advanced physics classes one learns that all matter is a sort of quantum wave, that indeed *everything* is really a manifestation of waves.

It therefore seems sensible to make a first pass at understanding waves and how they work in general, so that we can learn and understand more in future classes that go into detail.

The concept of a wave is simple – it is an *extended structure* that oscillates in both *space* and in *time*. We will study two kinds of waves at this point in the course:

- **Transverse Waves** (e.g. waves on a string). The displacement of particles in a transverse wave is *perpendicular* to the direction of the wave itself.

- **Longitudinal Waves** (e.g. sound waves). The displacement of particles in a longitudinal wave is in the same direction that the wave propagates in.

Some waves, for example water waves, are simultaneously longitudinal and transverse. Transverse waves are probably the most important waves to understand for the future; light is a transverse wave. We will therefore start by studying transverse waves in a simple context: waves on a stretched string.

3.3 Waves on a String

Suppose we have a uniform string (such as a guitar string) that is stretched so that it is under tension T. The string is characterized by its mass per unit length μ – thick guitar strings have more mass per unit length than thin ones. It is fairly harmless at this point to imagine that the string is fixed to pegs at the ends that maintain the tension.

Now imagine that we have plucked the string somewhere between the end points so that it is displaced in the y-direction from its equilibrium (straight) stretched position and has some curved shape. If we examine a short segment of the string of length Δx, we can write Newton's 2nd law for that segment. If $\theta is small$, in the x-direction the components of the tension nearly perfectly cancel. Each bit of string therefore moves more or less straight up and down, and a useful solution is described by $y(x, t)$, the y **displacement** of the string at position x and time t. The permitted solutions must be *continuous* if the string does not break.

In the y-direction, we find write the force law by considering the *difference* between the y-components at the ends:

$$F_y = T\sin(\theta_2) - T\sin(\theta_1) = \mu\Delta x a_y \tag{102}$$

We make the small angle approximation: $\sin(\theta) \approx \tan(\theta) \approx \theta$, divide out the $\mu\Delta x$, and note that $\tan(\theta) = \frac{dy}{dx}$ (the slope of the string is the tangent of the angle the string makes with the horizon). Then:

$$\frac{d^2y}{dt^2} = \frac{T}{\mu}\frac{\Delta(\frac{dy}{dx})}{\Delta x} \tag{103}$$

In the limit that $\Delta x \to 0$, this becomes:

$$\frac{d^2y}{dt^2} - \frac{T}{\mu}\frac{d^2y}{dx^2} = 0 \tag{104}$$

The quantity $\frac{T}{\mu}$ has to have units of $\frac{L^2}{t^2}$ which is a velocity squared.

We therefore formulate this as the one dimensional *wave equation*:

$$\frac{d^2y}{dt^2} - v^2\frac{d^2y}{dx^2} = 0 \tag{105}$$

where

$$v = \pm\sqrt{\frac{T}{\mu}} \tag{106}$$

are the velocity(s) of the wave on the string. This is a second order linear homogeneous differential equation and has (as one might imagine) well known and well understood solutions.

Note well: At the tension in the string increases, so does the wave velocity. As the mass density of the string increases, the wave velocity decreases. This makes *physical sense*. As tension goes up the restoring force is greater. As mass density goes up one accelerates less for a given tension.

3.4 Solutions to the Wave Equation

In one dimension there are at least three distinct solutions to the wave equation that we are interested in. Two of these solutions *propagate* along the string – energy is *transported* from one place to another by the wave. The third is a *stationary* solution, in the sense that the wave doesn't propagate in one direction or the other (not in the sense that the string doesn't move). But first:

3.4.1 An Important Property of Waves: Superposition

The wave equation is *linear*, and hence it is easy to show that if $y_1(x,t)$ solves the wave equation and $y_2(x,t)$ (independent of y_1) *also* solves the wave equation, then:

$$y(x,t) = Ay_1(x,t) + By_2(x,t) \tag{107}$$

solves the wave equation for arbitrary (complex) A and B.

This property of waves is most powerful and sublime.

3.4.2 Arbitrary Waveforms Propagating to the Left or Right

The first solution we can discern by noting that the wave equation equates a second derivative in time to a second derivative in space. Suppose we write the solution as $f(u)$ where u is an unknown function of x and t and substitute it into the differential equation and use the chain rule:

$$\frac{d^2 f}{du^2}(\frac{du}{dt})^2 - v^2 \frac{d^2 f}{du^2}(\frac{du}{dx})^2 = 0 \tag{108}$$

or

$$\frac{d^2 f}{du^2}\left\{(\frac{du}{dt})^2 - v^2(\frac{du}{dx})^2\right\} = 0 \tag{109}$$

$$\frac{du}{dt} = \pm v \frac{du}{dx} \tag{110}$$

with a simple solution:

$$u = x \pm vt \tag{111}$$

What this tells us is that *any* function

$$y(x, t) = f(x \pm vt) \tag{112}$$

satisfies the wave equation. *Any* shape of wave created on the string and propagating to the right or left is a solution to the wave equation, although not all of these waves will vanish at the ends of a string.

3.4.3 Harmonic Waveforms Propagating to the Left or Right

An interesting special case of this solution is the case of *harmonic* waves propagating to the left or right. Harmonic waves are simply waves that oscillate with a given harmonic frequency. For example:

$$y(x, t) = y_0 \sin(x - vt) \tag{113}$$

is one such wave. y_0 is called the *amplitude* of the harmonic wave. But what sorts of parameters describe the wave itself? Are there more than one harmonic waves?

This particular wave looks like a sinusoidal wave propagating to the right (positive x direction). But this is not a very convenient parameterization. To better describe a general harmonic wave, we need to introduce the following quantities:

- The **frequency** f. This is the number of cycles per second that pass a point or that a point on the string moves up and down.
- The **wavelength** λ. This the distance one has to travel down the string to return to the same point in the wave cycle at any given instant in time.

To convert x (a distance) into an angle in radians, we need to multiply it by 2π radians per wavelength. We therefore define the *wave number*:

$$k = \frac{2\pi}{\lambda} \tag{114}$$

and write our harmonic solution as:

$$\begin{aligned} y(x,t) &= y_0 \sin(k(x-vt)) && (115) \\ &= y_0 \sin(kx-kvt) && (116) \\ &= y_0 \sin(kx-\omega t) && (117) \end{aligned}$$

where we have used the following train of algebra in the last step:

$$kv = \frac{2\pi}{\lambda} v = 2\pi f = \frac{2\pi}{T} = \omega \qquad (118)$$

and where we see that we have two ways to write v:

$$v = f\lambda = \frac{\omega}{k} \qquad (119)$$

As before, you should simply *know* every relation in this set of algebraic relations between λ, k, f, ω, v to save time on tests and quizzes. Of course there is also the harmonic wave travelling to the left as well:

$$y(x,t) = y_0 \sin(kx+\omega t). \qquad (120)$$

A final observation about these harmonic waves is that because arbitrary functions can be *expanded* in terms of harmonic functions (e.g. Fourier Series, Fourier Transforms) and because the wave equation is linear and its solutions are superposable, the two solution forms above are not really distinct. One can expand the "arbitrary" $f(x-vt)$ in a sum of $\sin(kx-\omega t)$'s for special frequencies and wavelengths. In one dimension this doesn't give you much, but in two or more dimensions this process helps one compute the *dispersion* of the wave caused by the wave "spreading out" in multiple dimensions and reducing its amplitude.

3.4.4 Stationary Waves

The third special case of solutions to the wave equation is that of *standing waves*. They are especially apropos to waves on a string fixed at one or both ends. There are two ways to find these solutions from the solutions above. A harmonic wave travelling to the right and hitting the end of the string (which is fixed), it has no choice but to reflect. This is because the *energy* in the string cannot just disappear, and if the end point is fixed no work can be done by it on the peg to which it is attached. The reflected wave has to have the same amplitude and frequency as the incoming wave. What does the *sum* of the incoming and reflected wave look like in this special case?

Suppose one adds two harmonic waves with equal amplitudes, the same wavelengths and frequencies, but that are travelling in *opposite* directions:

$$\begin{aligned} y(x,t) &= y_0\left(\sin(kx-\omega t)+\sin(kx+\omega t)\right) && (121)\\ &= 2y_0\sin(kx)\cos(\omega t) && (122)\\ &= A\sin(kx)\cos(\omega t) && (123)\end{aligned}$$

(where we give the standing wave the arbitrary amplitude A). Since all the solutions above are independent of the phase, a second useful way to write stationary waves is:

$$y(x,t) = A\cos(kx)\cos(\omega t) \tag{124}$$

Which of these one uses depends on the details of the boundary conditions on the string.

In this solution a sinusoidal form oscillates harmonically up and down, but the solution has some very important new properties. For one, it is always *zero* when $x = 0$ for all possible *lambda*:

$$y(0,t) = 0 \tag{125}$$

For a given λ there are certain other x positions where the wave vanishes at all times. These positions are called *nodes* of the wave. We see that there are nodes for any L such that:

$$y(L,t) = A\sin(kL)\cos(\omega t) = 0 \tag{126}$$

which implies that:

$$kL = \frac{2\pi L}{\lambda} = \pi, 2\pi, 3\pi, \ldots \tag{127}$$

or

$$\lambda = \frac{2L}{n} \tag{128}$$

for $n = 1, 2, 3, ...$

Only waves with these wavelengths and their associated frequencies can persist on a string of length L fixed at both ends so that

$$y(0,t) = y(L,t) = 0 \tag{129}$$

(such as a guitar string or harp string). Superpositions of these waves are what give guitar strings their particular tone.

It is also possible to stretch a string so that it is fixed at one end but so that the *other* end is *free to move* – to slide up and down without friction on a greased rod, for example. In this case, instead of having a node at the free end (where the wave itself vanishes), it is pretty easy to see that the *slope* of the wave at the end has to vanish. This is because if the slope were not zero, the terminating rod would be pulling up or down on the string, contradicting our requirement that the rod be frictionless and not *able* to pull the string up or down, only directly to the left or right due to tension.

The slope of a sine wave is zero only when the sine wave itself is a maximum or minimum, so that the wave on a string free at an end must have an *antinode* (maximum magnitude of its amplitude) at the free end. Using the same standing wave form we derived above, we see that:

$$kL = \frac{2\pi L}{\lambda} = \pi/2, 3\pi/2, 5\pi/2 \ldots \tag{130}$$

for a string fixed at $x = 0$ and free at $x = L$, or:

$$\lambda = \frac{4L}{2n-1} \tag{131}$$

for $n = 1, 2, 3, \ldots$

There is a second way to obtain the standing wave solutions that particularly exhibits the relationship between waves and harmonic oscillators. One assumes that the solution $y(x,t)$ can be written as the *product* of a fuction in x alone and a second function in t alone:

$$y(x,t) = X(x)T(t) \tag{132}$$

If we substitute this into the differential equation and divide by $y(x,t)$ we get:

$$\begin{aligned}
\frac{d^2y}{dt^2} = X(x)\frac{d^2T}{dt^2} &= v^2\frac{d^2y}{dx^2} = v^2T(t)\frac{d^2X}{dx^2} && (133)\\
\frac{1}{T(t)}\frac{d^2T}{dt^2} &= v^2\frac{1}{X(x)}\frac{d^2X}{dx^2} && (134)\\
&= -\omega^2 && (135)
\end{aligned}$$

where the last line follows because the second line equations a function of t (only) to a function of x (only) so that both terms must equal a constant. This is then the two equations:

$$\frac{d^2T}{dt^2} + \omega^2 T = 0 \tag{136}$$

and

$$\frac{d^2X}{dt^2} + k^2 X = 0 \tag{137}$$

(where we use $k = \omega/v$).

From this we see that:

$$T(t) = T_0 \cos(\omega t + \phi) \tag{138}$$

and

$$X(x) = X_0 \cos(kx + \delta) \tag{139}$$

so that the most general stationary solution can be written:

$$y(x,t) = y_0 \cos(kx + \delta)\cos(\omega t + \phi) \tag{140}$$

3.5 Reflection of Waves

We argued above that waves have to reflect of of the ends of stretched strings because of energy conservation. This is true independent of whether the end is fixed or free – in neither case can the string do work on the wall or rod to which it is affixed. However, the behavior of the reflected wave is different in the two cases.

Suppose a wave *pulse* is incident on the fixed end of a string. One way to "discover" a wave solution that apparently conserves energy is to imagine that the string *continues* through the barrier. At the same time the pulse hits the barrier, an *identical* pulse hits the barrier from the other, "imaginary" side.

Since the two pulses are identical, energy will clearly be conserved. The one going from left to right will transmit its energy onto the imaginary string beyond at the same rate energy appears going from right to left from the imaginary string.

However, we still have two choices to consider. The wave from the imaginary string could be *right side up* the same as the incident wave or *upside down*. Energy is conserved either way!

If the right side up wave (left to right) encounters an upside down wave (right to left) they will always be *opposite* at the barrier, and when superposed they will *cancel* at the barrier. This corresponds to a *fixed string*. On the other hand, if a right side up wave encounters a right side up wave, they will *add* at the barrier with opposite *slope*. There will be a maximum at the barrier with zero slope – just what is needed for a free string.

From this we deduce the general rule that wave pulses *invert* when reflected from a fixed boundary (string fixed at one end) and reflect right side up from a free boundary.

When two strings of different weight (mass density) are connected, wave pulses on one string are both transmitted onto the other and are generally partially reflected from the boundary. Computing the transmitted and reflected waves is straightforward but beyond the scope of this class (it starts to involve real math and studies of boundary conditions). However, the following qualitative properties of the transmitted and reflected waves should be learned:

- Light string (medium) to heavy string (medium): Transmitted pulse right side up, reflected pulse inverted. (A fixed string boundary is the limit of attaching to an "infinitely heavy string").

- Heavy string to light string: Transmitted pulse right side up, reflected pulse right side up. (a free string boundary is the limit of attaching to a "massless string").

3.6 Energy

Clearly a wave can carry energy from one place to another. A cable we are coiling is hung up on a piece of wood. We flip a pulse onto the wire, it runs down to the piece of wood and knocks the wire free. Our lungs and larnyx create sound waves, and those waves trigger neurons in ears far away. The sun releases nuclear energy, and a few minutes later that energy, propagated to earth as a light wave, creates sugar energy stores inside a plant that are still later released while we play basketball. Since moving energy around seems to be important, perhaps we should figure out how a wave manages it.

Let us restrict our attention to a harmonic wave of known angular frequency ω. Our results will still be quite general, because arbitrary wave pulses can be fourier decomposed as noted above. Consider a small piece of the string of length dx and mass $dm = \mu dx$. This piece of string, displaced to its position $y(x,t)$, will have potential energy:

$$\begin{aligned}
dU &= \frac{1}{2}dm\omega^2 y^2(x,t) && (141)\\
&= \frac{1}{2}\mu dx\omega^2 y^2(x,t) && (142)\\
&= \frac{1}{2}A^2\mu\omega^2\sin^2(kx-\omega t)dx && (143)
\end{aligned}$$

We can easily integrate this over any specific interval. Let us pick a particular time $t = 0$ and integrate it over a single wavelength:

$$\begin{aligned}
U &= \int_0^\lambda \frac{1}{2}A^2\mu\omega^2\sin^2(kx)dx && (144)\\
&= \frac{1}{2k}A^2\mu\omega^2\int_0^\lambda \sin^2(kx)kdx && (145)\\
&= \frac{1}{2k}A^2\mu\omega^2\int_0^2 \pi\sin^2(\theta)d\theta && (146)\\
&= \frac{1}{4}A^2\mu\omega^2\lambda && (147)
\end{aligned}$$

Now we need to compute the kinetic energy in a wavelength at the same instant.

$$\begin{aligned}
dK &= \frac{1}{2}dm\left(\frac{dy}{dt}\right)^2 && (148)\\
&= \frac{1}{2}A^2\mu\omega^2\cos^2(kx-\omega t)dx && (149)
\end{aligned}$$

which has exactly the same integral:

$$K = \frac{1}{4}A^2\mu\omega^2\lambda \qquad (150)$$

so that the total energy in a wavelength of the wave is:

$$E_{\text{tot}} = \frac{1}{2}\mu\omega^2 A^2\lambda \qquad (151)$$

Study the dependences in this relation. Energy depends on the *amplitude squared!* (Emphasis to convince you to remember this! It is important!)

It depends on the mass per unit length times the length (the mass of the segment). It depends on the frequency squared.

This energy *moves* as the wave propagates down the string. If you are sitting at some point on the string, all the energy in one wavelength passes you in one period of oscillation. This lets us compute the *power* carried by the string – the energy per unit time that passes us going from left to right:

$$P = \frac{E}{T} = \frac{1}{2}\mu\omega^2 A^2 \lambda f = \frac{1}{2}\mu\omega^2 A^2 v \qquad (152)$$

We can think of this as being the *energy per unit length* (the total energy per wavelength divided by the wavelength) times the *velocity of the wave*. This is a very *good* way to think of it as we prepare to study light waves, where a very similar relation will hold.

4 Sound

4.1 Sound Summary

- **Speed of Sound** in a fluid

$$v = \sqrt{\frac{B}{\rho}} \tag{153}$$

where B is the bulk modulus of the fluid and ρ is the density. These quantities vary with pressure and temperature.

- **Speed of Sound** in air is $v_a \approx 340$ m/sec.

- **Doppler Shift: Moving Source**

$$f' = \frac{f_0}{(1 \mp \frac{v_s}{v_a})} \tag{154}$$

where f_0 is the unshifted frequency of the sound wave for receding (+) and approaching (-) source.

- **Doppler Shift: Moving Receiver**

$$f' = f_0(1 \pm \frac{v_s}{v_a}) \tag{155}$$

where f_0 is the unshifted frequency of the sound wave for receding (-) and approaching (+) receiver.

- **Stationary Harmonic Waves**

$$y(x,t) = y_0 \sin(kx)\cos(\omega t) \tag{156}$$

for displacement waves in a pipe of length L closed at one or both ends. This solution has a node at $x = 0$ (the closed end). The permitted resonant frequencies are determined by:

$$kL = n\pi \tag{157}$$

for $n = 1, 2...$ (both ends closed, nodes at both ends) or:

$$kL = \frac{2n-1}{2}\pi \tag{158}$$

for $n = 1, 2, ...$ (one end closed, nodes at the closed end).

- **Beats** If two sound waves of equal amplitude and slightly different frequency are added:

$$s(x,t) = s_0 \sin(k_0 x - \omega_0 t) + s_0 \sin(k_1 x - \omega_1 t) \quad (159)$$

$$= 2s_0 \sin(\frac{k_0 + k_1}{2} x - \frac{\omega_0 + \omega_1}{2} t) \cos(\frac{k_0 - k_1}{2} x - \frac{\omega_0 - \omega_1}{2} t) \quad (160)$$

  which describes a wave with the average frequency and twice the amplitude modulated so that it "beats" (goes to zero) at the difference of the frequencies $\delta f = |f_1 - f_0|$.

4.2 Sound Waves in a Fluid

Waves propagate in a fluid much in the same way that a disturbance propagates down a closed hall crowded with people. If one shoves a person so that they knock into their neighbor, the neighbor falls against *their* neighbor (and shoves back), and their neighbor shoves against their still further neighbor and so on.

Such a wave differs from the transverse waves we studied on a string in that the displacement of the medium (the air molecules) is *in the same direction* as the direction of propagation of the wave. This kind of wave is called a *longitudinal* wave.

Although different, sound waves can be related to waves on a string in many ways. Most of the similarities and differences can be traced to one thing: a string is a one dimensional medium and is characterized only by length; a fluid is typically a three dimensional medium and is characterized by a volume.

Air (a typical fluid that supports sound waves) does not support "tension", it is under pressure. When air is compressed its molecules are shoved closer together, altering its density and occupied volume. For small changes in volume the pressure alters approximately *linearly* with a coefficient called the "bulk modulus" B describing the way the pressure increases as the fractional volume decreases. Air does not have a mass per unit length μ, rather it has a mass per unit volume, ρ.

The velocity of waves in air is given by

$$v_a = \sqrt{\frac{B}{\rho}} \approx 343\text{m/sec} \quad (161)$$

The "approximately" here is fairly serious. The actual speed varies according to things like the air pressure (which varies significantly with altitude and with the weather at any given altitude as low and high pressure areas move around on the earth's surface) and the temperature (hotter molecules push each other apart more strongly at any given density). The speed of sound can vary by a few percent from the approximate value given above.

4.3 Sound Wave Solutions

Sound waves can be characterized one of two ways: as organized fluctuations in the *position* of the molecules of the fluid as they oscillate around an equilibrium displacement or as organized fluctuations in the *pressure* of the fluid as molecules are crammed closer together or are diven farther apart than they are on average in the quiescent fluid.

Sound waves propagate in one direction (out of three) at any given point in space. This means that in the direction *perpendicular* to propagation, the wave is spread out to form a "wave front". The wave front can be nearly arbitrary in shape initially; thereafter it evolves according to the mathematics of the wave equation in three dimensions (which is similar to but a bit more complicated than the wave equation in one dimension).

To avoid this complication and focus on general properties that are commonly encountered, we will concentrate on two particular kinds of solutions:

1. *Plane Wave* solutions. In these solutions, the entire wave moves in one direction (say the x direction) and the wave front is a 2-D plane perpendicular to the direction of propagation. These (displacement) solutions can be written as (e.g.):

   $$s(x, t) = s_0 \sin(kx - \omega t) \tag{162}$$

   where s_0 is the maximum displacement in the travelling wave (which moves in the x direction) and where all molecules in the entire *plane* at position x are displaced by the same amount.

   Waves far away from the sources that created them are best described as plane waves. So are waves propagating down a constrained environment such as a tube that permits waves to only travel in "one direction".

2. *Spherical Wave* solutions. Sound is often emitted from a source that is highly localized (such as a hammer hitting a nail, or a loudspeaker). If

the sound is emitted equally in all directions from the source, a spherical wavefront is formed. Even if it is not emitted equally in all directions, sound from a localized source will generally form a spherically curved wavefront as it travels away from the point with constant speed. The displacement of a spherical wavefront *decreases* as one moves further away from the source because the *energy* in the wavefront is spread out on larger and larger surfaces. Its form is given by:

$$s(r,t) = \frac{s_0}{r}\sin(kr - \omega t) \tag{163}$$

where r is the radial distance away from the point-like source.

4.4 Sound Wave Intensity

The energy density of sound waves is given by:

$$\frac{dE}{dV} = \frac{1}{2}\rho\omega^2 s^2 \tag{164}$$

(again, very similar in form to the energy density of a wave on a string). However, this energy per unit *volume* is propagated in a single direction. It is therefore spread out so that it crosses an *area*, not a single point. Just how much energy an object receives therefore depends on how much *area* it intersects in the incoming sound wave, not just on the energy density of the sound wave itself.

For this reason the energy carried by sound waves is best measured by *intensity*: the energy per unit time per unit area perpendicular to the direction of wave propagation. Imagine a box with sides given by ΔA (perpendicular to the direction of the wave's propagation) and $v\Delta t$ (in the direction of the wave's propagation. All the energy in this box crosses through ΔA in time Δt. That is:

$$\Delta E = (\frac{1}{2}\rho\omega^2 s^2)\Delta A v \Delta t \tag{165}$$

or

$$I = \frac{\Delta E}{\Delta A \Delta t} = \frac{1}{2}\rho\omega^2 s^2 v \tag{166}$$

which looks very much like the *power* carried by a wave on a string. In the case of a plane wave propagating down a narrow tube, it is very similar – the power of the wave is the intensity times the tube's cross section.

However, consider a spherical wave. For a spherical wave, the intensity looks something like:

$$I(r,t) = \frac{1}{2}\rho\omega^2 \frac{s_0^2 \sin^2(kr - \omega t)}{r^2} v \tag{167}$$

which can be written as:

$$I(r,t) = \frac{P}{r^2} \tag{168}$$

where P is the total power in the wave.

This makes sense from the point of view of energy conservation and symmetry. If a source emits a power P, that energy has to cross each successive spherical surface that surrounds the source. Those surfaces have an area that varies like $A = 4\pi r^2$. A surface at $r = 2r_0$ has 4 times the area of one at $r = r_0$, but the *same* total power has to go through both surfaces. Consequently, the intensity at the $r = 2r_0$ surface has to be 1/4 the intensity at the $r = r_0$ surface.

It is important to remember this argument, simple as it is. Think back to Newton's law of gravitation. Remember that gravitational field diminishes as $1/r^2$ with the distance from the source. Electrostatic field also diminishes as $1/r^2$. There seems to be a shared connection between symmetric propagation and spherical geometry; this will form the basis for *Gauss's Law* in electrostatics and much beautiful math.

4.5 Doppler Shift

Everybody has heard the doppler shift in action. It is the rise (or fall) in frequency observed when a source/receiver pair approach (or recede) from one another. In this section we will derive expressions for the doppler shift for moving source and moving receiver.

4.5.1 Moving Source

Suppose your receiver (ear) is stationary, while a source of harmonic sound waves at fixed frequency f_0 is approaching you. As the waves are emitted by the source they have a fixed wavelength $\lambda_0 = v_a/f_0 = v_a T$ and expand spherically from the point where the source was at the time the wavefront was emitted.

However, that point moves in the direction of the receiver. In the time between wavefronts (one period T) the source moves a distance v_sT. The distance between successive wavefronts in the direction of motion is thus:

$$\lambda' = \lambda_0 - v_sT \tag{169}$$

and (factoring and freely using e.g. $f = 1/T$):

$$\frac{v_a}{f'} = \frac{v_a - v_s}{f_0} \tag{170}$$

or

$$f' = \frac{f_0}{1 - \frac{v_s}{v_a}} \tag{171}$$

If the source is moving away from the receiver, everything is the same except now the wavelength is shifted to be bigger and the frequency smaller (as one would expect from changing the sign on the velocity):

$$f' = \frac{f_0}{1 + \frac{v_s}{v_a}} \tag{172}$$

4.5.2 Moving Receiver

Now imagine that the source of waves at frequency f_0 is stationary but the receiver is moving towards the source. The source is thus surrounded by spherical wavefronts a distance $\lambda_0 = v_aT$ apart. At $t = 0$ the receiver crosses one of them. At a time T' later, it has moved a distance $d = v_rT'$ in the direction of the source, and the wave from the source has moved a distance $D = v_aT'$ toward the receiver, and the receiver encounters the next wave front. That is:

$$\begin{aligned}
\lambda_0 &= d + D && (173)\\
&= v_rT' + v_aT' && (174)\\
&= (v_r + v_a)T' && (175)\\
v_aT &= (v_r + v_a)T' && (176)
\end{aligned}$$

We use $f_0 = 1/T$, $f' = 1/T'$ (where T' is the apparent time between wavefronts to the receiver) and rearrange this into:

$$f' = f_0(1 + \frac{v_r}{v_a}) \tag{177}$$

Again, if the receiver is moving away from the source, everything is the same but the sign of v_r, so one gets:

$$f' = f_0(1 - \frac{v_r}{v_a}) \tag{178}$$

4.5.3 Moving Source and Moving Receiver

This result is just the product of the two above – moving source causes one shift and moving receiver causes another to get:

$$f' = f_0 \frac{1 \mp \frac{v_r}{v_a}}{1 \pm \frac{v_s}{v_a}} \tag{179}$$

where in both cases *relative* approach shifts the frequency up and *relative* recession shifts the frequency down.

I do *not* recommend memorizing these equations – I don't have them memorized myself. It is *very* easy to confuse the forms for source and receiver, and the derivations take a few seconds and are likely worth points in and of themselves. If you're going to memorize anything, memorize the *derivation* (a process I call "learning", as opposed to "memorizing"). In fact, this is excellent advice for 90% of the material you learn in this course!

4.6 Standing Waves in Pipes

Everybody has created a stationary resonant harmonic sound wave by whistling or blowing over a beer bottle or by swinging a garden hose or by playing the organ. In this section we will see how to compute the harmonics of a given (simple) pipe geometry for an imaginary organ pipe that is open or closed at one or both ends.

The way we proceed is straightforward. Air cannot penetrate a closed pipe end. The air molecules at the very end are therefore "fixed" – they cannot displace into the closed end. The *closed* end of the pipe is thus a *displacement node*. In order *not* to displace air the closed pipe end has to exert a force on the molecules by means of pressure, so that the closed end is a pressure antinode.

At an open pipe end the argument is inverted. The pipe is open to the air (at fixed background/equilibrium pressure) so that there must be a pressure node at the open end. Pressure and displacement are $\pi/2$ out of phase, so that the *open* end is also a *displacement antinode*.

Actually, the air pressure in the standing wave doesn't instantly equalize with the background pressure at an open end – it sort of "bulges" out of the pipe a bit. The displacement antinode is therefore just *outside* the pipe end, not *at* the pipe end. You may still draw a displacement antinode (or pressure node) as if they occur at the open pipe end; just remember that the distance from the open end to the first displacement node is *not* a very accurate measure of a quarter wavelength and that open organ pipes are a bit "longer" than they appear from the point of view of computing their resonant harmonics.

Once we understand the boundary conditions at the ends of the pipes, it is pretty easy to write down expressions for the standing waves and to deduce their harmonic frequencies.

4.6.1 Pipe Closed at Both Ends

There are displacement nodes at both ends. This is just like a string fixed at both ends:

$$s(x,t) = s_0 \sin(k_n x) \cos(\omega_n t) \tag{180}$$

which has a node at $x = 0$ for all k. To get a node at the other end, we require:

$$\sin(k_n L) = 0 \tag{181}$$

or

$$k_n L = n\pi \tag{182}$$

for $n = 1, 2, 3....$ This converts to:

$$\lambda_n = \frac{2L}{n} \tag{183}$$

and

$$f_n = \frac{v_a}{\lambda_n} = \frac{v_a n}{2L} \tag{184}$$

4.6.2 Pipe Closed at One End

There is a displacement node at the closed end, and an antinode at the open end. This is just like a string fixed at one end and free at the other. Let's arbitrarily make $x = 0$ the closed end. Then:

$$s(x,t) = s_0 \sin(k_n x) \cos(\omega_n t) \tag{185}$$

has a node at $x = 0$ for all k. To get an antinode at the other end, we require:

$$\sin(k_n L) = \pm 1 \tag{186}$$

or

$$k_n L = \frac{2n-2}{2}\pi \tag{187}$$

for $n = 1, 2, 3...$ (odd half-integral multiples of π. This converts to:

$$\lambda_n = \frac{2L}{2n-1} \tag{188}$$

and

$$f_n = \frac{v_a}{\lambda_n} = \frac{v_a(2n-1)}{4L} \tag{189}$$

4.6.3 Pipe Open at Both Ends

There are displacement antinodes at both ends. This is just like a string free at both ends. We could therefore proceed from

$$s(x,t) = s_0 \cos(k_n x)\cos(\omega_n t) \tag{190}$$

and

$$\cos(k_n L) = \pm 1 \tag{191}$$

but we could *also* remember that there are *pressure* nodes at both ends, which makes them like a string fixed at both ends again. Either way one will get the same frequencies one gets for the pipe closed at both ends above (as the cosine is ± 1 for $k_n L = n\pi$) but the picture of the nodes is still different – be sure to draw displacement antinodes at the open ends!

4.7 Beats

If you have ever played around with a guitar, you've probably noticed that if two strings are fingered to be the "same note" but are really slightly out of tune and are struck together, the resulting sound "beats" – it modulates up and down in intensity at a low frequency often in the ballpark of a few cycles per second.

Beats occur because of the superposition principle. We can add any two (or more) solutions to the wave equation and still get a solution to the wave equation, even if the solutions have different frequencies. Recall the identity:

$$\sin(A) + \sin(B) = 2\sin(\frac{A+B}{2})\cos(\frac{A-B}{2}) \tag{192}$$

If one adds two waves with different wave numbers/frequencies and uses this rule, one gets

$$\begin{aligned} s(x,t) &= s_0 \sin(k_0 x - \omega_0 t) + s_0 \sin(k_1 x - \omega_1 t) & (193) \\ &= 2s_0 \sin(\frac{k_0+k_1}{2}x - \frac{\omega_0+\omega_1}{2}t)\cos(\frac{k_0-k_1}{2}x - \frac{\omega_0-\omega_1}{2}t) & (194) \end{aligned}$$

This describes a wave that has twice the maximum amplitude, the *average* frequency (the first term), and a second term that (at any point x) oscillates like $\cos(\frac{\Delta\omega t}{2})$.

The "frequency" of this second modulating term is $\frac{f_0-f_1}{2}$, but the ear *cannot hear* the inversion of phase that occurs when it is negative and the difference is small. It just hears maximum amplitude in the rapidly oscillating *average* frequency part, which goes to *zero* when the slowing varying cosine does, twice per cycle. The ear then *hears* two beats per cycle, making the "beat frequency":

$$f_{\text{beat}} = \Delta f = |f_0 - f_1| \tag{195}$$

4.8 Interference and Sound Waves

We will not cover interference and diffraction of harmonic sound waves in this course. Beats are a common experience in sound as is the doppler shift, but sound wave interference is not so common an experience (although it can definitely and annoyingly occur if you hook up speakers in your stereo out of phase). Interference *will* be treated next semester in the context of coherent light waves. *Just* to give you a head start on that, we'll indicate the basic ideas underlying interference here.

Suppose you have two sources that are at the *same* frequency and have the *same* amplitude and phase but are at different locations. One source might be a distance x away from you and the other a distance $x + \Delta x$ away from you. The waves from these two sources add like:

$$\begin{aligned} s(x,t) &= s_0 \sin(kx - \omega t) + s_0 \sin(k(x+\Delta x) - \omega t) & (196) \\ &= 2s_0 \sin(k(x + \frac{\Delta x}{2} - \omega t)\cos(k\frac{\Delta x}{2}) & (197) \end{aligned}$$

The sine part describes a wave with twice the amplitude, the same frequency, but shifted slightly in phase by $k\Delta x/2$. The cosine part is *time independent* and *modulates* the first part. For some values of Δx it can *vanish*. For others it can have magnitude one.

The intensity of the wave is what our ears hear – they are insensitive to the phase (although certain echolocating species such as bats may be sensitive to phase information as well as frequency). The average intensity is proportional to the wave amplitude *squared*:

$$I_0 = \frac{1}{2}\rho\omega^2 s_0^2 v \tag{198}$$

With two sources (and a maximum amplitude of two) we get:

$$\begin{aligned} I &= \frac{1}{2}\rho\omega^2(2^2 s_0^2 \cos^2(k\frac{\Delta x}{2})v & (199) \\ &= 4I_0 \cos^2(k\frac{\Delta x}{2}) & (200) \end{aligned}$$

There are two cases of particular interest in this expression. When

$$\cos^2(k\frac{\Delta x}{2}) = 1 \tag{201}$$

one has *four times* the intensity of one source at peak. This occurs when:

$$k\frac{\Delta x}{2} = n\pi \tag{202}$$

(for $n = 0, 1, 2...$) or

$$\Delta x = n\lambda \tag{203}$$

If the *path difference* contains an *integral number of wavelengths* the waves from the two sources arrive *in phase*, add, and produce sound that has twice the amplitude and four times the intensity. This is called *complete constructive interference.*

On the other hand, when

$$\cos^2(k\frac{\Delta x}{2}) = 0 \tag{204}$$

the sound intensity *vanishes.* This is called *destructive* interference. This occurs when

$$k\frac{\Delta x}{2} = \frac{2n+1}{2}\pi \tag{205}$$

(for $n = 0, 1, 2...$) or

$$\Delta x = \frac{2n+1}{2}\lambda \qquad (206)$$

If the path difference contains a *half integral* number of wavelengths, the waves from two sources arrive exactly out of phase, and *cancel*. The sound intensity vanishes.

You can see why this would make hooking your speakers up out of phase a bad idea. If you hook them up out of phase the waves *start* with a phase difference of π – one speaker is pushing out while the other is pulling in. If you sit equidistant from the two speakers and then harmonic waves with the same frequency from a single source coming from the two speakers *cancel* as they reach you (usually not perfectly) and the music sounds very odd indeed, because other parts of the music are not being played equally from the two speakers and don't cancel.

You can also see that there are many other situations where constructive or destructive interference can occur, both for sound waves and for other waves including water waves, light waves, even waves on strings. Our "standing wave solution" can be rederived as the superposition of a left- and right-travelling harmonic wave, for example. You can have interference from more than one source, it doesn't have to be just two.

This leads to some really excellent engineering. Ultrasonic probe arrays, radiotelescope arrays, sonar arrays, diffraction gratings, holograms, are all examples of interference being put to work. So it is worth it to learn the general idea as early as possible, even if it isn't assigned.

5 Fluids

5.1 Fluids Summary

- **Fluids** are states of matter characterized by a lack of long range order. They are characterized by their **density** ρ and their **compressibility**. Liquids such as water are (typically) relatively incompressible; gases can be significantly compressed. Fluids have other characteristics, for example viscosity (how "sticky" the fluid is). We will ignore these in this course.

- **Pressure** is the force per unit area exerted by a fluid on its surroundings:
$$P = F/A \tag{207}$$
Its SI units are *pascals* where 1 pascal = 1 newton/meter squared. Pressure is also measured in "atmospheres" (the pressure of air at or near sea level) where 1 atmosphere $\approx 10^5$ pascals. The pressure in an incompressible fluid varies with depth according to:
$$P = P_0 + \rho g D \tag{208}$$
where P_0 is the pressure at the top and D is the depth.

- **Pascal's Principle** Pressure applied to a fluid is transmitted undiminished to all points of the fluid.

- **Conservation of Flow** We will study only steady/laminar flow in the absence of turbulence and viscosity.
$$A_1 v_1 = A_2 v_2 \tag{209}$$

- **Bernoulli's Equation**
$$P + \frac{1}{2}\rho v^2 + \rho g h = \text{constant} \tag{210}$$

- **Toricelli's Rule:** If a fluid is flowing through a very small hole (for example at the bottom of a large tank) then the velocity of the fluid at the large end can be neglected in Bernoulli's Equation.

- **Archimedes' Principle** The buoyant force on an object
$$F_b = \rho g V_{\rm disp} \tag{211}$$
where frequency $V_{\rm disp}$ is the volume of fluid displaced by an object.

- **Venturi Effect** The pressure in a fluid *increases* as the velocity of the fluid *decreases.* This is responsible for e.g. the lift of an airplane wing.

5.2 Fluids

Fluids are the generic name given to two states of matter characterized by a lack of long range order and a high degree of mobility at the molecular scale. Fluids have the following properties:

- They usually assume the shape of any vessel they are placed in (exceptions are associated with surface effects such as surface tension and how well the fluid adheres to the surface in question).
- They are characterized by a mass per unit volume *density* ρ.
- They exert a *pressure* P (force per unit area) on themselves and any surfaces they are in contact with.
- The pressure can vary according to the dynamic and static properties of the fluid.
- The fluid has a measure of its "stickiness" and resistance to flow called *viscosity.* Viscosity is the internal friction of a fluid, more or less. We will treat fluids as being "ideal" and ignore viscosity in this course.
- Fluids are *compressible* – when the pressure in a fluid is increased, its volume descreases according to the relation:

  $$\Delta P = -B\frac{\Delta V}{V} \tag{212}$$

  where B is called the *bulk modulus* of the fluid (the equivalent of a spring constant).
- Fluids where B is a large number (so large changes in pressure create only tiny changes in fractional volume) are called *incompressible.* Water is an example of an incompressible fluid.
- Below a critical speed, the dynamic flow of a moving fluid tends to be laminar, where every bit of fluid moves parallel to its neighbors in response to pressure differentials and around obstacles. Above that

speed it becomes turbulent flow. Turbulent flow is quite difficult to treat mathematically and is hence beyond the scope of this introductory course – we will restrict our attention to ideal fluids either static or in laminar flow.

5.3 Static Fluids

A fluid in static equilibrium must support its own weight. If one considers a small circular box of fluid with area ΔA perpendicular to gravity and sides of thickness dx, the force on the sides cancels due to symmetry. The force pushing down at the top of the box is $P_0 \Delta A$. The force pushing up from the bottom of the box is $(P_0 + dP)\Delta A$. The weight of the fluid in the box is $w = \rho g \Delta A dx$. Thus:

$$\{(P_0 + dP) - P_0\} \Delta A = \rho g \Delta A dx \tag{213}$$

or

$$dP = \rho g dx \tag{214}$$

If ρ is itself a function of pressure/depth (as occurs with e.g. air) this can be a complicated expression to evaluate. For an incompressible fluid such as water, ρ is constant over rather large variations in pressure. In that case, integration yields:

$$P = P_0 + \rho g D \tag{215}$$

6 Zeroth Law of Thermodynamics

6.1 0th Law of Thermodynamics Summary

- **Thermal Equilibrium**

  A system with many microscopic components (for example, a gas, a liquid, a solid with many molecules) that is isolated from all forms of energy exchange and left alone for a "long time" moves toward a state of *thermal equilibrium*. A system in thermal equilibrium is characterized by a set of macroscopic quantities that depend on the system in question and characterize its "state" (such as pressure, volume, density) that do not change in time.

  Two systems are said to be in (mutual) thermal equilibrium if, when they are placed in "thermal contact" (basically, contact that permits the exchange of energy between them), their state variables do not change.

- **Zeroth Law of Thermodynamics**

  If system A is in thermal equilibrium with system C, and system B is in thermal equilibrium with system C, then system A is in thermal equilibrium with system B.

- **Temperature and Thermometers**

  The point of the Zeroth Law is that it is the basis of the thermometer. A thermometer is a portable device whose thermal state is related linearly to some simple property, for example its density or pressure. Once a suitable temperature scale is defined for the device, one can use it to measure the temperature of a variety of disparate systems in thermal equilibrium. Temperature thus characterizes thermal equilibrium.

- **Temperature Scales**

  1. **Fahrenheit**: This is one of the oldest scales, and is based on the coldest temperature that could be achieved with a mix of ice and alcohol. In it the freezing point of water is at 32° F, the boiling point of water is at 212° F.
  2. **Celsius or Centigrade**: This is a very sane system, where the freezing point of water is at 0° C and the boiling point is at 100° C. The degree size is thus 9/5 as big as the Fahrenheit degree.

3. **Kelvin or Absolute**: 0° K is the lowest possible temperature, where the internal energy of a system is at its absolute minimum. The degree size is the same as that of the Centigrade or Celsius scale. This makes the freezing point of water at atmospheric pressure 273.16° K, the boiling point at 373.16° K.

- **Thermal Expansion**
$$\Delta L = \alpha L \Delta T \tag{216}$$
where α is the *coefficient of linear expansion.* If one applies this in three dimensions:
$$\Delta V = \beta V \Delta T \tag{217}$$
where $\beta = 3\alpha$.

- **Ideal Gas Law**
$$PV = nRT = NkT \tag{218}$$
where $R = 8.315$ J/mol-K, and $k = R/N_A = 1.38 \times 10^{-23}$ J/K.

7 First Law of Thermodynamics

7.1 First Law of Thermodynamics Summary

- **Internal Energy**

  Internal energy is all the mechanical energy in all the components of a system. For example, in a monoatomic gas it might be the sum of the kinetic energies of all the gas atoms. In a solid it might be the sum of the kinetic and potential energies of all the particles that make up the solid.

- **Heat**

  Heat is a bit more complicated. It is internal energy as well, but it is internal energy that is *transferred* into or out of a given system. Furthermore, it is in some fundamental sense "disorganized" internal energy – energy with no particular organization, random energy. Heat flows into or out of a system in response to a temperature difference, always flowing from hotter temperature regions (cooling them) to cooler ones (warming them).

  Common units of heat include the ever-popular Joule and the *calorie* (the heat required to raise the temperature of 1 gram of water at 14.5° C to 15.5° C. Note that 1 cal = 4.186 J. Less common and more esoteric ones like the British Thermal Unit (BTU) and erg will be mostly ignored in this course; BTUs raise the temperature of one pound of water by one degree Fahrenheit, for example. Ugly.

- **Heat Capacity**

  If one adds heat to an object, its temperature usually increases (exceptions include at a state boundary, for example when a liquid boils). In many cases the temperature change is *linear* in the amount of heat added. We define the heat capacity C of an object from the relation:

  $$\Delta Q = C\Delta T \tag{219}$$

  where ΔQ is the heat that flows into a system to increase its temperature by ΔT. Many substances have a known heat capacity per unit mass. This permits us to also write:

  $$\Delta Q = mc\Delta T \tag{220}$$

where c is the *specific heat* of a substance. The specific heat of liquid water is approximately:

$$c_{\text{water}} = 1 calorie/gram -^\circ C \tag{221}$$

(as one might guess from the definition of the calorie above).

- **Latent Heat** As noted above, there are particular times when one can add heat to a system and not change its temperature. One such time is when the system is changing state from/to solid to/from liquid, or from/to liquid to/from gas. At those times, one adds (or removes) heat when the system is at fixed temperature until the state change is complete. The specific heat may well change across phase boundaries. There are two trivial equations to learn:

$$\Delta Q_f = mL_f \tag{222}$$
$$\Delta Q_v = mL_v \tag{223}$$

where L_f is the *latent heat of fusion* and L_v is the *latent heat of vaporization*. Two important numbers to keep in mind are $L_f(H_2O) = 333$ kJ/kg, and $L_v(H_2O) = 2260$ kJ/kg. Note the high value of the latter – the reason that "steam burns worse than water".

- **Work Done by a Gas**

$$W = \int_{V_i}^{V_f} PdV \tag{224}$$

This is the area under the $P(V)$ curve, suggesting that we draw lots of state diagrams on a P and V coordinate system. Both heat transfer and word depend on the *path* a gas takes $P(V)$ moving from one pressure and volume to another.

- **The First Law of Thermodynamics**

$$\Delta E_{\text{int}} = \Delta Q - W \tag{225}$$

In words, this is that the change in total mechanical energy of a system is equal to heat put into the system plus the work done *on* the system (which is minus the work done *by* the system, hence the minus above).

This is just, at long last, the fully *generalized* law of conservation of energy. All the cases where mechanical energy was not conserved in previous chapters because of nonconservative forces, the missing energy appeared as *heat*, energy that naturally flows from hotter systems to cooler ones.

- **Cyclic Processes** Most of what we study in these final sections will lead us to an understanding of simple heat engines based on gas expanding in a cylinder and doing work against a piston. In order to build a true engine, the engine has to go around in a repetitive *cycle*. This cycle typically is represented by a closed loop on a state e.g. $P(V)$ curve. A direct consequence of the 1st law is that **the net work done by the system per cycle is the area inside the loop of the $P(V)$ diagram**. Since the internal energy is the same at the beginning and the end of the cycle, it also tells us that:

$$\Delta Q_{\text{cycle}} = W_{\text{cycle}} \tag{226}$$

  the heat that flows into the system per cycle must exactly equal the work done by the system per cycle.

- **Adiabatic Processes** are processes (PV curves) such that no heat enters or leaves an (insulated) system.

- **Isothermal Processes** are processes where the temperature T of the system remains constant.

- **Isobaric Processes** are processes that occur at constant pressure.

- **Isovolumetric Processses** are processes that occur at constant volume.

- Work done by an Ideal Gas: Recall,

$$PV = NkT \tag{227}$$

  where N is the number of gas atoms or molecules. Isothermal work at (fixed) temperature T_0 is thus:

$$\begin{aligned} W &= \int_{V_1}^{V_2} \frac{NkT_0}{V} dV && (228) \\ &= NkT \ln(\frac{V_2}{V_1}) && (229) \end{aligned}$$

Isobaric work is trivial. $P = P_0$ is a constant, so

$$W = \int_{V_1}^{V_2} P_0 dV = P_0(V_2 - V_1) \tag{230}$$

Adiabatic work is a bit tricky and depends on some of the internal properties of the gas (for example, whether it is mono- or diatomic). We'll examine this in the next section.

8 Equipartition and Adiabatic Processes

8.1 Equipartition and Adiabatic Processes

- **Equipartition Theorem**

  There is $\frac{1}{2}kT$ of energy per degree of freedom of a molecule in a system. A monoatomic gas has three degrees of freedom (three dimensions of kinetic energy). A diatomic gas typically has five (three translational degrees, two rotational degrees, skip vibration and the third rotation). A solid typically has six (three translational degrees, three vibrational degrees). The internal energy of N molecules of a monoatomic gas are thus:

  $$E_{\rm tot} = \frac{3}{2}NkT \tag{231}$$

  and so forth.

- **Molecular Model of Ideal Gas** leads to $PV = NkT$ using the equipartition theorem.

- **Molar Specific Heats**

  At constant volume, no work is done and all heat that goes into a system increases its internal energy. At constant pressure, heat going into a system can both do work and increase internal energy and typically does both. We define:

  $$\Delta Q = nC_v\Delta T(\text{constantvolume}) \tag{232}$$
  $$\Delta Q = nC_p\Delta T(\text{constantpressure}) \tag{233}$$

  where C_v is the molar specific heat at constant volume and C_p is the molar specific heat at constant pressure (and n is the number of moles of gas). These specific heats are also often expressed per molecule with an obvious conversion based on N_A.

- $\mathbf{C_v}$

  $$C_v = \frac{1}{n}\frac{dE_{\rm tot}}{dT} = \frac{d_f}{2}R \tag{234}$$

  where d_f is the number of degrees of freedom of a molecule in the system (e.g. 3 for monoatomic gas, 5 for diatomic gas, 6 for solid).

- $\mathbf{C_p}$ From the first law:

  $$dQ = dE + dW = \frac{d_f}{2}nRdT + PdV = \frac{d_f}{2}nRdT + nRdT \tag{235}$$

(where the last follows because with P constant, dV must be proportional to dT) so

$$C_p = \frac{1}{n}\frac{dQ}{dT} = \frac{d_f}{2}R + R = C_v + R \tag{236}$$

- γ We define:
$$\gamma = \frac{C_p}{C_v} \tag{237}$$
Thus $\gamma = 5/3$ for monoatomic gases, $7/5$ for diatomic gases.

- Adiabatic Processes are characterized by $\Delta Q = 0$. From the first law, this means that:
$$dE = -dW \tag{238}$$
or
$$nC_v dT = nRdT = -PdV \tag{239}$$
If we take the full derivative of the ideal gas law:
$$d(PV) = PdV + VdP = nRdT \tag{240}$$
and substitute in the first law and rearrange, we get:
$$PdV + VdP = -\frac{R}{C_v}PdV \tag{241}$$
Collecting terms, rearranging, and dividing by PV we get:
$$\begin{aligned}
(1 + \frac{R}{C_v})PdV + VdP &= 0 && (242)\\
(\frac{C_v + R}{C_v})PdV + VdP &= 0 && (243)\\
(\frac{C_p}{C_v})PdV + VdP &= 0 && (244)\\
\gamma PdV + VdP &= 0 && (245)\\
\gamma\frac{dV}{V} + \frac{dP}{P} &= 0 && (246)
\end{aligned}$$
and integrating it leads us to
$$\ln P + \gamma \ln V = \text{constant} \tag{247}$$

or

$$PV^{\gamma} = \text{constant} \tag{248}$$

This is the equation for the PV curve of an adiabatic process. There are lots of ways to manipulate this algebraically by e.g. combining it with the ideal gas law to eliminate P or V in favor of T and the remaining one.

We will need the $PV^{\gamma} = \text{constant}$ result in order to solve problems involving adiabatic processes in cyclic heat engines, so learn it well.

9 Second Law of Thermodynamics

9.1 Second Law of Thermodynamics Summary

- **Heat Engines**

  A *heat engine* is a cyclic device that takes heat Q_H in from a *hot reservoir*, converts *some* of it to work W, and rejects the rest of it Q_C to a *cold reservoir* so that at the end of a cycle it is in the same state (and has the same internal energy) with which it began. The net work done per cycle is (recall) the area inside the PV curve.

  The *efficiency* of a heat engine is defined to be

$$\epsilon = \frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H} = 1 - \frac{Q_C}{Q_H} \tag{249}$$

- **Kelvin-Planck statement of the Second Law of Thermodynamics**

  It is impossible to construct a cyclic heat engine that produces no other effect but the absorption of energy from a hot reservoir and the production of an equal amount of work.

- **Refrigerators (and Heat Pumps)**

  A *refrigerator* is basically a cyclic heat engine run backwards. In a cycle it takes heat Q_C in from a cold reservoir, does work W on it, and rejects a heat Q_H to a hot reservoir. Its net effect is thus to make the cold reservoir colder (refrigeration) by removing heat from inside it to the warmer warm reservoir (warming it still further, e.g. as a heat pump). Both of these functions have practical applications – cooling our homes in summer, heating our homes in winter.

  The *coefficient of performance* of a refrigerator is defined to be

$$\mathrm{COP} = \frac{Q_C}{W} \tag{250}$$

  It is not uncommon for heat pumps to have a COP of 3-5 (depending on the temperature differential) giving them a significant economic advantage over resistive heating. The bad side is that they don't work terribly well when the temperature difference is large in degrees K.

- **Clausius Statement of the Second Law of Thermodynamics**

It is impossible to construct a cyclic refrigerator whose sole effect is the transfer of energy from a cold reservoir to a warm reservoir without the input of energy by work.

- bf Reversible Processes Reversible processes are ones where no friction or turbulence or dissipative forces are present that represent an additional source of energy loss or gain for a given system. For the purposes of this book, both adiabatic and isothermal processes are reversible. Irreversible processes include the transfer of heat energy from a hot to a cold reservoir in general – heat engines and refrigerators can be constructed whose steps in a cycle are all reversible, but the overall effect of transferring heat one way or the other is irreversible.

- **Carnot Engine**

  The *Carnot Cycle* is the archetypical reversible cycle, and a Carnot Cycle-based heat engine is one that does not dissipate any energy internally and uses only reversible steps. *Carnot's Theorem* states that no real heat engine operating between a hot reservoir at temperature T_H and a cold reservoir at temperature T_C can be more efficient than a Carnot engine operating between those two reservoirs.

  The Carnot efficiency is easy to compute (see text and lecture example). A Carnot Cycle consists of four steps:

  1. Isothermal expansion (in contact with the heat reservoir)
  2. Adiabatic expansion (after the heat reservoir is removed)
  3. Isothermal compression (in contact with the cold reservoir)
  4. Adiabatic compression (after the cold reservoir is removed)

  The efficiency of a Carnot Engine is:

$$\epsilon_{\text{Carnot}} = 1 - \frac{T_C}{T_H} \tag{251}$$

- **Entropy**

  Entropy S is a measure of disorder. The change in entropy of a system can be evaluated by integrating:

$$dS = \frac{dQ}{T} \tag{252}$$

between successive infinitesimally separated equilibrium states (the weasel language is necessary because temperature should be constant in equilibrium, but systems in equilibrium have constant entropy). Thus:

$$\Delta S = \int_{T_i} T_f \frac{dQ}{T} \tag{253}$$

has limited utility except for particularly simple processes (like the cooling of a hot piece of metal in a body of cold water.

We extend our definition of reversible processes. A reversible process is one where the entropy of the system does not change. An irreversible process increases the entropy of the system and its surroundings.

- **Entropy Statement of the Second Law of Thermodynamics**

  The entropy of the Universe never decreases. It either increases (for irreversible processes) or remains the same (for reversible processes).