Physics 202 Lecture 1

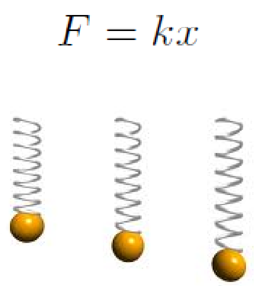
Stress, Strain, and Pressure

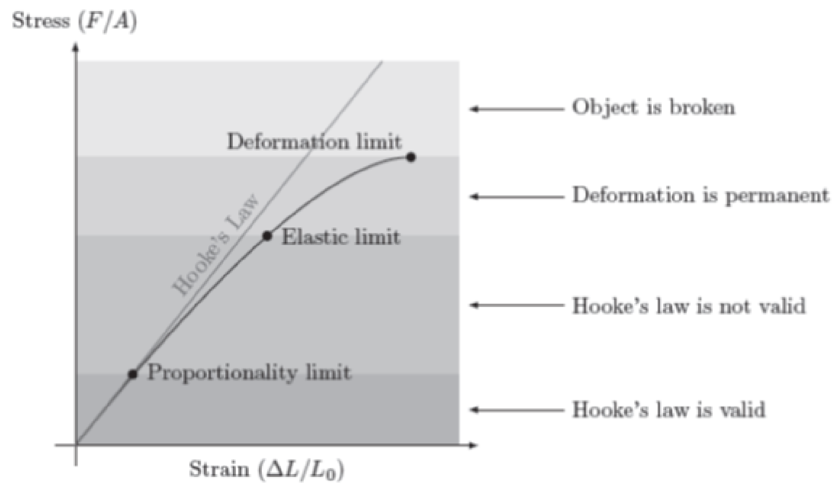
- In this term we move beyond the simple particle mechanics of the first term.
- As we examine more complex physical systems, we quickly reach a point where we cannot derive every detail exactly. But we will do our best.
- We start with a discussion of the properties of matter: solids, liquids, gases—then heat.
- Half-way through the term we switch gears and the theme for the rest of the term is wave motion with applications to sound and light.
- Light presents peculiar problems.... But we will talk about that when we get there.

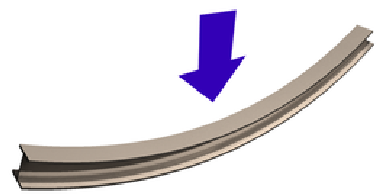
The Phases of Matter

- For a moment, forget everything you've learned so far in class.
- When we look around, I think the most obvious physical characteristic we can use to distinguish material things is their phase: whether it's a solid, liquid or gas.
- We know that at a molecular level, the phase of a substance is determined by the strength of its inter-molecular forces relative to its temperature.
- Traditionally one says:
  - Solids retain their shape and volume
  - Liquids take the shape of their container but retain their volume
  - Gases expand in volume to fill a container
- From a more mechanical perspective, this distinction shows up in that each responds differently to force...
  - Solids deform (or change shape)
  - Fluids flow
- In this way, liquids and gases are similar. When we speak of a **fluid**, we are referring to either one.
- We will discuss fluid flow in the next lecture. For now we focus on solids.

Elasticity: Stress and Strain







- One common picture of the molecules in a solid is a bunch of balls connected by springs. This model is more appropriate than it may first appear.
- The reason is that a spring is an ideal elastic system. Any system in equilibrium will act similar to a spring—including the molecules in a solid.
- A spring that is displaced from equilibrium will experience a force that pulls it back to equilibrium. The larger the displacement, the larger the restorative force.
- An **ideal spring** responds with a force that is directly proportional to this displacement:

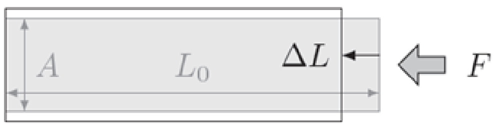
  $$F = kx$$

  where k is called the **spring constant**. This formula is sometimes called **Hooke's Law**.
- Any system in equilibrium will respond in this linear fashion for a certain range of (small) displacement. This will become important again when we begin to discuss wave motion later in the term.
- In particular, a solid object will also obey Hooke's Law. We write:

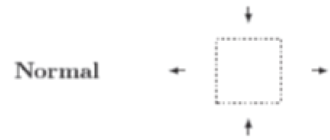
  $$\frac{F}{A} = Y\left(\frac{\Delta L}{L_0}\right)$$

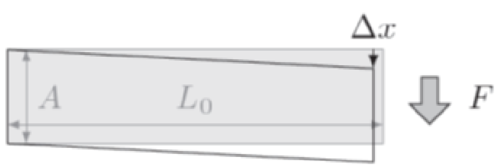
  where L_0 represents the original length of the solid, ΔL its change in length, and Y is called **Young's modulus** which is characteristic of the material being stretched or compressed.
- In this formula, the F/A is called the **stress** on the object, and $\Delta L/L_0$ is called its **strain**. Note that strain is a ratio: so it has no unit.

The Three Components of Stress

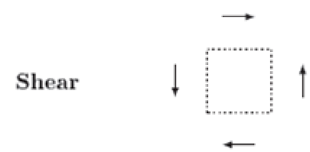


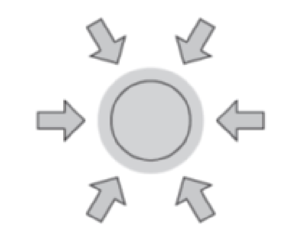
$$\frac{F}{A} = Y\left(\frac{\Delta L}{L_0}\right)$$



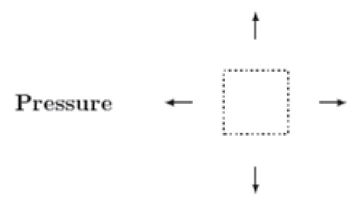


$$\frac{F}{A} = S\left(\frac{\Delta x}{L_0}\right)$$





$$\Delta P = -B\left(\frac{\Delta V}{V_0}\right)$$



- Consider a person standing on the edge of a diving board. The deflection in the board is caused by a force that is applied perpendicular to the linear dimension of the board.
- This type of stress is called **shear stress** while stretching or compression is sometimes called **normal stress**.
- For a small deflection :math‘Delta x‘, shear stress will also obey a formula similar to Hooke's Law:

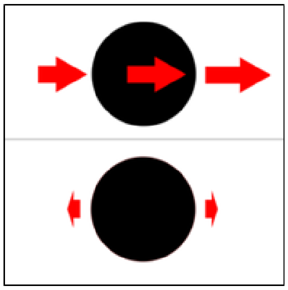
  $$\frac{F}{A} = S\left(\frac{\Delta x}{L_0}\right)$$

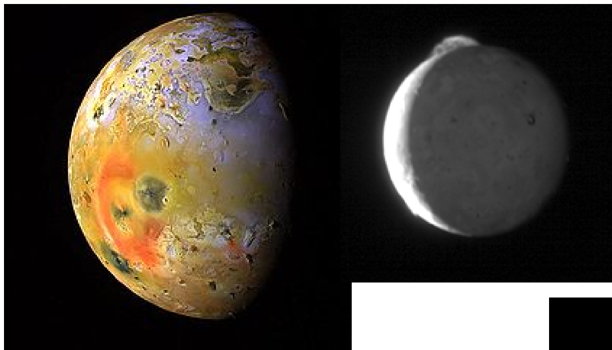
  where S is called the **shear modulus** for the material.
- Remember that the Δx is perpendicular to the length L_0 and parallel to the cross-section A. In general, the strain in an object measures the relative displacement of the parts of the system.
- Clearly any force on a long bar can be broken into shear and normal components. When we consider a three-dimensional object, we also need to include the **pressure** surrounding an object. Again, we have a version of Hooke's Law:

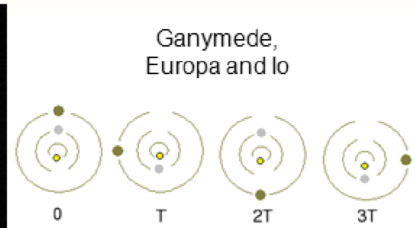
  $$\Delta P = -B\left(\frac{\Delta V}{V_0}\right)$$

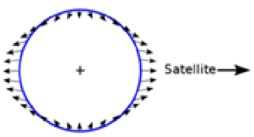
  in this case B stands for **bulk modulus** and is negative because and increase in pressure corresponds to a decrease in volume.
- It can be shown that these three types of stress are all that is necessary to describe an arbitrary set of (small) forces in applied to an object. In this way, we can characterize stress by three simple numbers. This method of analysis is called **continuum mechanics**. It applies in the region between the macroscopic and the molecular scale.

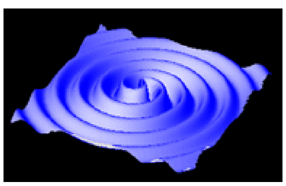
Tides, Gravity Waves, and the Solar System

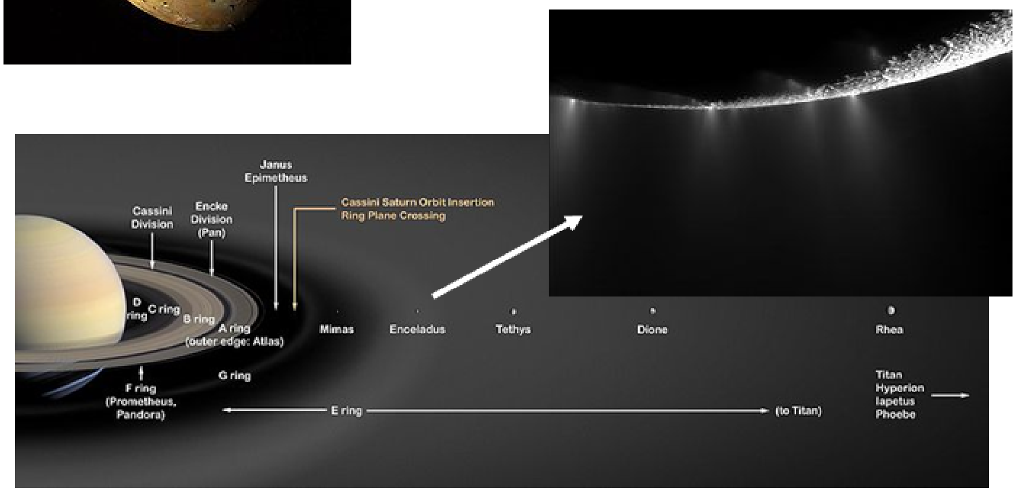






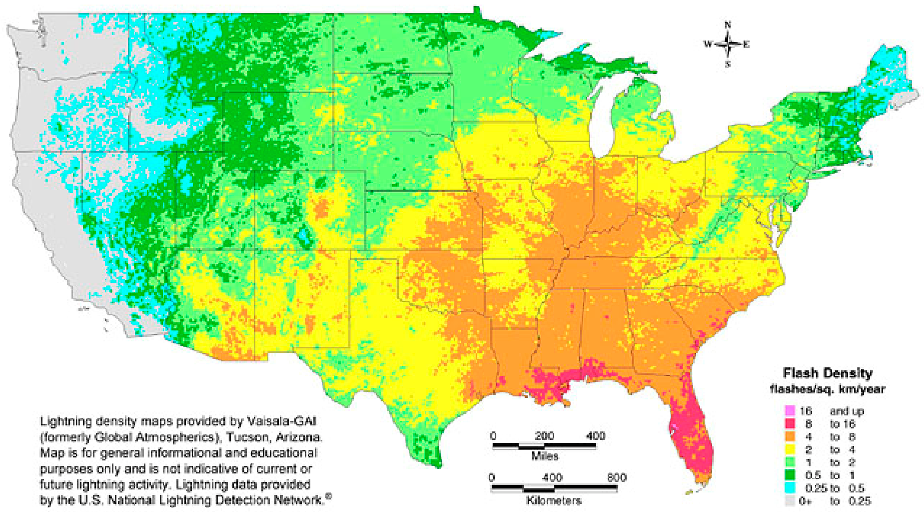






- One unusual example of stress is that on an astronomic scale. Because Newton's law of gravity is not uniform, gravitational differentials will exert stress on a body large enough to notice—like the ocean tides. Newton was able to show how the tides are a result of the Moon's gravitational stress on the Earth.
- According to general relativity, the fingerprint of gravitational influence is normal stresses like the ocean tides (no shear stress or net pressure). Gravitational radiation, when discovered, will literally be "tidal waves" of space-time curvature.
- Of course, the Earth also affects the Moon—the Moon itself bulges. But since the Moon's surface is solid, its rotation will twist this bulge. The gravitational field from the Earth actually torques against this twist—in effect "locking" the rotation of the Moon into one revolution per orbit. This is why we only see one side of the Moon.
- Many other satellites are locked to their parents (including Mercury to the Sun). One of the most dramatic is Io. Not only it this moon locked to Jupiter but its orbit is fairly eccentric. As a consequence gravitational stresses are constantly squeezing the planet back and forth as it travels along its orbit like a huge stress ball.
- This squeezing generates heat which normally would dissipate into space, lowering its orbital energy, thereby reducing its eccentricity. But the eccentricity of Io is also driven by its neighbors Europa and Ganymede. This three-body resonance makes Io the most volcanically active body in the Solar System.
- Another object of interest is Saturn's moon Enceladus. Recently some have said it "is emerging as the most habitable spot beyond Earth in the Solar System for life as we know it." Why? Liquid water.
- In July 14, 2005 the Cassini spacecraft actually flew through a geyser of water spewing from the satellite. The source of this explosive volcanism is believed to be a combination of tidal stress and radioactivity (which is the Earth's source of geological activity).

Intrinsic and Extrinsic Properties

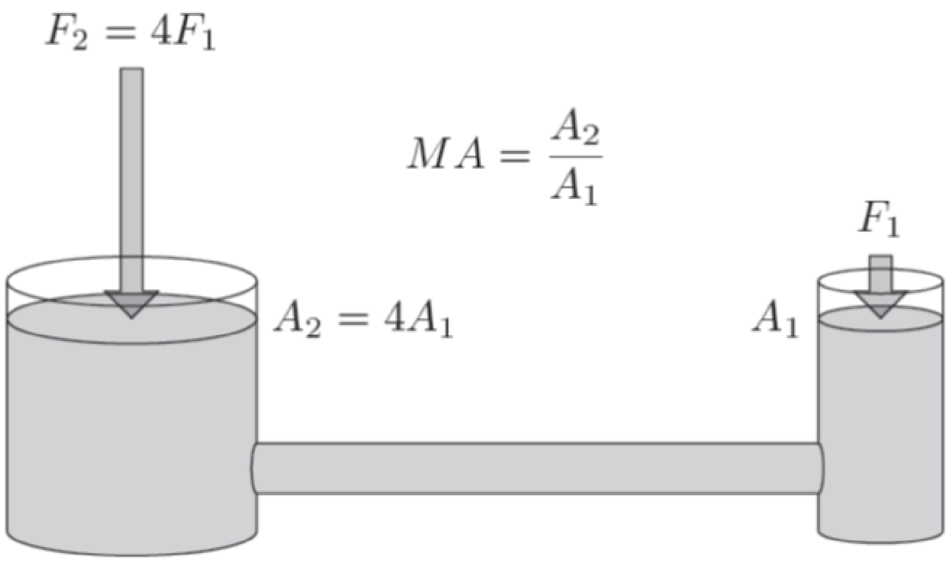


- Having spoken about elasticity, we now move on to the other phases of matter: liquids and gases. Though fluids are characterized by their ability to flow, we will table that discussion for next lecture. For now, we will wrap up this lecture talking about fluids at rest.
- Because fluids do not have definite shapes, it is often convenient to characterize them by their **intrinsic properties**—in particular, density.
- The difference between intrinsic and extrinsic properties is the difference between the density and the mass of an object. Essentially, we divide the extrinsic property by volume to arrive at the intrinsic property.
- Intrinsic properties are characteristic of the *substance* rather than the object. In this way, we separate issues related to substance from the issues related to the overall motion of the system.
- Typically, intrinsic properties are defined as the ratio of two extrinsic properties. Thus, density is

$$\rho = M/V$$

- Frequently, an intrinsic property will have the word "density" in it: e.g., energy density, number density, charge density, etc.
- Usually one imagines an intrinsic property as associated with a particular place in the system—with potentially different values at different places. Mathematically, we say that it is a function across the space.

Fluids and Pascal's Principle



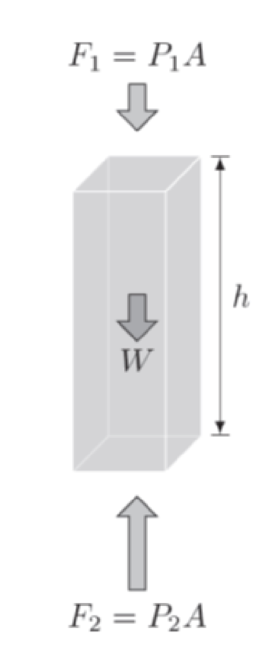
- Ideal fluids cannot support any kind of normal or shear stress. Real fluids deviate from this ideal through viscosity and other ways (e.g., non-Newtonian fluids). But in the ideal case, the only kind of stress a fluid can support is pressure.
- If the fluid is at rest, it cannot have any pressure differentials, because any differential will cause the fluid to flow.
- From this follows **Pascal's principle** which states that if the pressure in a fluid is increased anywhere, it increases everywhere.
- This is the principle behind the hydraulic press (invented by Pascal). Remember that pressure is the perpendicular force per unit area

  $$P = F/A$$

  (so it is an intrinsic property). If two cylinders of fluid are connected to one another, the pressures must be equal.
- If a force is applied to one of the cylinders, the incremental pressure is transmitted to the second cylinder and the force is potentially multiplied based on the cross-sections of the cylinders.
- The mechanical advantage of an ideal hydraulic press is simply

  $$MA = A_2/A_1$$

Buoyancy: Archimedes' Principle



$$\Delta P = \rho g h$$

$$B = \rho V g$$

- Hydrostatics is one of the few branches of physics that can be traced back to ancient times. (Others are astronomy, optics, and statics.) The last great before the Dark Ages in Europe was the physicist/mathematician Archimedes.
- Archimedes is credited with understanding the principle of the lever, building optical "death rays", and discovering a rudimentary form of calculus. He is also one of the few (only?) ancient persons with a physical principle in modern use named after him.
- **Archimedes' principle** is states that the force of buoyancy in a fluid is equal to the weight of fluid that is displaced by the submerged object.
- This follows from the hydrostatic equation

  $$\Delta P = \rho g h$$

  where ΔP represents the pressure differential due to the increased fluid depth h.
- You can see that I lied earlier when I said that a static fluid cannot support a pressure differential. It must if it is subject to an all-pervasive force like gravity. This is simply due to the fact that the fluid below must support the the weight of the fluid that is above with increased pressure. The hydrostatic equation follows.
- And this is why Archimedes' principle works. A submerged object is subject to the fluid pressure that surrounds it—this pressure is in equilibrium with the fluid that would be there if the object were absent. Since the total force is pressure multiplied by surface area, the force of buoyancy must be

  $$B = (\Delta P)(A) = \rho g V$$

  which is the weight of the displaced fluid.

Atmospheric Pressure

- Both Pascal's and Archimedes' principles apply to gases too. Atmospheric pressure (at sea level) is about 101.3 kilopascals. (The SI unit for pressure is the pascal: one newton per meter squared.)
- It is easy to forget about atmospheric pressure because it's everywhere. But without it helium balloons would fall to the ground, soda straws would not work, and we wouldn't have weather. Then again, you would not have to suffer your ears popping when you change elevation either.
- In our everyday life, atmospheric pressure is practically constant. As a consequence most pressure gauges are calibrated to ignore it. Your tire gauge is actually measuring the pressure *differential* between the tire and the atmosphere.
- So be careful when solving homework problems: you may need to add 101.3 kilopascals to this **gauge pressure** in order to get the absolute pressure of the gas.
- The density of a gas is related to its pressure. The hydrostatic equation gives us a pressure differential, but in a gas this differential is proportional to the absolute pressure (through its density). Mathematically, this implies that the pressure will depend exponentially on the altitude. The formula is

  $$P = P_0 \exp(-kz)$$

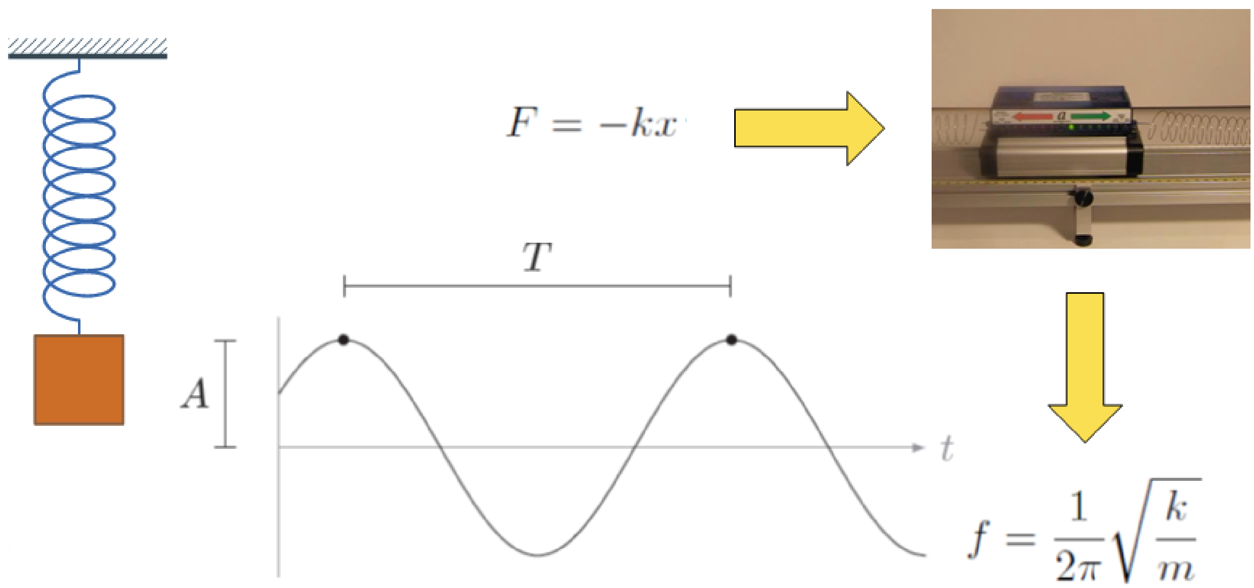
  where k is related to the molar mass and temperature of the gas (which affect the density of the gas).
- In addition, pressure affects the boiling point of liquids, which is why recipes often need to be adjusted for high elevations. Early explorers used the boiling point of water as a rough estimate of their altitude.

Physics 202 Lecture 2

Vibration and Harmonic Motion

- For the rest of the term, we will be discussing the nature of waves: either explicitly or indirectly through the lectures on light.
- In order for a wave to exist we need to things: a source and a medium. The source creates the initial disturbance and the medium propagates that disturbance through space.
- Today we start with the source. The waves source is a vibrating element of the medium. This is why we talk about why and how things vibrate.
- Of course, the internal motion of a vibrating object can be quite complicated. In this lecture we talk mostly about the vibration of a particle—which of course has no internal vibration.

The Simple Harmonic Oscillator



$$\psi(t) = A\cos(2\pi f t + \phi)$$

- Remember that as ideal spring involves a restorative force that is proportional to displacement: $F = -kx$.
- When we displace the system from equilibrium, this force act and the system accelerates back. But once the system reaches equilibrium, its inertia will cause it to "overshoot" creating another displacement in the opposite direction.
- We end up with a back-and-forth oscillation we call **simple harmonic motion**.
- A system can vibrate along any of its degrees of freedom, whether that is a spring, a string, the surface of a drum, a metal beam, a column of air, or the electromagnetic field.
- Mathematically, the form of this motion is

$$\psi(t) = A\cos(\omega t + \phi)$$

- This motion involves three parameters:
  - **Amplitude** (A): The maximum displacement, related to energy. We will see later that

$$E = \tfrac{1}{2}kA^2$$

  - **Frequency** (f): The number of vibrations per second, related to the properties of the system. We will see later that

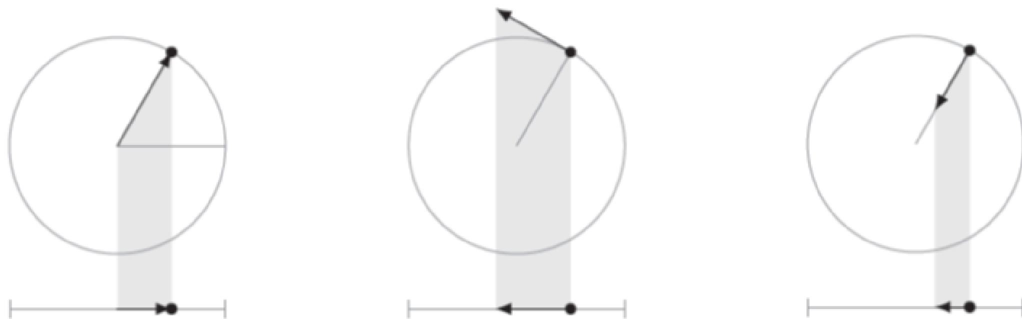
$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$$

  - **Phase shift** (ϕ): This will be important later when we discuss wave motion. We will see later that

$$\phi = \frac{2\pi}{\lambda}x$$

A Fortunate Fluke

$$r = A \qquad v = \frac{2\pi r}{T} = A\omega \qquad a = \frac{v^2}{r} = A\omega^2$$
$$\theta = \omega t$$

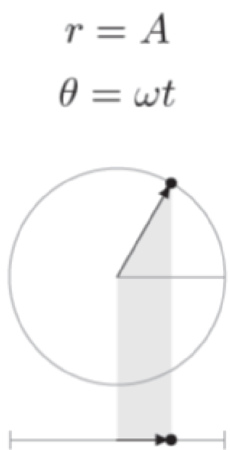


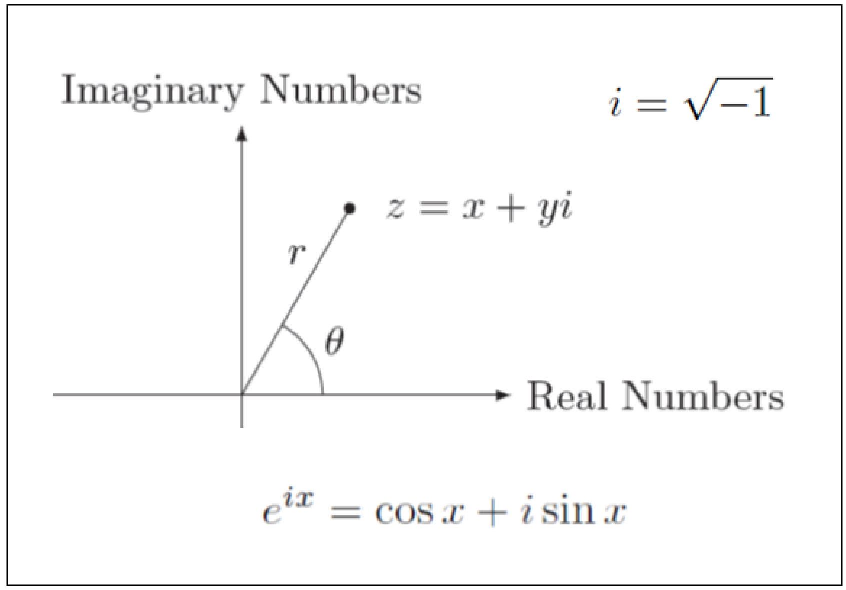
$$\omega = 2\pi f$$

$$\psi(t) = A\cos(\omega t + \phi)$$

- It so happens that the components of uniform circular motion obey simple harmonic motion. Remember that the centripetal force is $F = -mv^2/r = -mr\omega^2$ where ω is the constant angular speed. If we just look at the x-component, we have
$$F_x = -(m\omega^2)(x)$$
- This not only establishes the result, but gives us a relationship for frequency:
$$k = m\omega^2 \implies \omega = \sqrt{k/m}$$
- Since ω is radians per second and the oscillation repeats every 2π radians, we have $\omega = 2\pi f$ and
$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$$
- Furthermore, we can apply our knowledge of uniform circular motion to understand simple harmonic motion. The circle associated with our motion has a radius of A and the angle is given by ωt.
- We know that the speed of uniform circular motion is $v = 2\pi r/T$. Combining this with $f = 1/T$ and $\omega = 2\pi f$, we can also say
$$v = A\omega$$
- The velocity of the simple harmonic motion is only the projection of this velocity, so this represents the *maximum* speed as the system passes through equilibrium.
- The acceleration of the circular motion is $a = v^2/r$ which can be rewritten as
$$a = A\omega^2$$
- Again, this represents the maximum acceleration which is experienced by the system at maximum displacement.

Spinning Arrows and Complex Numbers





$$\psi(t) = \mathrm{Re}(Ae^{i\omega t})$$

- The association of uniform circular motion with simple harmonic motion is more than just a trick. It is a mathematical technique called **phasors**. They greatly simplify the mathematics when combining vibrations from different sources is required. We will see them again when we study AC electronics.
- The only bad thing about phasors is that they are complex numbers: a combination of a real number and an imaginary number. An **imaginary number** is the square root of a negative real number. The fundamental one is

  $$i = \sqrt{-1}$$

  called the imaginary unit. Every other imaginary number can be written with the imaginary unit. For example, the square root of -9 is $3i$.
- A **complex number** is usually written as

  $$z = a + bi$$

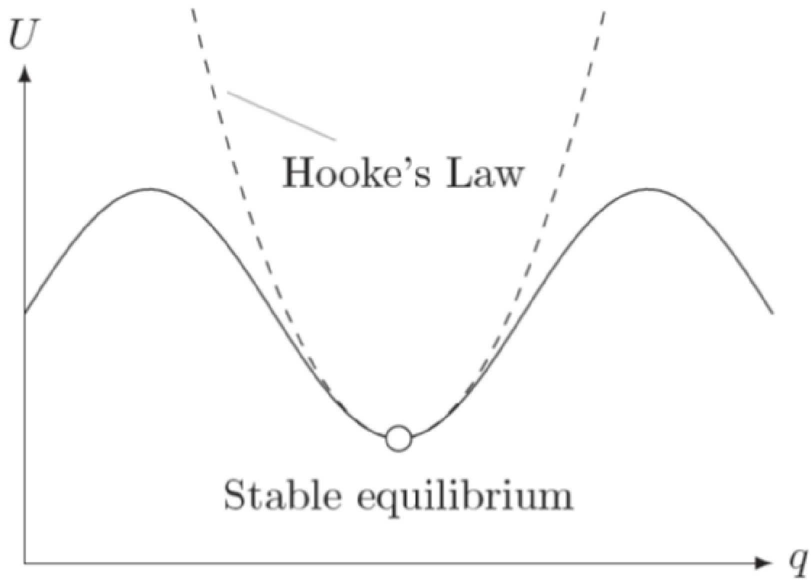
  where a is the real part and b is the imaginary part.
- Complex numbers act a lot like two-dimensional vectors. We add them by component. But, unlike vectors, we can also multiply them. Treat the i like a variable, FOIL the product, and remember that $i^2 = -1$.
- We can also exponentiate with imaginary numbers. Feynman calls this Euler's jewel:

  $$e^{ix} = \cos x + i \sin x$$

- Any complex number can be rewritten in its so-called "polar" form $z = re^{i\theta}$ using Euler's formula. Multiplying complex numbers is easiest in this form because we multiply the magnitudes r and add the angles θ.
- So we can represent simple harmonic motion as the real component of a "spinning arrow" (ala Feynman):

  $$\psi(t) = \mathrm{Re}(Ae^{i\omega t})$$

Energy in Oscillation



$$PE = \frac{1}{2}kx^2$$

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

- Hooke's law is conservative and the potential energy associated with it is

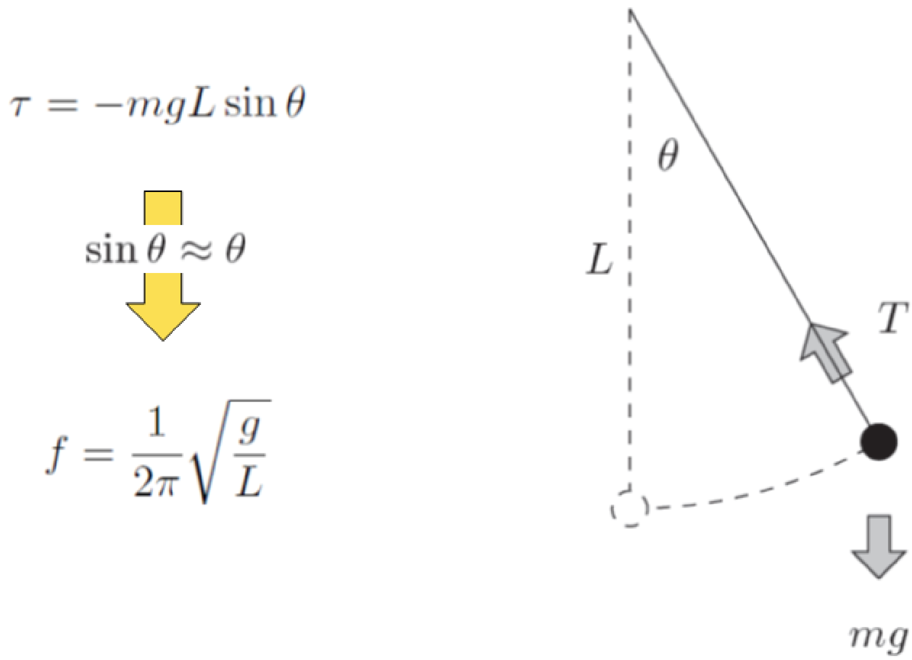
$$PE = \tfrac{1}{2}kx^2$$

- In general, the total energy of the ideal spring system is the sum of the kinetic energy of the mass and the potential energy in the spring.
- But when the system is at maximum extension the velocity of the mass is zero (for a moment). Since $x = A$ at this point, we can write

$$E_{\text{tot}} = \tfrac{1}{2}kA^2$$

- The fact that the energy of vibration is proportional to the square of the amplitude of the vibration explains why laser light is so much more powerful than regular light.
- Hooke's law is an idealization for a real spring, and even more so for complex physical systems. However, the forces maintaining any state of stable equilibrium are approximately linear for small displacements.
- One way to see this is by investigating the energy diagram. For a system to be in stable equilibrium, the potential energy curve must create a trough or potential well. The point of equilibrium is at the bottom of this well.
- We fix the vertex of a parabola to this point and fit its curvature to the curvature of the potential energy function. In this way any potential well can be approximated by a parabola associated with some version of Hooke's law.
- The project for this term uses this trick to calculate an estimate for the anomalous precession of Mercury due to general relativity.

Pendulum Motion



- An important example of a non-linear system is the pendulum. In this case we consider our displacement to be the angle of the pendulum from vertical. The weight of the pendulum bob creates a torque that has a tendency to push the pendulum back to center.
- The tension from the string has no lever arm, so it introduces no torque into the system. The tangential component of the weight is simply $-mg\sin\theta$, so the total torque on the pendulum is

$$\tau = -mgL\sin\theta$$

- Newton's second law for rotation states that this will cause an angular acceleration ($I = mL^2$):

$$-mgL\sin\theta = mL^2\alpha$$

- Now if the angle involved is small, we can use the approximation that $\sin\theta \approx \theta$ (when θ is measured in radians).
  - Small means less than 10°, or less than 0.1 radian. This is the point in the derivation where we replace the natural potential energy function with the ideal parabola for simple harmonic motion.
- We can also absorb an L on both sides to convert the angles into tangential variables. We get

$$-gs = La$$

  where s represents the arc-length of the pendulum's swing. This equation is similar to $-kx = ma$ for an ideal spring. So, we can immediately write the frequency for the pendulum swing as

$$f = \frac{1}{2\pi}\sqrt{\frac{g}{L}}$$

- Remember, this equation is only valid when the angular displacement is small.

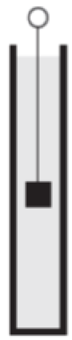
Damped Harmonic Motion

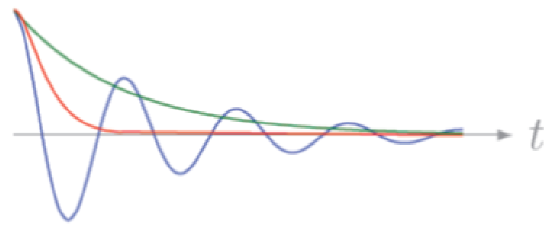
$$F_{\text{drag}} = -cv$$

$$\gamma = c/m.$$

$$A(t) = A_0 \exp(-\tfrac{1}{2}\gamma t)$$

$$E(t) = E_0 \exp(-\gamma t)$$





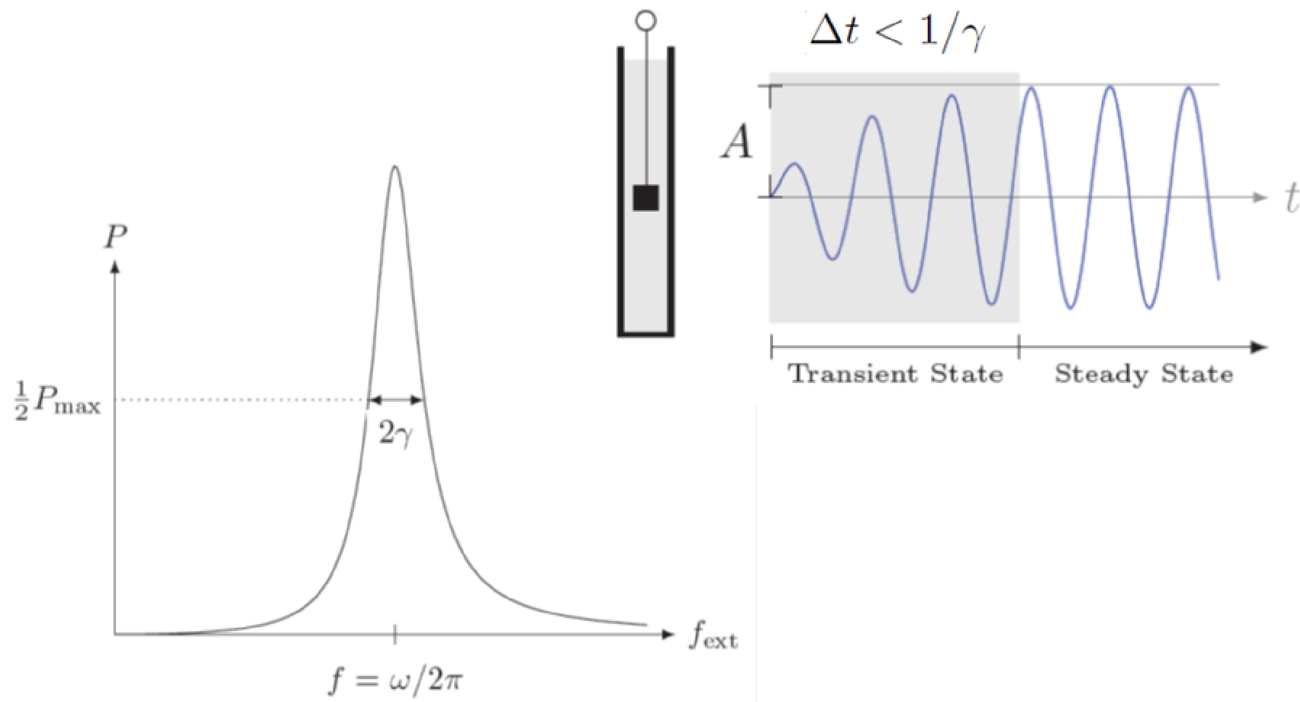
Light damping $(\gamma < 2\omega)$

Heavy damping $(\gamma > 2\omega)$

Critical damping $(\gamma = 2\omega)$

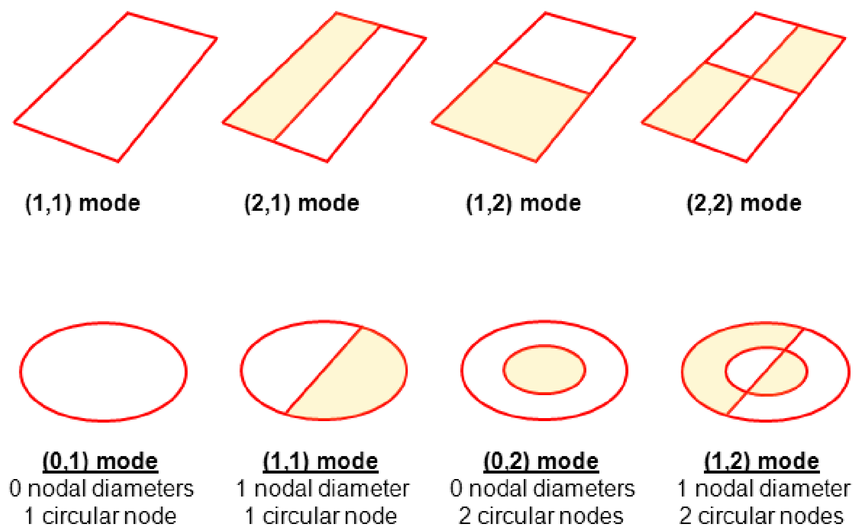
- The ideal spring is a good model for any vibration. We can improve the model by adding an element of friction to the system. There are different ways to do this, but one very common way is by introducing a drag term which is proportional to the velocity of the system:
$$F_{\text{drag}} = -cv$$
where c is the "stiffness", or viscosity of the drag.
- This is the type of friction an object experiences as it pushes slowly through a fluid. The shock absorbers in your car use this kind of a set up to minimize vibration. This damping effect introduces a non-conservative force into the system which will reduce the amplitude of the motion over time.
- Since the amplitude is related to the energy of this system, this reflects the fact that energy is being lost through friction. The equation for the amplitude and energy is
$$A(t) = A_0 \exp(-\tfrac{1}{2}\gamma t) \qquad \text{and} \qquad E(t) = E_0 \exp(-\gamma t)$$
where $\gamma = c/m$.
- The displacement of the damped oscillator over time is simple harmonic motion with a continually decreasing amplitude.
  - When $\gamma = 2\omega$, the system is said to be **critically damped**. This is the circumstance which brings the system back to equilibrium the fastest.
  - If the system is **under-damped** ($\gamma < 2\omega$), then it will oscillate a bit before coming to rest, as described above.
  - If the system is **over-damped** ($\gamma > 2\omega$), then the viscosity will be so thick it will actually drag against the system being brought to rest.

Driven Harmonic Motion and Resonance



- We have one last topic to consider for oscillation in one dimension: forced vibrations. If a damped system is driven by an external vibration, the motion of the system will eventually match the frequency of the driver.
- Initially it will move through a temporary ($\Delta t < 1/\gamma$) **transient state** until it reaches a **steady state** of sinusoidal motion. This steady state will have an amplitude which will depend upon both the frequency of the external driver and the natural frequency of the system.
- What happens is that the external vibration does work on the system when the force is aligned with the displacement of the system. When the frequency of the external vibration is matched with the natural frequency of the system, the work done in each cycle accumulates. It does not take long for system to absorb a lot of energy from the external source even if the magnitude of the source is small.
- Eventually the system is moving fast enough that the damping force (proportional to this speed) increases to drain the energy that is being absorbed. This steady state is called **resonance**.
- The peak power absorption occurs at the natural frequency of the system. The width of the peak is related to the "stiffness" of the damping: at one-half of the maximum power absorption it is equal to 2γ.
- Resonance can be used to explain the scattering of light (why the sky is blue), how microwaves work, and is important in certain electrical circuits.
- The **quality factor** of an electrical circuit is defined as the ratio of the resonant frequency and the half-maximum width of the power absorption curve. A sharp peak will make a good radio receiver as it will only react to a particular range of radio frequencies.

Vibration of Extended Objects



http://www.kettering.edu/~drussell/demos.html

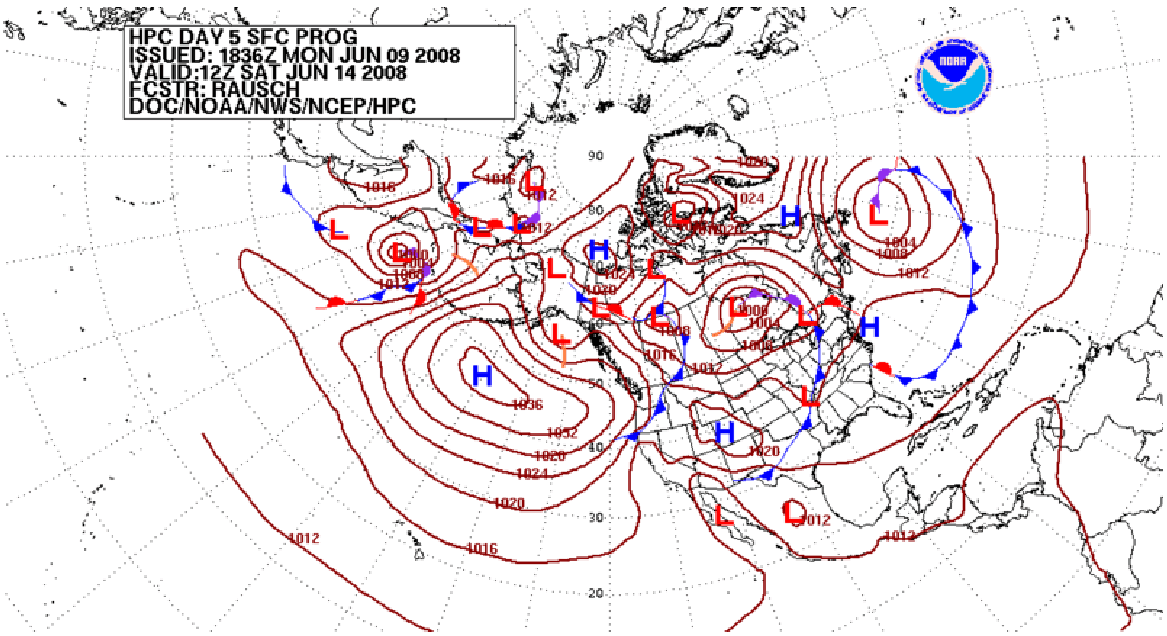
- A one-dimensional oscillator can only vibrate in one way: back-and-forth with its natural frequency.
- An extended object has many degrees of freedom in which to vibrate. In general, a displacement (like giving it a good whack) will activate many of these degrees of freedom.
- A single vibrational degree of freedom is called vibrational **mode**. The vibrational modes of an object are controlled by its shape.
- For each mode, there are certain spots, lines or planes in which there is no motion—these are called **nodes** and define the mode.
- For example, on a circular drum the modes are defined by the number of rings and radial lines which are nodes. In general, the vibrational modes of an object are identified by one, two, or three integers depending on whether the object is one, two, or three dimensional.
- With each mode corresponds a particular natural frequency of vibration. The pattern of these natural frequencies can be used to analyze the composition of an object.
- When a string is plucked or a surface is struck is that many modes are activated. All are damped, but some more than others.
- The ones with the smallest damping constants remain and gives the object a characteristic "ring". Every musical instrument works this way.

Physics 202 Lecture 3

Fluid Flow

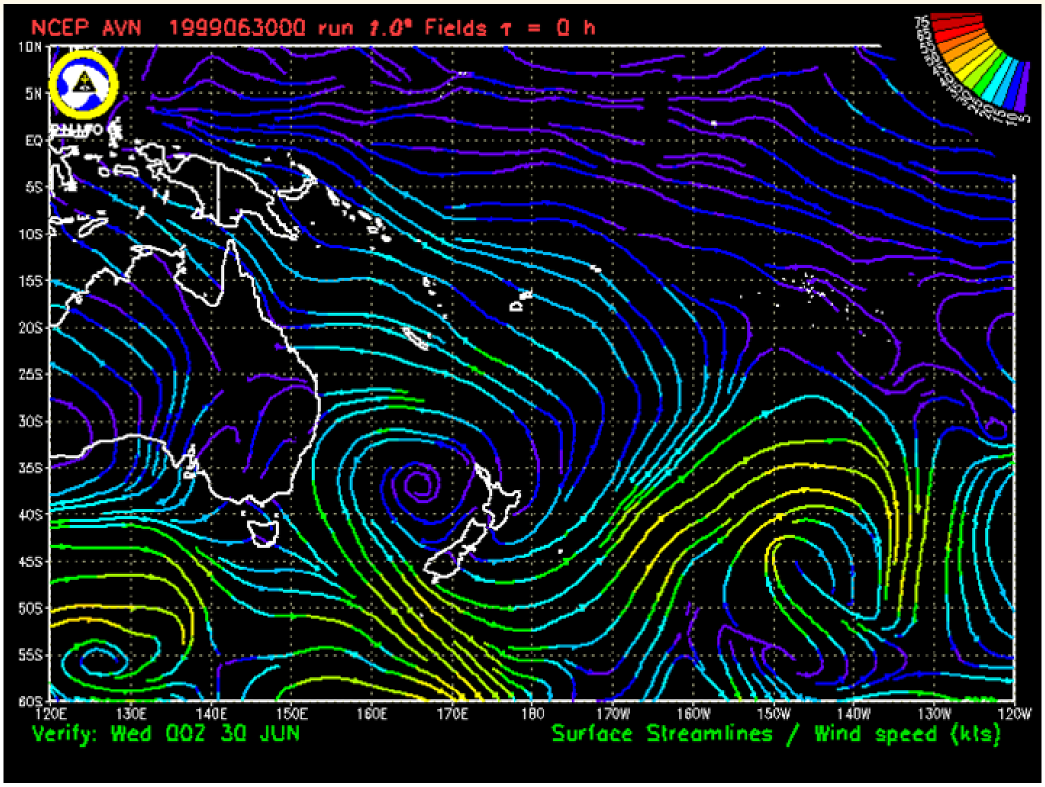
- In this lecture we will cover topics related to the movement of liquids and gases.
- The goal is to characterize the relative motion of the parts of the fluid over time and space. With solids we had merely to calculate the relative position of the parts—with flow, we need a more complicated model.
- The mathematics of fluid flow is some of the most difficult in physics—clearly we won't be able to do more than touch the surface of the subject. But we will do what we can do.
- Later this term we will discuss radiation and wave motion and radiation. These topics involve the flow of *energy*. In this lecture we discuss the flow of matter, but the parts of the mathematical analysis of how things flow is similar for both cases.
- In addition, some of the concepts will come up again in the context of field theory at the beginning of the third term when we talk about electromagnetism.

Pressure Differentials Cause Flow



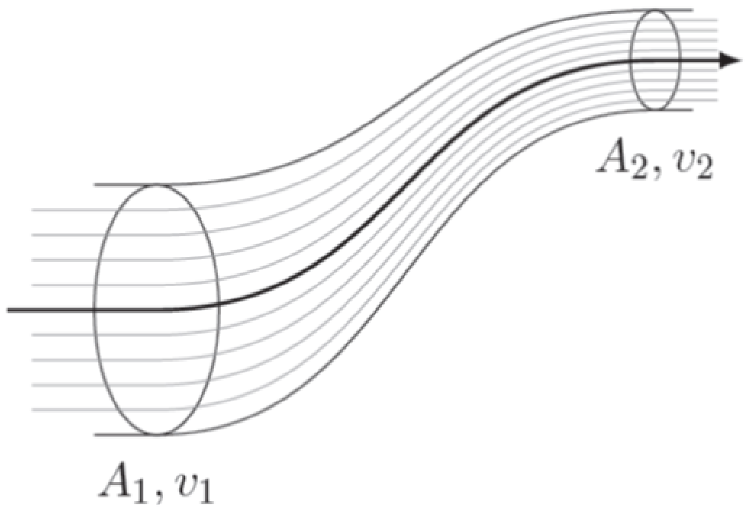
- In order to start, we would like to begin with some idealized form of fluid. Unlike solids, you will see there are many levels to our idealization.
- First of all we must decide whether electromagnetic effects are important or not. The most important application for **magnetohydrodynamics** is in plasma physics, which has many important industrial and astronomical applications. We will assume our fluids are electrically neutral.
- The second level of consideration is **viscosity**, or how "thick" the fluid is. We will only touch viscous fluids superficially in this class—because the math is extremely complicated.
  - Viscous fluids divide themselves into Newtonian and non-Newtonian fluids. Non-Newtonian fluids exhibit extremely odd behavior, defying our intuition about what fluids "ought" to do.
  - Unfortunately, by eliminating viscosity from our models, we leave out an *essential* component of the fluid. John von Neumann called this the study of "dry" water. Nonetheless, we will do so and see how far we get.
- Finally, we have one more consideration: is the fluid **compressible** or not? Most liquids are incompressible, but even gases may be considered incompressible for our purposes. The key is whether the density of the fluid changes significantly with the flow. If not, the fluid is considered incompressible.
  - Another equivalent condition is that the fluid speed if much less than the speed of sound in the fluid (less than 30%). One consequence of this condition is that we will be leaving out the ability to study sound waves or shock waves.
- Thus, we have an **ideal fluid**: an electrically neutral, inviscid, and incompressible fluid.
  - You may wonder about thermodynamic considerations. In general, the pressure and density of a fluid determine its temperature at any point. This is called the **equation of state** for the substance. Since we will only study incompressible fluids, we can ignore this issue.

Streamlines and Laminar Flow



- Perhaps the most obvious way to analyze the motion of a fluid is to model each little "piece" of the fluid as a particle. This can be done using Lagrangian mechanics, but a more common approach goes back to Euler.
- In Euler's method, the pattern of the fluid flow is primary. The fluid is defined mathematically by functions across space and time. In particular, a velocity field is defined for each place in the fluid.
- For example, we define **steady flow** as the fluid state for which the velocity field does not change with time. Fluid elements are constantly flowing through these points, but they are consistently replaced in a way that the overall flow pattern does not change.
- If the flow is steady, we can map out the flow pattern using **streamlines** where we trace out the velocity field lines in a kind of sophisticated connect-the-dots. Note that the streamlines represent a snapshot in time—the fluid elements do not follow these lines if the flow is not steady.
- Steady or not, another characteristic of the velocity field is whether it is **irrotational**. The simplest way to define this is whether or not the fluid carries angular momentum. Important examples of rotational fluid flow includes
  - Whirlpools, tornadoes, and smoke rings
  - Coriolis force in weather patterns (hurricanes)
- We define the **circulation** of the velocity field by calculating the sum of the component of the field around a closed loop. If this value is zero, the fluid is irrotational and there is no angular momentum in the fluid at that point.
  - Mathematically, no circulation implies a "potential" velocity field something like the way a conservative force implies a potential energy. If the flow is steady, this potential is proportional to the pressure in the fluid.
  - No circulation does not mean no rotation at all. The steady flow pattern around a curve ball is a combination of straight steady flow and a circular flow around the ball ($v \propto 1/r$).

Flow Rates, Current, and Continuity



$$Q = Av$$

$$\Delta m/\Delta t = \rho v A$$

$$\rho A_1 v_1 = \rho A_2 v_2 \implies A_1 v_1 = A_2 v_2$$

- The **mass flow rate** is the amount of fluid that crosses a particular surface area per second.
  - For incompressible fluids, the density is constant. So the mass of the fluid is directly related to its volume. In that case we may speak of a "volume flow rate." For some reason, this has acquired the symbol Q, but the more fundamental mass flow rate has not.
  - The mass flow rate is also known as the **flux** of fluid across the surface. The fluid **current** is the flux density, or fluid flux divided by the area involved.
- It's pretty easy to show that the flow rates are related to the speed of the fluid according to

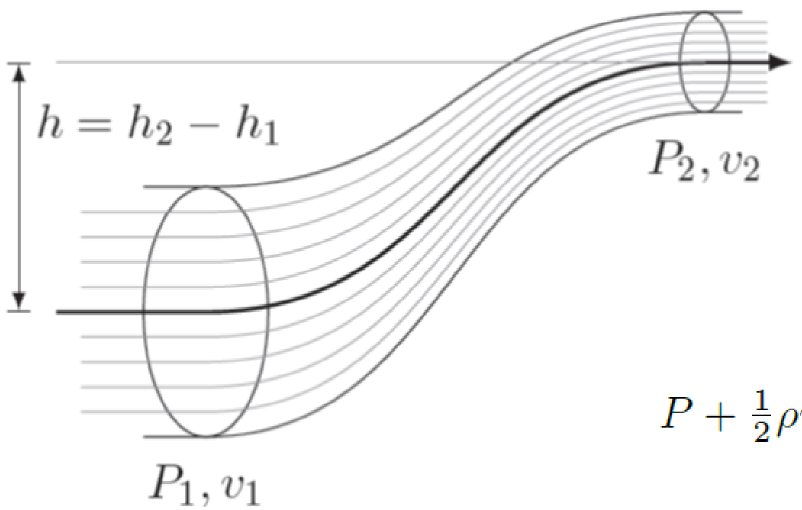
$$\Delta m/\Delta t = \rho v A \qquad \text{and} \qquad Q = Av$$

- The **equation of continuity** is the recognition that the total mass flow rate (or total flux) through the boundary of a volume changes the amount of fluid within the volume by just that amount. The fluid cannot "teleport" into or out of the volume.
  - Consider the example of fluid flowing through a pipe and mark out two different cross sections. These cross sections enclose a volume of the fluid.
  - For an incompressible fluid, the mass flow rate must be zero for the constant volume. Therefore, the fluid flux flowing in one side must equal to the fluid flux flowing out the other (the "traffic jam" equation):

$$\rho A_1 v_1 = \rho A_2 v_2 \implies A_1 v_1 = A_2 v_2$$

- Continuity forces the streamlines to spread evenly through any change in volume. Fluid "sources" will cause the streamlines to diverge, and fluid "sinks" will cause them to converge. In this way sources, sinks and boundaries completely determine the streamline pattern for an steady ideal irrotational fluid.

Bernoulli's Equation



$$P + \frac{1}{2}\rho v^2 + \rho g h = \text{const.}$$

- So far we have really only discussed the velocity flow pattern of the fluid. We have not discussed pressure which is driving the fluid flow. (Using technical jargon: we have only described the kinematics of fluid flow.)
- Viscosity is how friction manifests itself in a fluid. Because we have agreed to exclude viscosity from our discussion, we can assume that there is no loss in energy due to internal friction.
- **Bernoulli's equation** is essentially the application of the work-energy theorem to our ideal fluid. We have

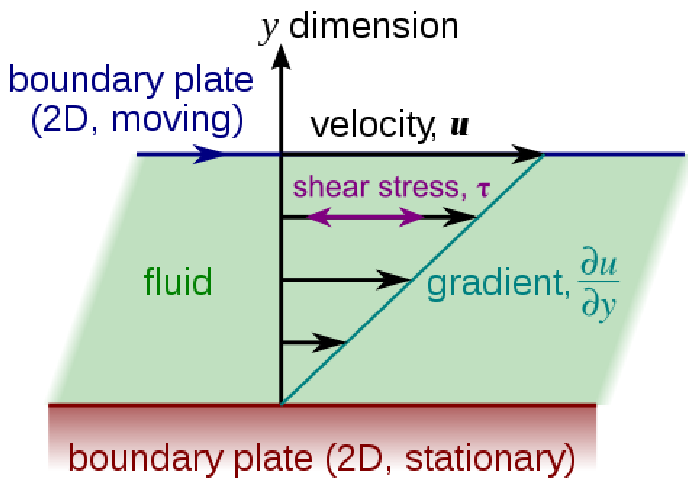
  $$P + \tfrac{1}{2}\rho v^2 + \rho g h = \text{const.}$$

- The first term (pressure) corresponds to work, the second term corresponds to kinetic energy, and the third term corresponds to potential energy.
- There are two special cases of this equation to consider. The first is a static fluid: $v = 0$. In that case, the middle term drops out and we are left with the hydrostatic equation we discussed last lecture.
- The second case is more interesting. Suppose there is negligible height differential between two points in the fluid flow so that the third term in Bernoulli's equation can be ignored. We have

  $$P + \tfrac{1}{2}\rho v^2 = \text{const.}$$

  also known as **Bernoulli's principle**: high speed means low pressure.
- This is the principle behind many surprising fluid behaviors like low pressure hurricanes, curve-balls, the lift on an airplane wing, and plumbing vents.

Beyond Bernoulli: Viscosity



$$F = \eta A v / d \qquad F = 6\pi\eta R v \qquad \Delta P = \frac{8\eta L Q}{\pi r^4}$$

- Here is a surprising experimental fact about fluids: the fluid directly adjacent to the boundary is always at rest. Nothing we have done so far allows us to anticipate this. But why is there always a fine layer of dust on the blades of a fan? You can never blow all the dust off a table!
- Consider a fluid trapped between two parallel plates with cross-section A separated by a distance d. Fix the bottom one and drag the top one at a constant velocity. The force required to maintain this speed is given by

  $$F = \eta A v / d$$

  where the proportionality constant η is the viscosity of the fluid.

  – The viscosity for air is about 1.5×10^{-5} pascal-seconds, water is about 10^{-3} pascal-seconds at room temperature, honey is a few pascal-seconds.
- The drag on a small sphere falling through a fluid with no turbulence is given by **Stokes' law**:

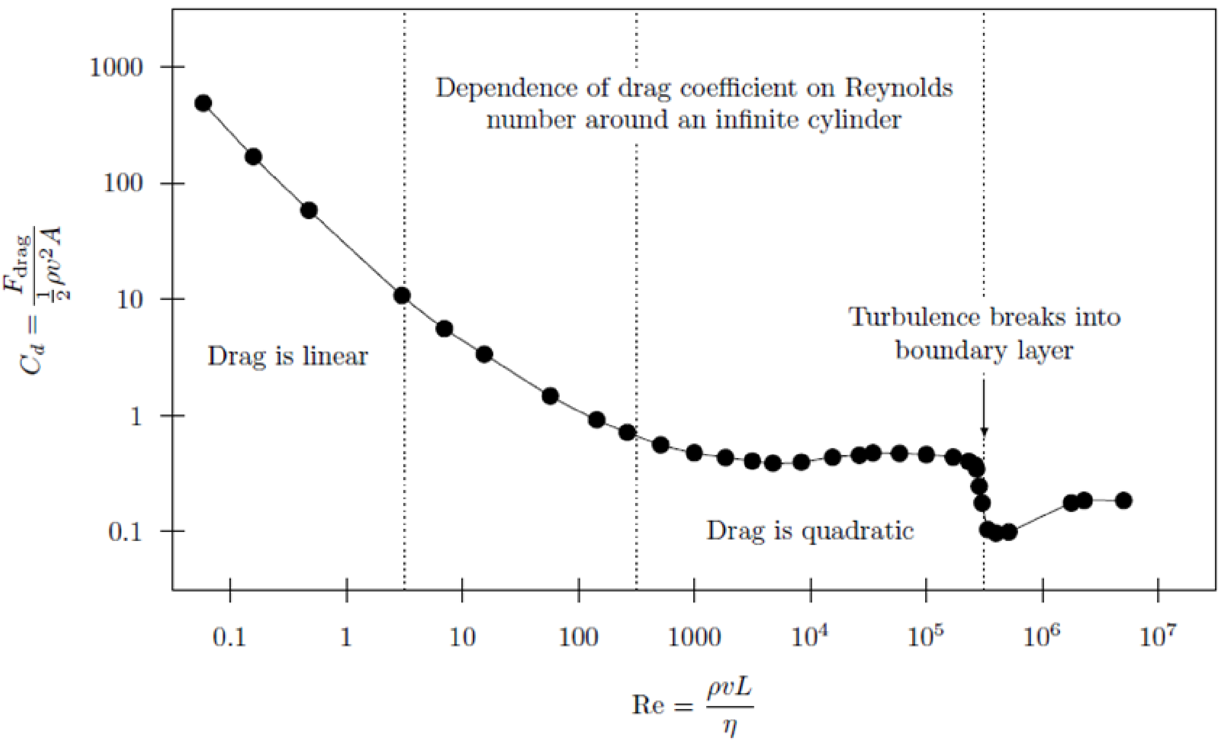
  $$F = 6\pi\eta R v$$

- Because of viscosity, any steady fluid will lose pressure as it flows. **Poiseuille's equation** defines how much of a pressure differential is required to overcome the frictional effect of viscosity through a tube of constant radius:

  $$\Delta P = \frac{8\eta L Q}{\pi r^4}$$

  where L and r are the length and radius of the tube and Q is the volume flow rate through the tube.
- We have assumed that η is constant for a given temperature. The viscosity of some fluids depend on pressure—the harder you push, the more they resist. This sets up a kind of non-linearity which can cause some truly bizarre, behavior. These are called **non-Newtonian** fluids. Corn starch and water (2:1 ratio) is pretty fun to play with!

Turbulence and Reynolds Number



- So far, we have left out one of the most characteristic things about the flow of fluids: **turbulence**. Turbulence involves some of the last unsolved problems in classical mechanics, so we can be excused for not going into too much detail here.

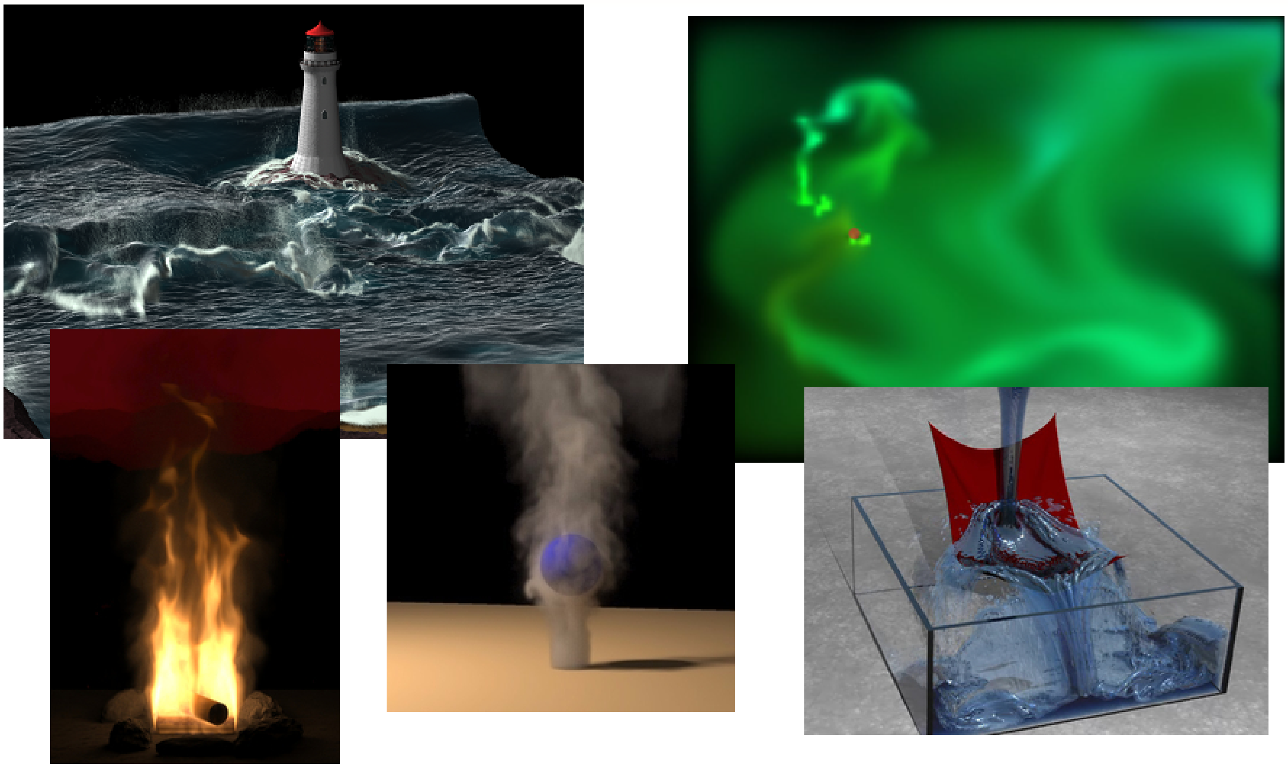
- In general, the level of turbulence is driven by a combination of the speed and size of the object and the viscosity and density of the fluid. These quantities can be combined together to yield the dimensionless **Reynolds number**:

$$\text{Re} = \rho v L/\eta$$

- It can be shown that the fluid equations for different situations with similar Reynolds numbers are similar.

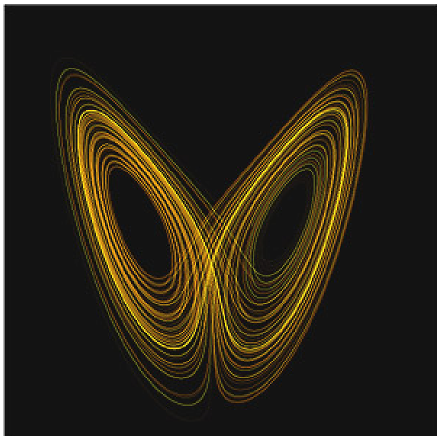
- The chart shows some experimental data for an "infinite" cylinder. There are a few clear regimes in the chart.
  - For Reynolds number under 10, the viscosity of the fluid maintains a laminar flow and the resistance is roughly linear with velocity. This is comparable to the drag modeled by Stokes' law we mentioned earlier for a sphere. The energy is lost primarily through friction.
  - For Reynolds number over 100, turbulence dominates the drag. The energy is lost through the complex convection currents generated in the flow rather than friction. The drag is almost all quadratic.
  - In between these regions, there is a mixture of laminar and turbulent flow. Different patterns are evident with the trend moving into more turbulence for higher Reynolds number. A "boundary layer" of steady fluid flow forms around the cylinder.
  - The around $\text{Re} = 30,000$ the turbulence breaks into the boundary layer. This has the effect of enveloping the turbulence into a streamlined wake. A dramatic drop in drag occur here.

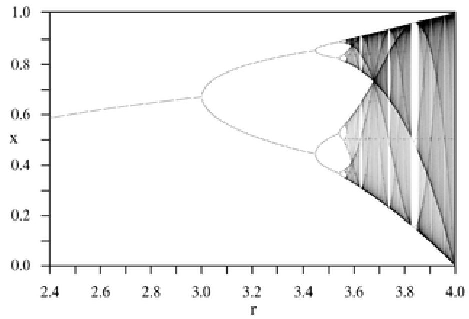
Beyond Bernoulli: Navier-Stokes

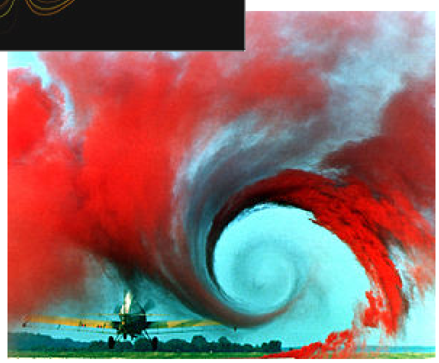


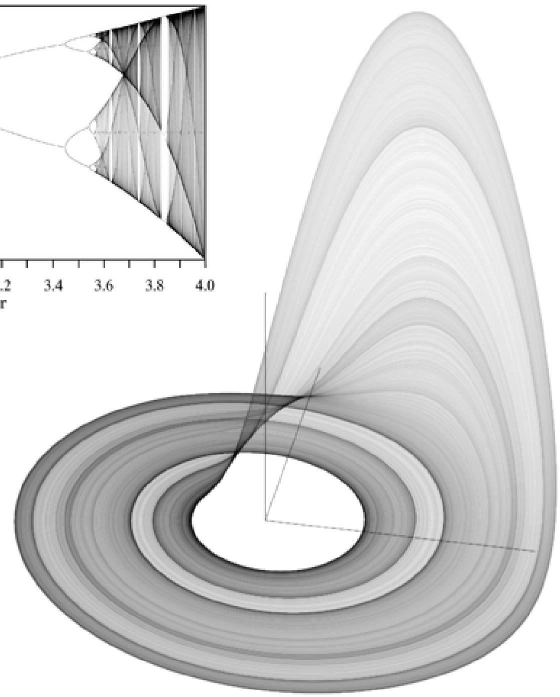
- The Navier-Stokes equations are the application of Newton's second law to fluid dynamics. They take into account the pressure gradients, velocity profiles, viscosity, rotational effects, and any other external forces that act on the fluid.
- In principle, these equations determine the physical properties of any fluid. The problem is that they have only been solved for the most simplest of cases.
- But with computing power increasing every day, we have reached the point where computer simulations can be used to model realistic looking fluid flow with the Navier-Stokes equations. This field of study is called "computational fluid dynamics" and is under active development.
- Conceptually, the solution is there in the Navier-Stokes equations, but the reality is that it is very difficult to build an efficient calculation algorithm. A lot of thought must go into what one must keep and what one can throw out of the simulation.
- When modelling is done for games, only the visual effect is important. In that case, we can drop the information regarding pressure and focus on solving for the velocity field pattern over time.
- One of the millennium prize awards from the Clay Mathematics Institute is for a general solution to the Navier-Stokes equations: http://www.claymath.org/millennium/Navier-Stokes_Equations/.
- Also: notice that we haven't even touched on issues related to surface tension...

Deterministic Chaos









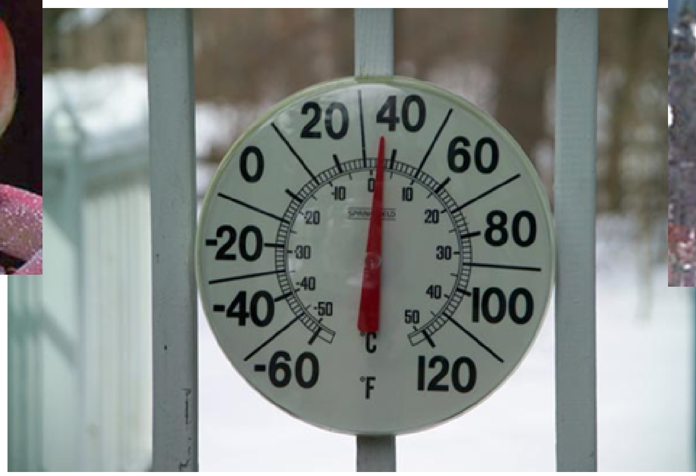
- In 1963, Lorenz was studying weather patterns using an extremely simplified version of Navier-Stokes. Though highly idealized, his mathematical model kept one element he felt critical for more accurate weather prediction: non-linearity.
- He discovered that in certain non-linear mathematical models, the simulated outcome was exponentially sensitive to the initial conditions. A slight change in the 10th decimal place eventually infected the entire system causing the dynamics to diverge wildly. He called this the **butterfly effect**.
- One of the simplest mathematical models exhibiting this **deterministic chaos** was discovered by Rossler and shows the two necessary elements: explosive sensitivity and contained mixing. The sensitivity makes sure the system does not repeat its path, but the containment means that the paths do not diverge off to infinity either. They are "attracted" to a complex system of solutions never leaving and never quite repeating.
- It is important to point out that this is a *mathematical* issue. This does not have to do with experimental error—the path is completely determined by the equations. The "mixing" that occurs is completely reversible. (See here.)
- The real problem is we cannot determine the end result without performing all the intermediate calculations. Now, technically for any mathematical model only infinite precision will guarantee infinite predictability.
- For linear systems, a doubling of precision will double the time you can expect decent results. The output is proportional to the input.
- For chaotic systems, a doubling of precision may only give you one more iteration, one more step in time. This is why the weatherman cannot forecast more that two or three days. It's mathematically impossible.
- Chaos theory show us that, in the end, it is easier to control the weather than predict it!

Physics 202 Lecture 4

Heat Flow and Temperature

- Now for a complete change of subject.
- Questions regarding the nature of temperature and heat flow have ancient roots. The Greeks included heat as one of the four basic elements of the universe.
- In the modern era, attempts were made to reduce the study of heat and temperature to mechanical models. A fluid model was developed, but eventually rejected.
- We now understand that temperature is the manifestation of the random perpetual motion at the molecular level. This explanation is called **kinetic theory** and we will talk about this is in a future lecture.
- For now, we are interested in simply establishing facts. And the beginning of any investigation requires a method of measurement. This is where we will start.

Temperature Scales



$$F = (9/5)C + 32$$

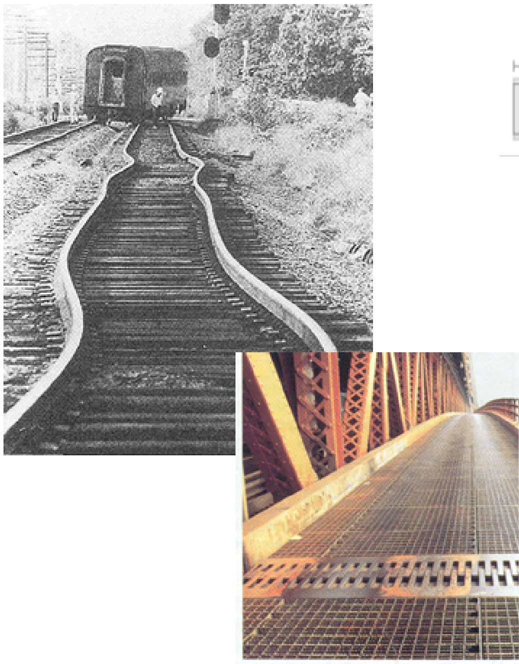
$$C = (5/9)(F - 32)$$

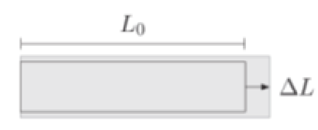
$$K = C + 273.15$$

$$C = K - 273.15$$

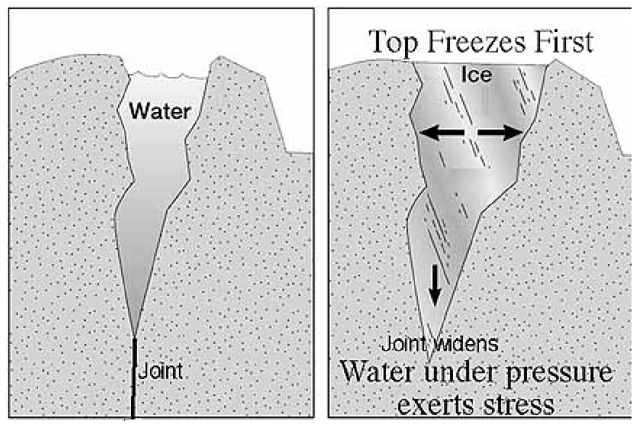
- There are three temperature scales in common use today.
- The first was the Fahrenheit scale with 180 degrees between the melting point and boiling point of water. Fahrenheit calibrated the scale so that body temperature was 100 °F. But he was a bit off: typical body temperature is 98.6 °F. The boiling point of water is 212 °F and the melting point is 32 °F.
- Later, Celsius built a scale with 100 degrees between the melting point and boiling point of water. He calibrated his scale so that the boiling point of water was 100 °C, so the melting point of water is 0 °C. Therefore, the conversion between the two is
$$C = (5/9)(F - 32) \qquad \text{and} \qquad F = (9/5)C + 32$$
- Still later Kelvin came along and calibrated the Celsius scale to **absolute zero**.
  - In the early 1700s it was discovered that the pressure of a gas is directly proportional to its temperature. Its pretty easy to extrapolate backward to the temperature at which the pressure is zero. Since pressure cannot be negative, the gas can't get colder than this temperature.
  - Furthermore, experiments showed that this extrapolation yields the same minimum temperature.
  - Kelvin, working in the middle 1800s used very general thermodynamic arguments to show that this minimum temperature is -273.15 °C.
- There are still 100 degrees between the melting point and boiling point of water in the Kelvin scale, but water melts at 273.15 kelvin, so
$$K = C + 273.15 \qquad \text{and} \qquad C = K - 273.15$$
- Notice that by definition, a *change* in temperature has the same value in Celsius or Kelvin.

Thermal Expansion





$$\frac{\Delta L}{L_0} = \alpha \Delta T$$

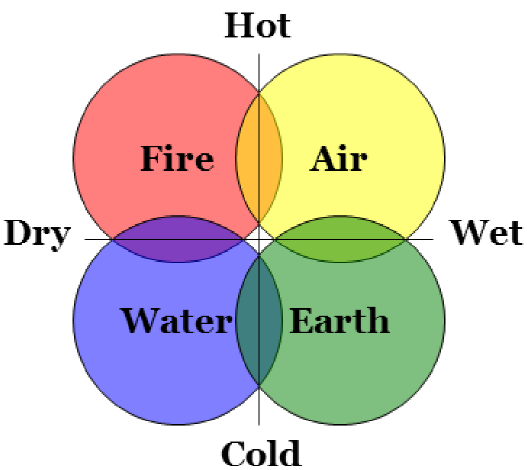


- In order to measure temperature, we need to identify a **thermometric property**—that is, some physical effect that is controllably and consistently changed by temperature levels. Examples include
  - The volume of a liquid
  - The length of a metal rod
  - The pressure of a constant volume of gas
  - The electrical resistance of a piece of wire
  - The speed of sound
  - The color of a hot stove
- Any thermometric property can be represented mathematically as a function of temperature. And any function can be considered linear for small increments (cf. binomial theorem).
- In particular, we can consider the linear expansion of an object due to temperature. The formula for it is

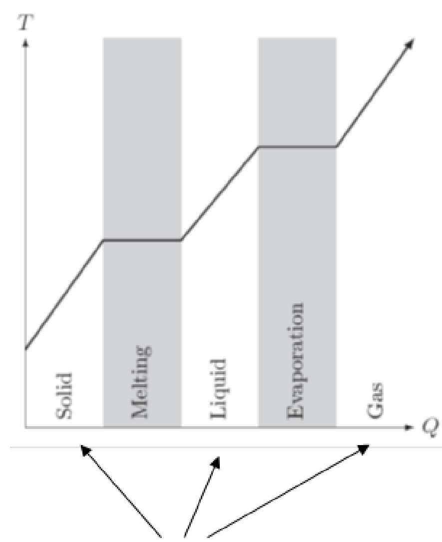
  $$\frac{}{L_0} = \alpha \Delta T$$

  where α is called the coefficient of **thermal expansion**.
- Notice that the left-hand side of the equation is the same percent strain we introduced when discussing elasticity. Even though this percentage may be small, the thermal expansion accumulates over the total length.
- Usually the value of α is positive indicating that things tend to expand with increasing temperature. Hot air rises because it is less dense due to this thermal expansion.
- However, cold water (below 4 °C) will expand with decreasing temperature. So ice floats too.

Heat Capacity and Internal Energy



$$1 \text{ cal} = 4.186 \text{ J}$$



$$Q = C\Delta T$$
$$Q = cm\Delta T$$

- Usually when an object is subjected to a source of heat, its temperature rises. How much the temperature rises is a property of the object and is called the **heat capacity** of the material. Thus,

$$Q = C\Delta T$$

  where Q is the heat absorbed and ΔT is the change in temperature.

- Of course, the more material there is the more the object is able to hold heat. The specific heat capacity is the intensive material property of which the object is composed. We have

$$Q = cm\Delta T$$

  where c is the specific heat capacity and m is the mass of the object.

- During the 18th century, it was unclear what the mechanical nature of heat was. This specific heat formula gives us an "agnostic" way of quantifying heat through our definition of temperature.

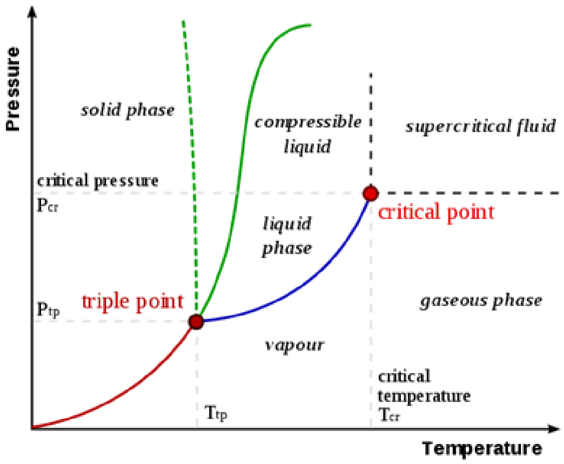
- This is why one **calorie** of heat is defined as the amount required to raise one gram of water one degree Celsius. So, the specific heat of water is by definition 1000 cal/kg-°C. (By the way, one food calorie equal 1000 heat calories).

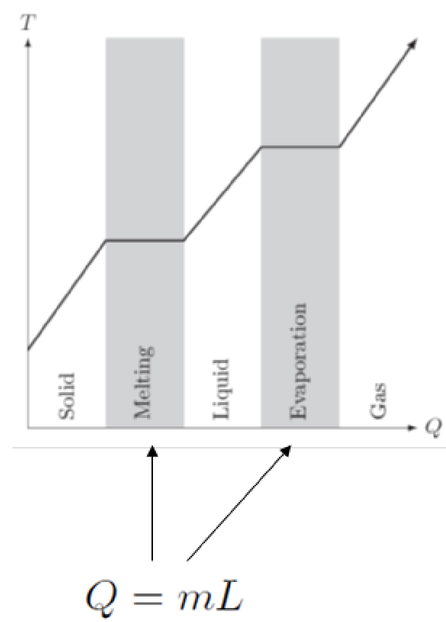
- Now we know that heat is associated with internal molecular energy. As such, there is a conversion factor between calories and joules:

$$1 \text{ cal} = 4.186 \text{ J}$$

- Therefore the specific heat capacity of water is 4186 kJ/kg-°C. We will have more to say about this equality next lecture.

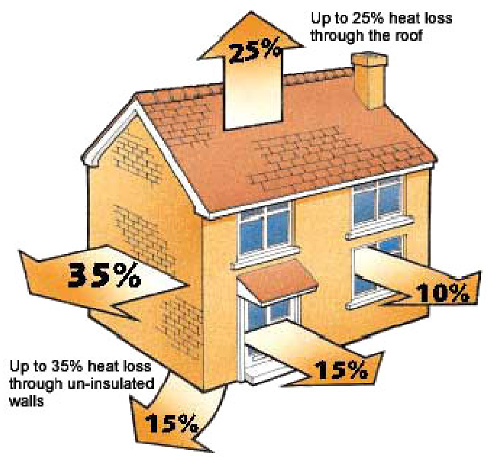
Phase Changes and Latent Heat





- Usually when an object is subjected to a source of heat, its temperature rises. But sometimes it doesn't. When a substance melts or boils, it absorbs or releases heat without changing temperature. This is called **latent heat**.
- Every phase change has a corresponding latent heat value.
  - Even non-standard ones like the changes between the different molecular arrangements in some solids.
  - One extreme example is plutonium. It has six main solid phases for different temperatures. This makes is fairly difficult to work with and is commonly slightly alloyed with gallium to force it into its most stable (and therefore most easily workable) solid form.
  - Another example is water-ice with 15 identifiable phases at various levels of pressure and temperature.
- The formula involving latent heat is simply
$$Q = mL$$
where L is the specific latent heat for the phase change. For water we have:
  - Latent heat of fusion (freezing/melting) : $L_f = 334$ kJ/kg
  - Latent heat of vaporization (boiling/condensation) : $L_v = 2260$ kJ/kg
- Under low pressure it is possible for a substance to transition directly from solid to gas: this is called sublimation. The opposite process is called deposition. These details can be seen on a pressure-temperature phase diagram.
- During any phase change, the substance is in a combination of the two phases. At a certain value of pressure and temperature, it is possible for a substance to be in all three phases simultaneously. This is called the **triple point**. (For water: 273.16 K and 0.6117 kPa.)
- Beyond the triple point, we have the **critical point** in which the pressures and temperatures are so high that the distinction between the gas and liquid phase disappears. (For water: 647.1 K and 22,060 kPa.)

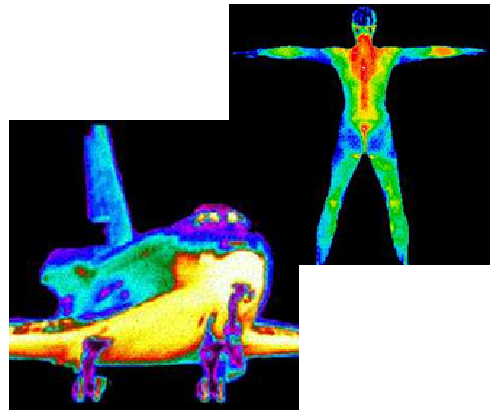
Conduction and Convection



$$P = \frac{Q}{t} = \frac{kA}{L}(\Delta T) \qquad R = \frac{L}{k}$$

- So far we have discussed how the transfer of heat affects an object either through an increase in temperature or a change of phase. Now we will to discuss the flow of heat between objects—the rate at which heat flows (the heat current, if you will).
- Heat flow is traditionally divided into three types: conduction, convection, and radiation. These distinctions are due to the medium through which the heat flows.
  - If the heat flows through a solid object, the main method of transporting heat is through **conduction**. This is a kind of diffusion of the heat energy because the medium itself will not move.
  - If the heat flows through a fluid system, the main method of transporting heat is through **convection**. This is simply when the hot fluid flows through convection currents. Typically hot flows up and cold flows down because of thermal expansion and buoyancy.
  - Heat can also flow through vacuum by accessing the electromagnetic field itself. This is called heat **radiation**.
- The formula for the conduction of heat through a solid is
$$P = \frac{Q}{t} = \frac{kA}{L}(\Delta T) \qquad \text{and} \qquad R = \frac{L}{k}$$
where k is called the **thermal conductivity** for the solid material. Sometimes the insulating power of an object is given by its R-value which is equal to L/k. When heat passes through multiple layers of material, the total insulating power is the sum of these R-values.
- There is no compact formula for the convection of heat because it involves the complexities of fluid flow. For simple situations, we may assume that the heat flow due to convection is proportional to the temperature differential (like in conduction)—this is Newton's law of cooling. But the validity of this rule is inconsistent.

Heat Flow Through Radiation



$$P = \frac{Q}{t} = \epsilon\sigma AT^4$$

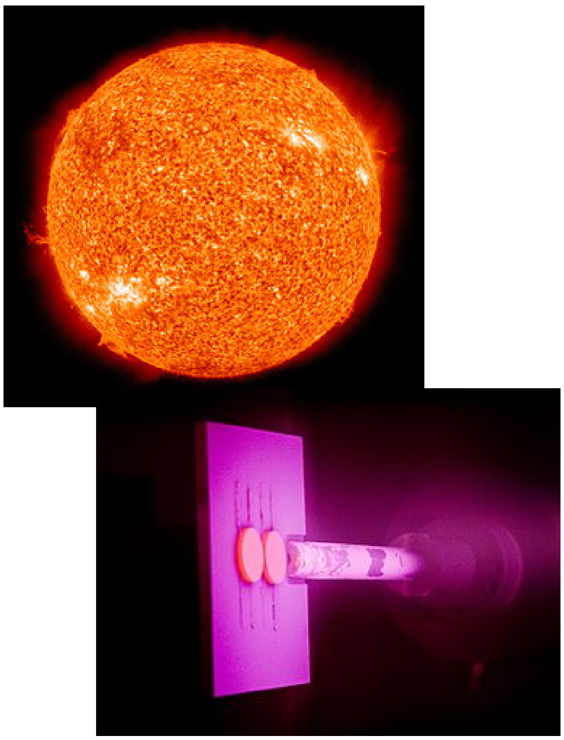
5.67×10^{-8}

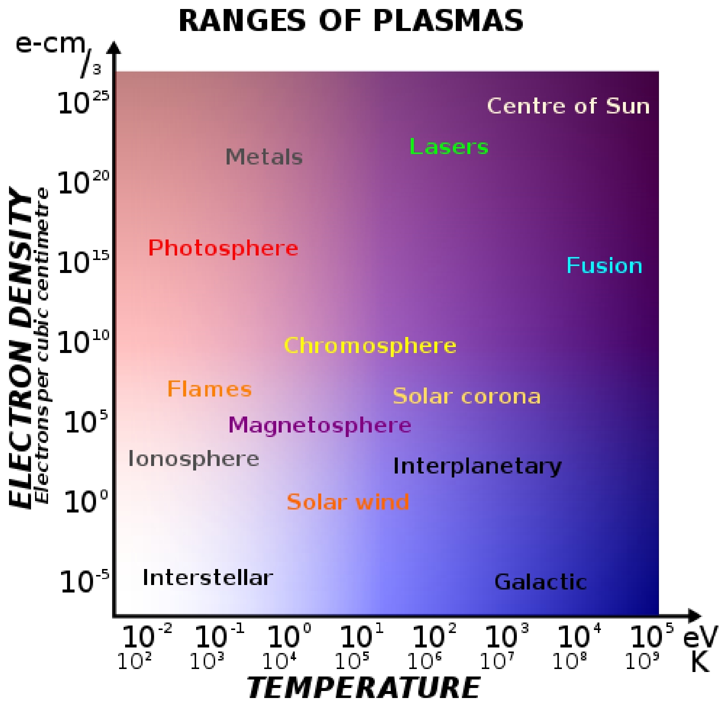
- The third form of heat flow is through the long-range force fields.
  - We usually talk about radiation through the electromagnetic field, but there is no reason we could not talk about radiation through the gravitational field.
  - However, gravity is much weaker than electromagnetism, so we can safely ignore it now.
- All matter emits thermal radiation based on its temperature. The formula is called the **Stefan-Boltzmann** law:

$$P = \frac{Q}{t} = \epsilon\sigma AT^4$$

  - The ϵ is the emissivity of the object—a number that varies from zero to one. Graphite and carbon black have emissivities over 0.95 while polished silver is at 0.02. The ideal value of one corresponds to a so-called **blackbody**.
  - The σ is called the Stefan-Boltzmann constant and is a characteristic of electromagnetic radiation. Its value in SI units is 5.670×10^{-8}.
  - The A is the exposed surface area of the object.
  - The T is the absolute temperature of the object in Kelvin. You must convert from Celsius to Kelvin to use this formula.
- This law was experimentally derived by Stefan and theoretically derived by Boltzmann—whose argument was based on the thermodynamics of a heat engine using light(!) as the working matter rather than a gas.
- Just as every object radiates heat, so does every object absorb heat radiation. This law also describes the rate at which an object *absorbs* radiative heat from its surroundings.

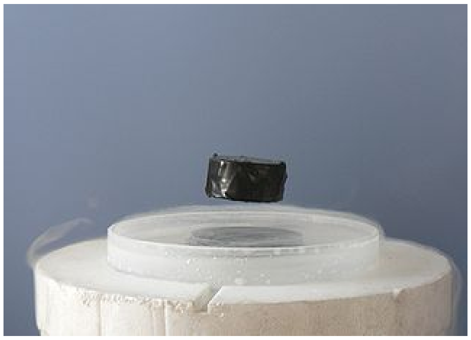
Things Very Hot: Plasma Physics

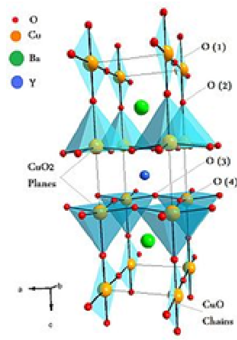


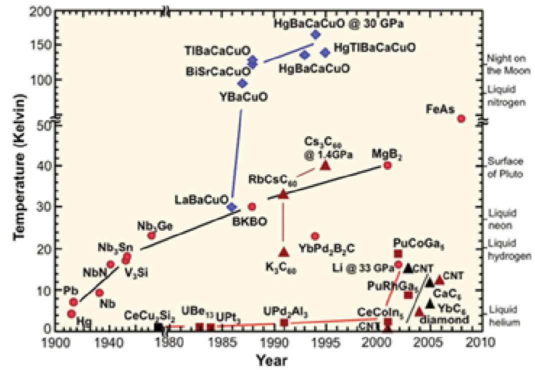


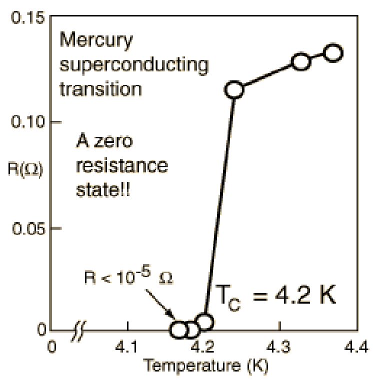
- Finally, I wanted to round out this lecture with the two extremes in temperature. We'll start with the top end.
- As temperatures increase, the internal energy of the gas increases. Eventually, this energy becomes large enough to start ripping the atoms apart in the gas.
- The condition is that the thermal energy must be within a percent or so of the ionization energy of the atoms. The thermal energy is simply kT, where k is the Boltzmann constant equal to 1.381×10^{-23} in SI units.
  - For hydrogen, the atomic ionization energy 13.6 eV (which is equivalent to 1312 kJ/mol).
  - One percent of the corresponding thermal energy is 1600 kelvin. Increasing the temperature exponentially increases the density of plasma versus non-plasma gas.
- When a substance enters into the plasma phase, significant difference occur.
  - The plasma is electrically active and is a mixture of electrons, ions, protons, and neutrons.
  - The velocities of the particles are frequently much faster than the neutral gas as a consequence.
  - Although collisions dominate the dynamics of a gas, the electromagnetic effects of the plasma create much more complex and long-range interactions which are extremely difficult to model and control.
- Artificially created plasmas are frequently is use. Plasma screen TVs and neon lights are common examples. Industrial application like semiconductor deposition techniques and one of the most promising methods harnessing fusion energy uses magnetically confined plasma.
- Plasma is also an important element in astronomy. The Sun itself is a huge burning ball of plasma. The solar wind is essentially this plasma with very low density. The Earth is surrounded by a plasma called the ionosphere which protect life from cosmic radiation. The planet Jupiter and its moon Io are connected by a plasma "flux tube" due to its excessive volcanic activity. Plasma effects even exist on interstellar and intergalactic levels.

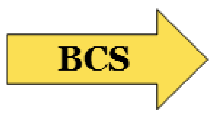
Things Very Cold: Superconductivity

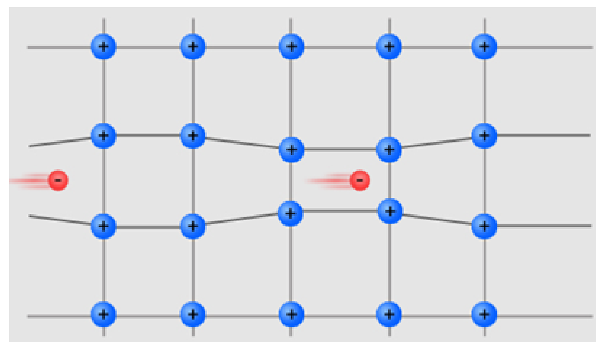












- Exploring the extreme cold end of the temperature spectrum began in earnest in the late 1800s. Superconductivity first discovered in 1911.
- In general, electrical resistance decreases with decreasing temperature. But most metals have a characteristic temperature at which this resistance drops dramatically to zero. Literally zero. For solid mercury (the first one discovered), the critical temperature is 4.2 kelvin.
- Another important effect associated with this zero resistance is called the **Meissner effect**. A superconducting metal will expel magnetic fields from its interior. This leads to a repulsion which can be used to levitate the superconductor above another stationary magnet.
- The theoretical explanation for superconductivity was completed in 1957.
  - Quantum mechanics plays an essential role in the explanation. Usually only for the atomic scale or smaller, these quantum effects accumulate in the superconductor.
  - The idea is that the electrons pair up through an interaction with the metal substrate. As a consequence of this relatively weak interaction, the pair acts more like light than matter.
  - Only when the temperature is very small can this slight interaction overcome the random thermal motion of the material. But once it does, the system locks into place.
- Over the years, investigators have sought superconductors with a higher and higher critical temperature. A breakthrough occurred in 1986 with the discovery of a crystalline compound **YBCO** with a critical temperature of 93 kelvin. This is important because it is above the boiling point of liquid nitrogen (77 kelvin) which is a relatively easy to use refrigerant.

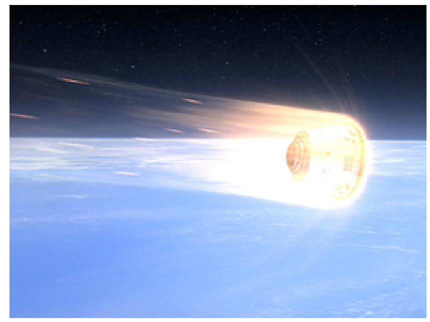
Physics 202 Lecture 5

Thermodynamics

- This year, I've decided to try and mix this up a bit.
- Usually a conservation about kinetic theory precedes a formal introduction to thermodynamics. But the reality is that most of the theory was worked out be physicists and engineers who had the *wrong* mechanical model of heat.
- In addition, the kinetic theory we will describe in the next lecture is also wrong! We will use a classical model—which is good enough in general, but limited. The full theory of quantum mechanics is necessary for more precision.
- This shows that the thermodynamics arguments in this lecture based on very general principles and independent of the statistical mechanics underneath. This is my attempt to show you how this general construct works.

The First Law: Heat and Work

$$\Delta U = Q - W$$

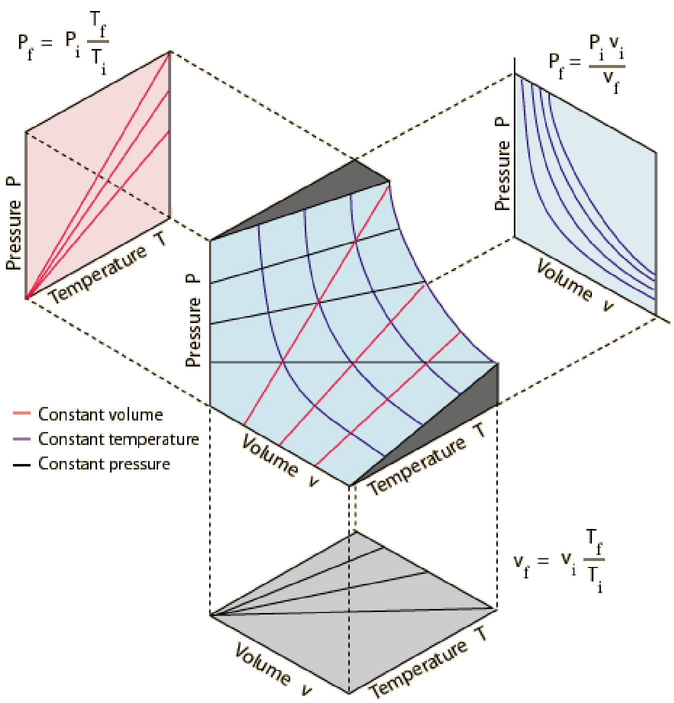


- Thermodynamics was initially about the study of high pressure steam engines. As the temperature of a gas increases, it expands (a manifestation of thermal expansion). This expansion is then harnessed mechanically to move pistons, axles, etc.
- By definition, the work done by any engine is $F\Delta x$. For a gas, it is easier to talk about its pressure and volume. The work done by an expanding gas is $P\Delta V$. A pump is usually used to transport the steam/water through the cycle, but the work output far exceeds this work input by a factor of 50-to-1 or so.
- That energy comes from somewhere—the internal energy of the gas. We also know that the flow of heat affects the internal energy of a system. This is the **First Law of Thermodynamics**:

$$\Delta U = Q - W$$

- Mathematically this equation only makes sense if heat is a form of energy, like work. Joule was the person to establish the mechanical equivalent of heat. Friction converts mechanical energy into heat. Through precise experimentation, Joule was able to determine that 4.186 joules of mechanical energy lost through friction will increase the temperature of a gram of water one degree Celsius. Thus, one calorie of heat is 4.186 joules of energy.
- So the first law is really a statement concerning the conservation of energy. The sign on W is negative because work done *by* the system is at the expense of its internal energy (like an engine). Work done *on* the system will increase the internal energy (like friction).

The Ideal Gas Law

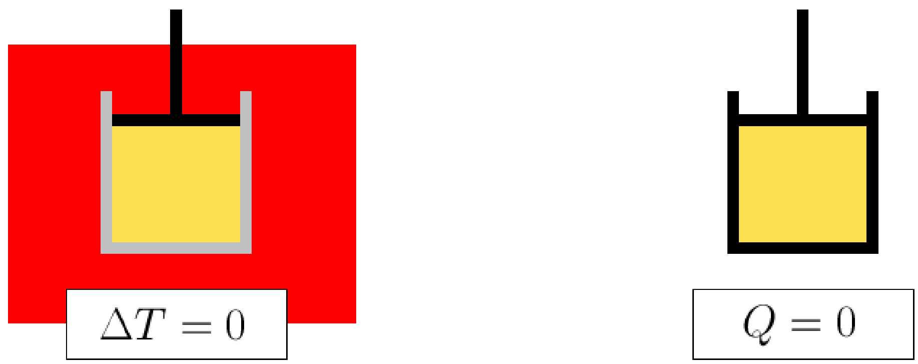


$$PV = nRT$$

$$U = \frac{3}{2}nRT$$

- Gases are the best material for a heat engine because of they are easy to manipulate and respond aggressively to temperature changes. The **ideal gas law** shows how temperature, pressure, volume and the amount of gas are related to one another:
$$PV = nRT$$
where n is the amount of gas in **moles** (mass divided by molecular weight), and R is the universal gas constant equal to 8.31 in SI units.
- The dependence of the internal energy on the temperature for an ideal gas is
$$U = \tfrac{3}{2}nRT$$
- Each of these are called an **equation of state** for our system—it relates the physical quantities in the system with its temperature.
- It's the job of statistical mechanics to provide these equations for our system. But once they are had, we proceed with the thermodynamics.
- For real gases, both of these equations need to be modified. For example, if the gas is diatomic (two-atoms: like oxygen gas or hydrogen gas) the internal energy is
$$U = \tfrac{5}{2}nRT$$
- Also, the **van der Walls** equation is a common modification for the ideal gas law:
$$\left(P + a\frac{n^2}{V^2}\right)(V - nb) = nRT$$
where the values of a and b must be experimentally determined for each gas.

Isothermal and Adiabatic Processes



$$W_{\mathrm{isoth}} = nRT\ln(V/V_0) \qquad W_{\mathrm{adiab}} = -\tfrac{3}{2}nR\Delta T$$

$$PV = \mathrm{const.} \qquad PV^\gamma = \mathrm{const.}$$

$$\gamma = 5/3$$

- The work we can extract from an ideal gas depends upon how we do it. There are two important processes to consider. The first process is the **isothermal** process (constant temperature). Since both pressure and temperature are changing we need a bit of calculus in order to use the ideal gas law and derive the work done. The answer is

$$W_{\mathrm{isoth}} = nRT\ln(V/V_0)$$

- The second process is the **adiabatic** process (no heat flow). If there is no flow of heat, the first law tells us that all the work comes from the internal energy of the gas. Since $U = \frac{3}{2}nRT$ and $Q = 0$, we have

$$W_{\mathrm{adiab}} = -\tfrac{3}{2}nR\Delta T$$

- In a way, both of these processes divide the connection between heat and temperature. In the adiabatic process temperature changes without heat flow, and in the isothermal process heat flows without a change in temperature.
- One graphical way to represent the work done in any thermodynamic process is with a PV diagram, with volume on the horizontal and pressure on the vertical. At any one time, the gas occupies a particular spot on this chart. As an engine runs through its cycle, it draws a closed curve on the chart. The area enclosed is the net work performed by the engine for that cycle.
- According to the ideal gas law, an isothermal process will trace out the curve

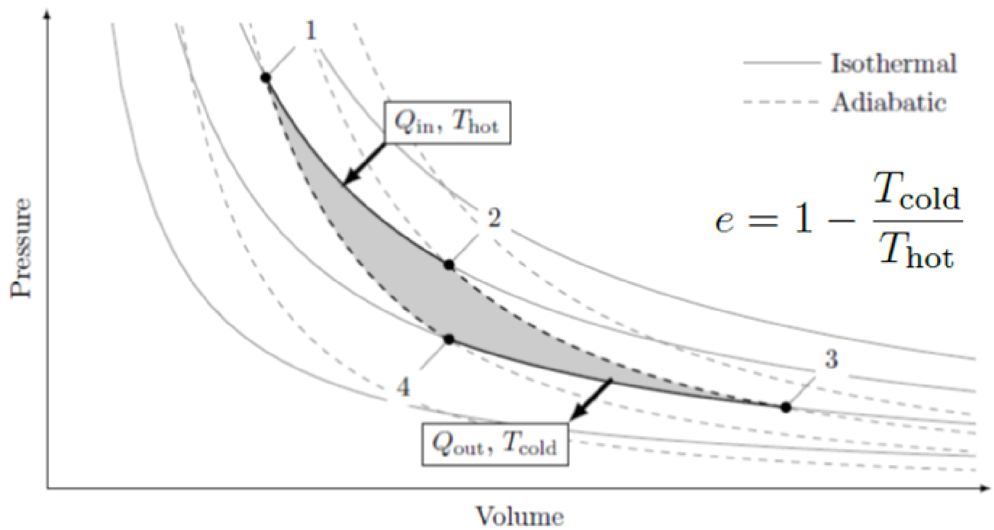
$$PV = \mathrm{const.}$$

- But for an adiabatic process on an ideal gas, the pressure-volume relationship is

$$PV^\gamma = \mathrm{const.}$$

with $\gamma = 5/3$ for a monatomic ideal gas. This says that under compression the pressure in an adiabatic process will rise faster than in an isothermal one. Under expansion the pressure will fall faster too.

A Reversible Heat Engine



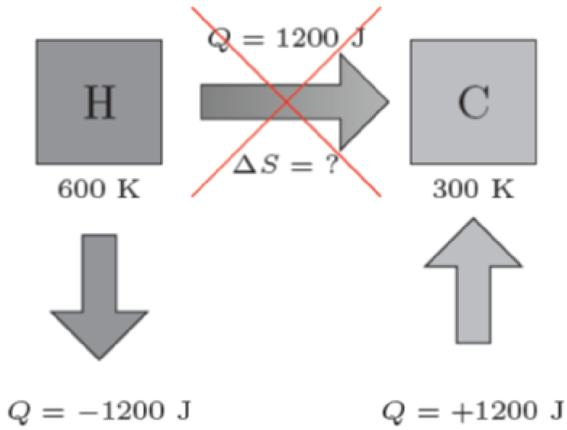
- The **Second Law of Thermodynamics** simply states heat will flow spontaneously from hot to cold.
- The key word in this statement is "spontaneously"—which is closely related to **irreversibility**. Of course, heat can flow against a temperature differential (this is what a refrigerator does after all). But when the heat flow is spontaneous, the system does not control the flow of energy.
- This is unlike a mechanical system in which the energy is flowing due to the internal forces between the parts. The mechanical energy distribution is directly related to the configuration of the system, so it is possible (in principle) to reverse the flow of energy by simply reverting this configuration back to its initial state. There is nothing a mechanical system can do to reverse the spontaneous flow of heat in this way.
- But the isothermal process does not involve a temperature differential and the adiabatic process does not involve the flow of heat. Therefore both of these are reversible.
- The simplest non-trivial cycle which combines these two processes is called a **Carnot cycle**. It moves between the hot and cold temperatures during the adiabatic stages and harnesses the flow of heat during the isothermal stages.
- Now, the efficiency for any heat engine is
$$e = \frac{W_{\text{net}}}{Q_{\text{in}}} = 1 - \frac{Q_{\text{out}}}{Q_{\text{in}}}$$
- For the Carnot cycle, it's not too difficult to show that
$$e = 1 - \frac{T_{\text{cold}}}{T_{\text{hot}}}$$
You will need to remember that there is no change to the internal energy during the isothermal stage, so the first law says that $Q = -W$. Also you will have to show that the pressure-volume relations imply that $V_2/V_1 = V_4/V_3$.

Irreversibility and Entropy

- Here's the kicker (and also the insight that put Carnot in the history books). This is the best any heat engine can do between heat reservoirs of these two temperatures. The argument has two parts.
  - First, every reversible engine will have the same efficiency. One way to see this is by supposing we have two reversible engines with different efficiencies. Since they are truly reversible, run the more efficient one backward by using the work generated from the first one. In this way we generate excess work without any net heat flow at all!
  - Second, any irreversible engine will have a lower efficiency than a reversible one. This is pretty obvious as any spontaneous heat flow will simply add to the wasted heat without any productive work.
- Although our efficiency formula was derived for an ideal gas, it represent the maximum efficiency of any heat engine. Obviously a real heat engine will be worse. A real heat engine involves some level of spontaneous heat loss.
- **Entropy** is a measure of the distribution of internal energy in a system and is related to how closely the system is to thermal equilibrium. The broader the distribution, the larger the entropy.
  - For example, when one hot and one cold object are thermally connected (forming one system) heat will flow. The internal energy of the combined system redistributes itself until it reaches thermal equilibrium (uniform temperature). Another way to say this is that the system moves from a low to a high entropy state.
- Whenever heat flows in a system, its entropy changes. In a reversible isothermal process, the change in entropy is

$$\Delta S = Q/T$$

Entropy Quantifies the Second Law



$$\Delta S = Q/T$$

$$\boxed{\Delta S \geq 0}$$

- Entropy helps identify what is special about the Carnot cycle—the net entropy change is zero (which follows from the formula for efficiency). This is why the Second Law of Thermodynamics is often written as

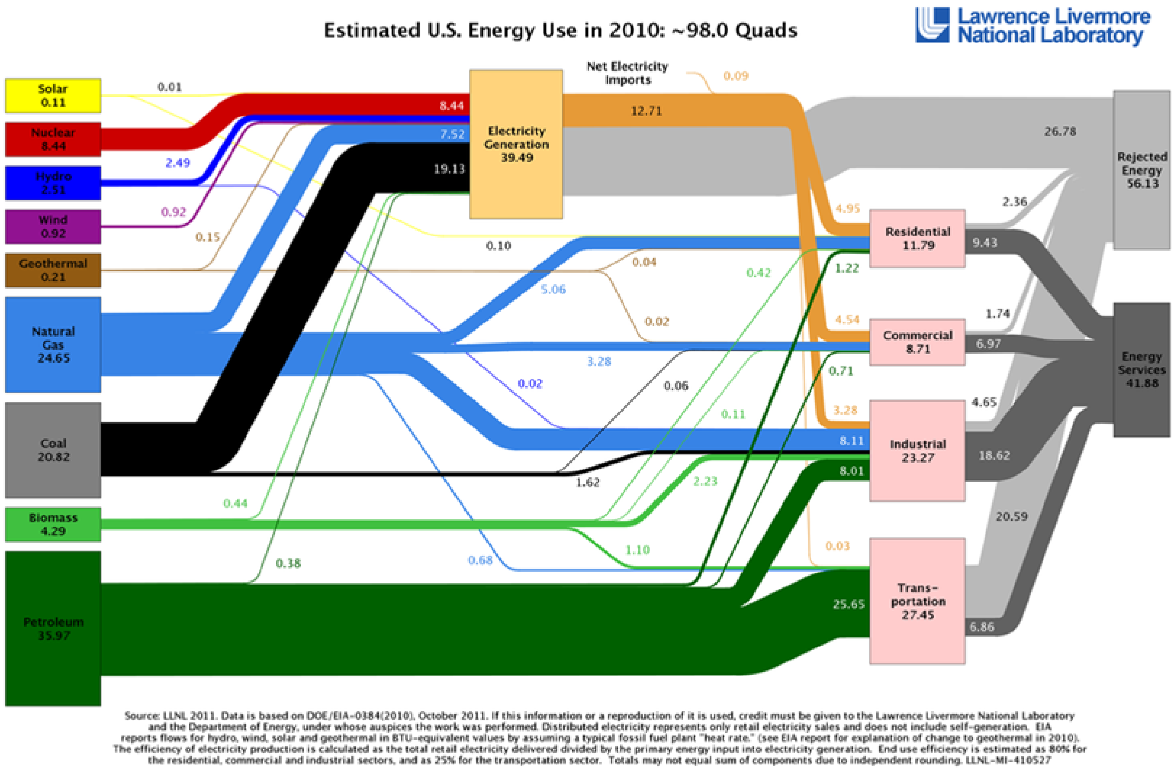
$$\Delta S \geq 0$$

- Be aware what this statement does not say: it does not say entropy never decreases—it often does. But whenever the internal energy of a system is pushed away from thermal equilibrium, the entropy also increases in the external environment that is doing the pushing.
- When this process is reversible, these two entropy changes are equal. If the process is irreversible, the decrease in entropy in the system is accompanied by a larger increase in entropy somewhere else.
- How do we calculate this increase? The trick is to imagine a corresponding *reversible* process with the same initial and final distributions of internal energy. For example, consider spontaneous heat flow.
  - Imagine that we isolate the hot and cold parts of the system from each other and connect each of the parts to a heat bath with corresponding hot and cold temperatures.
  - Then force equal amounts of heat to flow isothermally from the hot part of the system to the hot bath and from the cold bath to the cold part of the system.
  - Notice how the distribution of internal energy is now the same as with the spontaneous heat flow. The entropy change in the system due to the isothermal heat flow is

$$\Delta S = -\frac{Q}{T_{\text{hot}}} + \frac{Q}{T_{\text{cold}}} > 0$$

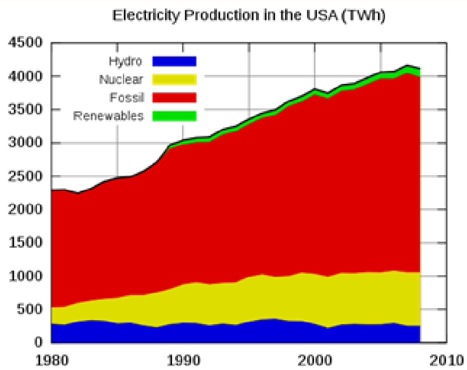
  - For this imaginary scenario, the entropy of the surrounding environment decreases by the same amount. But in the spontaneous process, the only entropy change is the increase in the system.

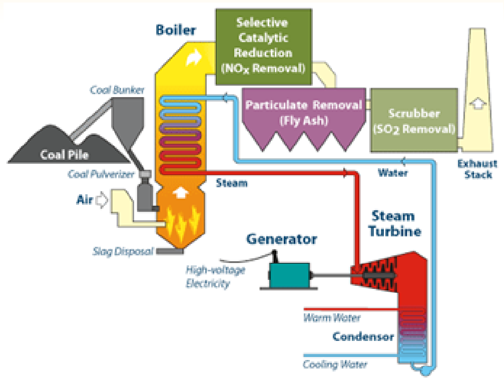
Energy Consumption

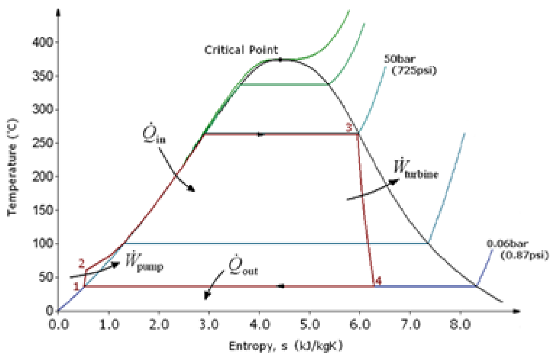


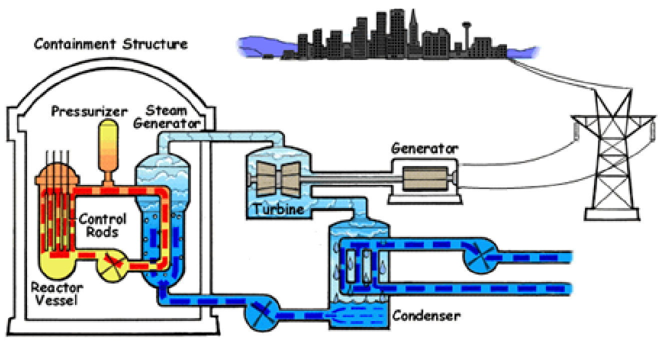
- Supply side:
  - Passive sources of energy (wind, geothermal, solar) add up to just over 1% of total energy supply.
  - Other non-fossil fuel sources of energy (nuclear, biomass, hydro) add up to about 15% of the total.
  - Fossil fuel supply comes primarily from petroleum (37%), then natural gas (25%), and coal (22%).
- Demand side:
  - 40% of all our energy supply is used to generate electricity. The remaining 60% is used directly.
  - Nearly half of the electricity is generated from burning coal—and this accounts for nearly all the coal demand.
  - Residential consumption of energy accounts for about 17% of total energy demand.
- Efficiency:
  - Overall energy efficiency is 42%.
  - Transportation sector is very inefficient at 20%.
  - Electricity generation is only about 32% efficient.
- One "quad" is 10^{15} BTU or 1.055×10^{18} joules. The average household uses about 2000 joules per second.

Electricity Generation









- Hydro, wind, and solar generate electricity directly. For the first two, electromagnetic induction is used to convert motion into AC current. Solar converts light energy into electricity.
- Every other means of generating electricity utilizes steam to drive a heat engine whether renewable or not. This is about 90% of the electricity generation in the US and it is all limited by the Carnot efficiency.
- Most steam engines operate using the **Rankine cycle**.
- From http://en.wikipedia.org/wiki/Rankine_cycle:
  - **Compression**. The working fluid is pumped from low to high pressure. As the fluid is a liquid at this stage the pump requires little input energy.
  - **Heating**. The high pressure liquid enters a boiler where it is heated at constant pressure by an external heat source to become a dry saturated vapor.
  - **Expansion**.The dry saturated vapor expands through a turbine, generating power. This decreases the temperature and pressure of the vapor, and some condensation may occur.
  - **Condensation**.The wet vapor then enters a condenser where it is condensed at a constant temperature to become a saturated liquid.
- One of the principal advantages the Rankine cycle holds over others is that during the compression stage relatively little work is required to drive the pump, the working fluid being in its liquid phase at this point.
- By condensing the fluid, the work required by the pump consumes only 1% to 3% of the turbine power and contributes to a much higher efficiency for a real cycle.

Physics 202 Lecture 6

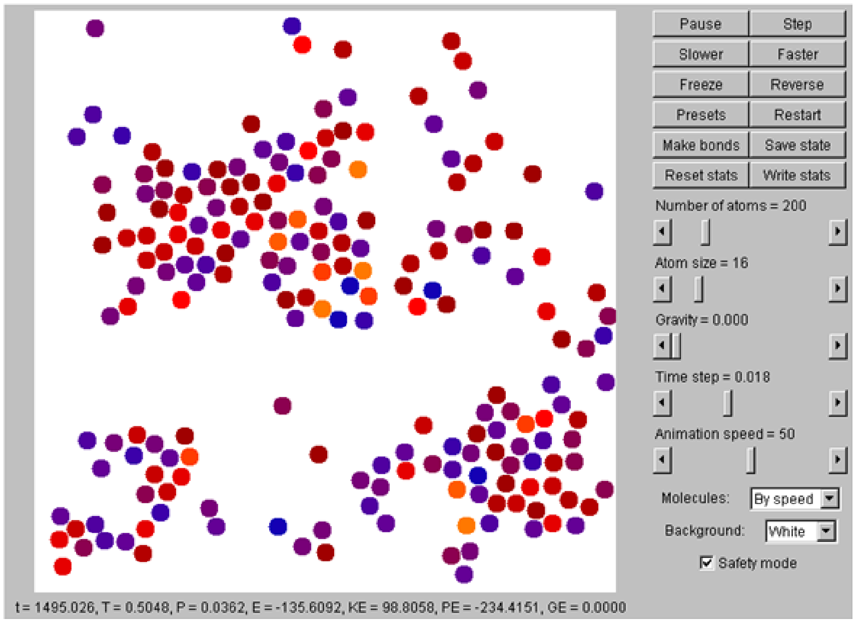
Kinetic Theory

- In the previous lectures we have investigated some of the basic properties of heat and temperature by focusing on what heat "does": how to measure it, how it flows, etc. In this lecture we focus on what heat "is": the random motion of molecules.
- Richard Feynman has said,

  > If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis that...
  >
  > > All things are made of atoms—little particles that that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another.
  >
  > In that one sentence, you will see, there is an enormous amount of information about the world, if just a little imagination and thinking are applied.

- Well, today we will think a bit about this in the context of the **kinetic theory of gases**. The basic idea is that the higher the temperature of a gas, the more kinetic energy the molecules possess.
- We can even incorporate phase changes into this framework by associating latent heat with the potential energy created through inter-molecular interactions.
- A truly phenomenal website with a simple molecular dynamics applet is located at `http://physics.weber.edu/schroeder/software/MDApplet.html`

Molecules and Moles



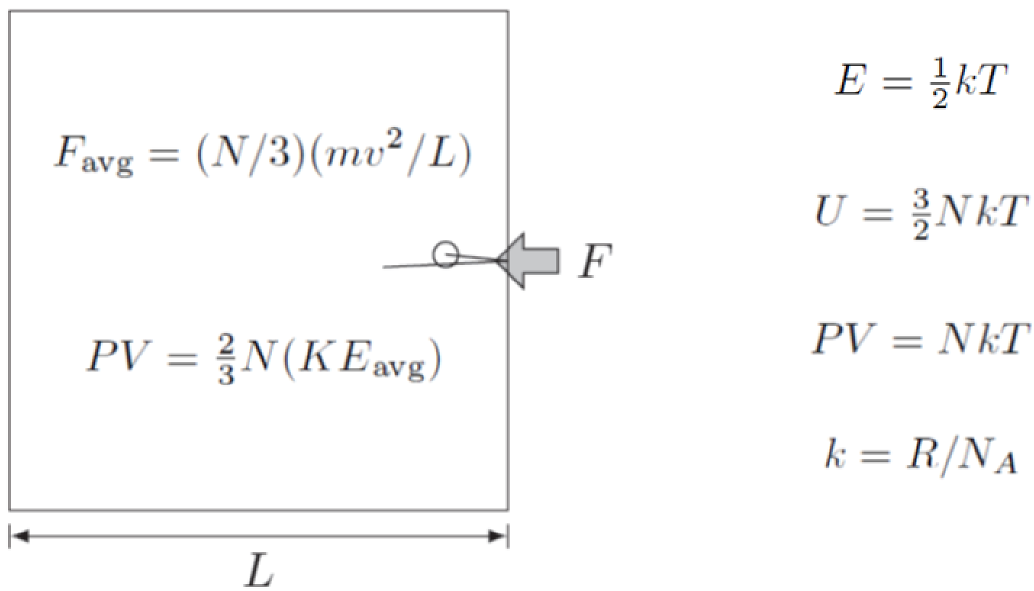
http://physics.weber.edu/schroeder/software/MDApplet.html

- Before we can speak in detail about molecular motion, we need to deal with their size. The radius of a typical atom is on the order of the nanometer. This means that the number of atoms in any laboratory sample is huge. It also means that they are essentially unobservable without highly specialized equipment.
- The fundamental unit of the mole has been developed to deal with this atomic size. By definition, the mole is the number of atoms in 12 grams of carbon-12 (a specific isotope of carbon) and has a value of 6.022×10^{23}. This number is called Avagadro's number labeled N_A.
- You may wonder why we need a special unit for something that is just a number. Isn't it just a naming convention, like a dozen eggs?
- In a fundamental sense that is true, but one must remember that we are human. No matter how sophisticated our machinery becomes, ultimately measurements must be translated to and from this human scale. There is no way to literally count the number of atoms in a particular sample.
- Maybe with the advances in nanotechnology this will be possible. If that day ever comes, we will still need to define the counting process in a way that is reproducible and independent of external variables.
  - In other words, the mole is not just a number that we simply arbitrarily choose—it is the relationship between the lab and the atomic world.
- This relationship is best summarized in the following formula:

$$\frac{\text{grams}}{\text{mole}} = \frac{\text{molecular mass}}{\text{molecule}}$$

where the molecular mass is the mass of each molecule in so called atomic mass units—listed on any periodic chart.

The Ideal Gas Law, Again



- There is a general result from classical statistical mechanics called the **equipartition principle**: when a system is in thermal equilibrium, each degree of freedom (or mode of energy) has an average energy of

$$E = \frac{1}{2}kT$$

  where k is Boltzmann's constant equal to 1.381×10^{-23}.

- In kinetic theory, we model an ideal gas as simple molecules with negligible interaction and size. The only mechanism to distribute the internal energy is through elastic collisions. Since there is no potential energy, the equipartition principle states that the internal energy is distributed to each molecule in all three dimensions so that

$$U = \frac{3}{2}NkT$$

  where N is the number of molecules.

- From Newton's second law, the average force of these randomly distributed molecules on a surrounding wall is

$$F_{\text{avg}} = (N/3)(mv^2/L)$$

- Since pressure is average force over area, we may write this in terms of the average molecular kinetic energy:

$$PV = \frac{2}{3}N(KE_{\text{avg}})$$

- Realizing that all of the internal energy of an ideal gas is kinetic, we can use the previous expression for internal energy to get the ideal gas law:

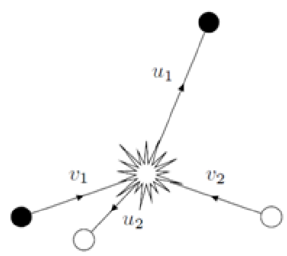
$$PV = NkT$$

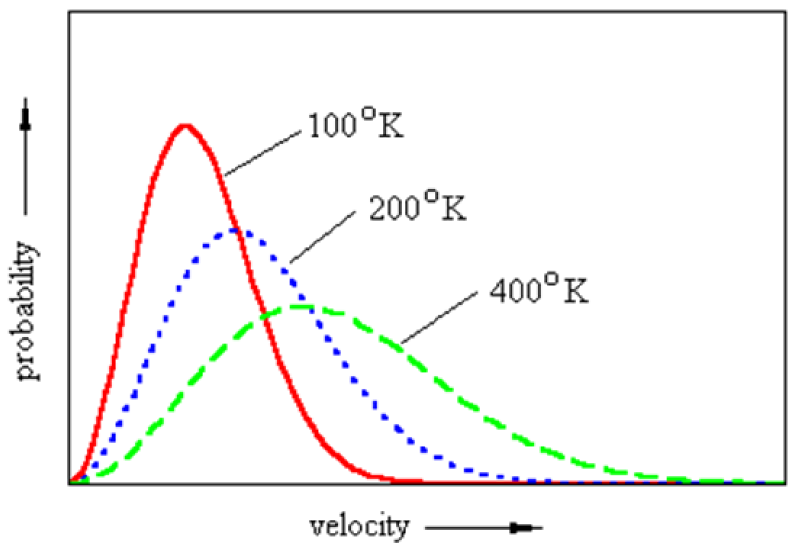
  which also gives us another relationship between the lab and the microscopic: $k = R/N_A$.

Maxwell-Boltzmann Distribution

$$v_{\text{avg}} = \sqrt{3kT/M}$$

$$N \propto v^2 \exp(-\tfrac{1}{2}mv^2/kT)$$





- According to the equipartition principle, the average molecule in an ideal gas has a kinetic energy of $\frac{3}{2}kT$. Therefore the average speed of these molecules is
$$v_{\text{avg}} = \sqrt{3kT/M}$$
where M is the molecular mass.

- But this is just an average. There is a natural spread of these speeds: the molecules are all moving with different speeds, some fast and some slow.

- This spread in speed helps explain evaporation. Even in a liquid, there is always a certain (small) percentage of molecules with enough energy to leave the attractive forces that hold the liquid together. These molecules escape as gas—the higher the temperature the more are able to leave.

- Also, the random collisions create a certain distribution. Even if we could somehow start with all the molecules at the same speed, the collisions would tend to spread them out. The distribution adjusts until a certain equilibrium distribution is reached.

- This distribution is based on a fundamental fact from kinetic theory. When a system is in thermal equilibrium, the number of molecules with a certain energy is proportional to $\exp(-E/kT)$.

- Since the molecular energy in an ideal gas is simply $E = \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2)$, we can factor the exponential into three components. Each of these vector components have a classic ("Gaussian") bell-shape curve.

- Calculating the overall speed distribution yields
$$N \propto v^2 \exp(-\tfrac{1}{2}mv^2/kT)$$

Molecular Entropy

E_1	E_2	Ω_1	Ω_2	Ω
0	5	0	25	0
1	4	1	16	16
2	3	8	9	72
3	2	27	4	108
4	1	64	1	64
5	0	125	0	0

$$\Omega \propto E^{d/2}$$

$$S = k \ln \Omega$$

- Our experience with time is very directional. Watching a film in reverse is hilarious because it looks so ridiculous. Yet Newton's laws of mechanics do not prefer one direction in time versus the other—projectile motion in reverse (without the friction) still works.
- The only place in physics where the direction of time seems to matter is the Second Law of Thermodynamics. It would seem we may be able to uncover the "**arrow of time**" by learning why entropy increases in kinetic theory.
- Recall that entropy measures the distribution of internal energy. Distribution across what? Across the degrees of freedom. The **microstate** of the system refers to its particular distribution of energy across its degrees of freedom.
- For an ideal gas, the energy is distributed throughout the velocities of the molecules as kinetic energy. Therefore the number of microstates is proportional to the square root of the energy. The total number of microstates for the system is the product of all these degrees of freedom. Thus,

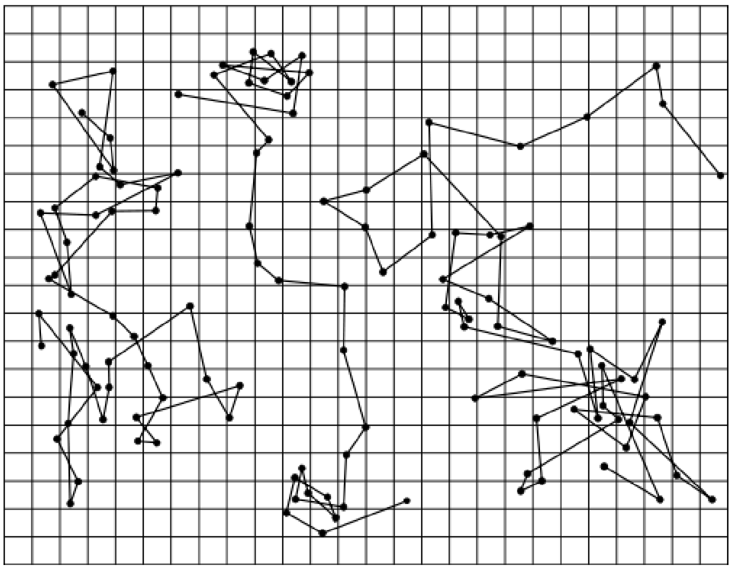
  $$\Omega \propto E^{d/2}$$

  where Ω counts the number of microstates for the energy and d is the number of degrees of freedom in the system.
- The number of ways energy can be distributed uniformly across the degrees of freedom is magnitudes larger than any other way. A fundamental postulate of kinetic theory is that each microstate is equally probable. From *this* it follows that a non-uniform distribution is astronomically improbable for a macroscopic system.
- Boltzmann realized that the connection between entropy and these microstates is simply

  $$S = k \ln \Omega$$

- Entropy increases because the system always moves to maximize the number of microstates available for its energy consistent with its equation of state. Boltzmann was so proud of this formula that it is engraved on his tombstone.

Diffusion and Brownian Motion

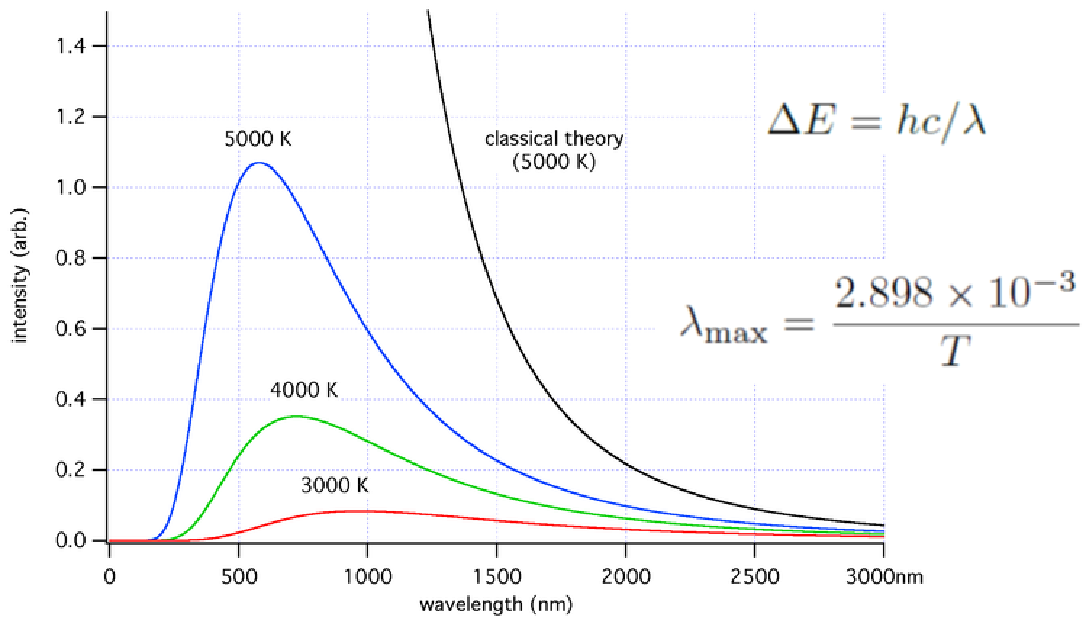


$$\frac{M}{t} = DA\frac{\Delta C}{\Delta x}$$

$$D = kT/6\pi\eta r$$

- Diffusion is the process in which material flows from a region of high concentration to low concentration. The substrate through which the flow occurs could be solid, liquid or gaseous. Our sense of smell depends on diffusion.
- The mechanism of diffusion is easy to imagine using kinetic theory. Some of the molecules from the high concentration region have a random motion away from that region. Though molecules are leaving the low concentration area also, they do so at a smaller rate. More flow in than out. This is another manifestation of entropy—but in this case the system moves to distribute its mass uniformly.
- The equation governing diffusion is Fick's law:
$$\frac{M}{t} = DA\frac{\Delta C}{\Delta x}$$
where M/t is the amount of mass transported per second (a.k.a., mass flow rate), ΔC is the concentration differential across the distance Δx through the cross-section area A. The quantity D called the diffusion constant.
  - Notice the similarity to thermal conduction. It is also similar to Ohm's law of electric current. Even Pouiselle's law of viscous flow can be rewritten in this way. This formula describes any flow through a resistive medium.
- Diffusion also offers us a view into the molecular realm through a phenomena called **Brownian motion**: the jostled motion of a small particle suspended in fluid. Around 60 BC, Lucretius noticed this motion in dust suspended in air and correctly attributed this to the bombardment of air molecules.
- Brownian motion was the subject of one of the 1905 "miracle year" papers of Einstein. He was able to show from kinetic theory that the diffusion constant is
$$D = kT/6\pi\eta r$$
- This analysis was one of the first to provide direct physical evidence of atoms through kinetic theory.

The Ultraviolet Catastrophe



- Imagine a box full of electromagnetic radiation. It's not hard: your bedroom is an example. Or better yet, an oven.
- Once equilibrium is reached, we can apply thermodynamic principles to analyze the radiation spectrum in the box.
- When trapped in a box, the number of ways that the electromagnetic energy can exist is constrained (we will see this later—they are called "standing waves"). These are the degrees of freedom for the EM field.
- The geometry is important for the details but at a basic level this number is proportional to square of the frequency of the waves. If we apply the equipartition principle, we expect each of these degrees of freedom to have equal amounts of energy proportional to kT. This result is called the **Rayleigh-Jeans law**.
- This is the prediction of classical physics and it is completely false. Something is fundamentally, powerfully, and absolutely wrong. This is called the **ultraviolet catastrophe** because the blue/violet end of the spectrum is higher frequency. Hot things glow red, not blue!
- In 1893, a result known as Wien's Law was empirically discovered for the radiation peak:

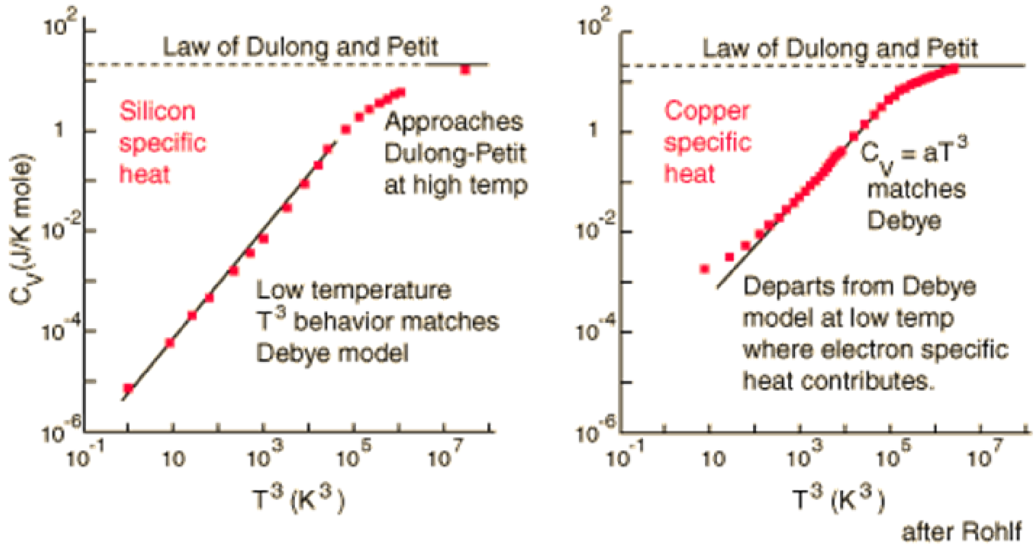
$$\lambda_{\text{max}} = \frac{2.898 \times 10^{-3}}{T}$$

- Later, in 1900, Planck uncovered the correct mathematical curve for the radiation spectrum. He could only explain the curve only by assuming the EM radiation was transferred only in discrete units equal to

$$\Delta E = hc/\lambda$$

  with h equal to 6.626×10^{-34} now known as Planck's constant. ($c = 3 \times 10^8$ is the speed of light.)
- Classical physics cannot justify this assumption, but it does explain why the lower wavelengths are less probable—if $\Delta E > kT$, these degrees of freedom cannot be "activated". They are said to be **frozen out**.

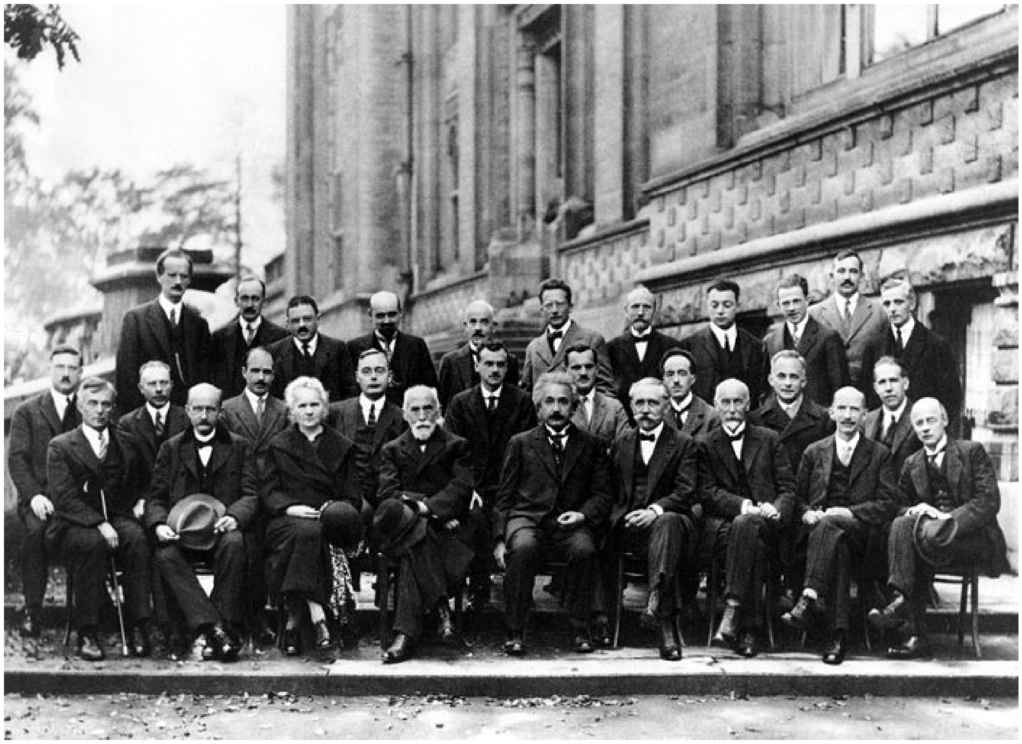
Specific Heat: Equipartition Fails Again



$$\cancel{C/n = 3R \sim 25}$$

- There is another area in which classical thermodynamics and kinetic theory fail to describe reality. The specific heat of solids. This is a less obvious conflict but illustrates the issue more directly than the ultraviolet catastrophe.
- In a solid, the molecules are locked into their relative positions. Each molecule is trapped in a stable equilibrium by the forces of the other molecules surrounding it.
- When a system is in equilibrium, its potential energy can be approximated by $\frac{1}{2}kx^2$ where k is proportional to the strength of the surrounding forces and x is its displacement from equilibrium (we will see why next lecture).
- This means that the energy for a molecule in a solid has six degrees of freedom: three for its motion (kinetic energy) and three for its position (potential energy). Just like the ideal gas, the equipartition theorem says that each degree of freedom should have $\frac{1}{2}kT$ units of energy.
- So the internal energy of a solid in thermal equilibrium ought to be $U = 3NkT$. For this reason, we expect the heat capacity of any solid to be simply $3R$ per mole.
- This works okay for room temperatures. But as the temperature drops below 200 kelvin or so the prediction gets worse and worse. The heat capacity falls and the internal energy becomes more and more sensitive to temperature differentials. This is why reaching absolute zero is so difficult.
- Einstein was able to explain the discrepancy with a "back-of-the-envelope" calculation using Planck's formula to show that the degrees of freedom are begin frozen out just like with the EM spectrum. Debye later cleaned up the calculation to get an exact match with experiment.
- Actually, we have an even more immediate example of the same thing: the air we breathe. A diatomic ideal gas has seven degrees of freedom because it can store energy in rotation and vibration as well as motion. But the vibrational modes are "frozen" at temperatures less than 1000 kelvin. This is why the heat capacity is only $\frac{5}{2}nR$.

The Birth of Quantum Mechanics



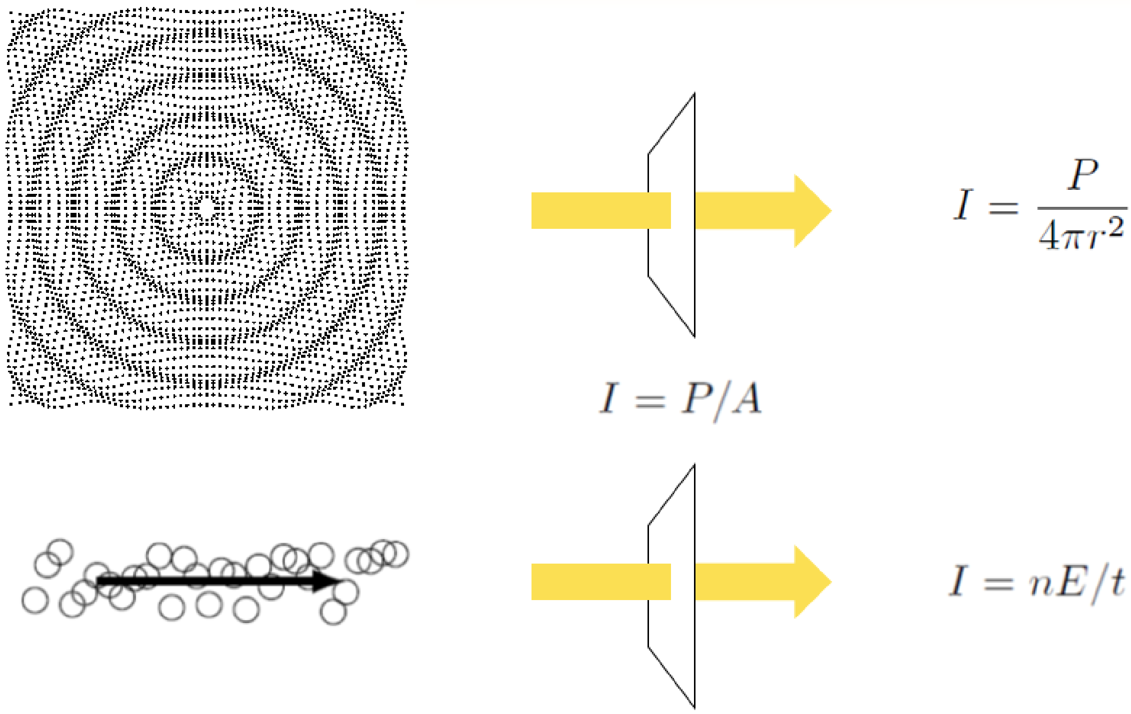
- So around the beginning of the 20th century, thermodynamics was paving the way to quantum mechanics. Technology had also advanced enough to begin to investigate the nature of the atom more directly.
- Collecting and making sense of these results was a complicated collaborative process. Traditionally, the year 1927 is used to mark the completion of the non-relativistic quantum theory.
- Every abstract mathematical tool from classical mechanics was thrown on the table. Over time physicists have combined these tools in different ways and in different patterns to reproduce the same theory. The first combination that worked is called the **Schroedinger equation**. Every other "remix" of quantum theory is equivalent to this.
- The basic idea is that at a fundamental level, elementary particles are not particles at all—they are waves. This wave nature is only evident at the microscopic level due to the small value of Planck's constant.
- But once we get down to the micron level or so (one-thousandth of a millimeter), we begin to see it. When we get to the nanometer level or so, we cannot avoid it. We need quantum mechanics to explain...
  - Physical chemistry: quantum mechanics explains the covalent bond and the electron shells of the atom.
  - Semiconductors: both why certain elements are semiconductors and how to manipulate them to create certain electrical properties.
  - Lasers: how the electromagnetic field "lines up" to create laser light far more powerful than ordinary light.
  - Low temperature physics: including superconductivity and superfluidity.
  - The existence of the atom itself: according to classical theory, the orbiting electrons ought to radiate energy, spiral inward and collapse within nanoseconds.

Physics 202 Lecture 7

Radiation: Particles, Waves, Rays

- Radiation can be divided into two classes: that which transports mass and that which transports energy.
- An example of the first class is nuclear radiation. When nuclear energy is released, each atomic nucleus acts like a little bomb. The atomic particles go flying in all directions carrying residual kinetic energy. This atomic shrapnel we call radiation. We will see later that there are three types of natural nuclear radiation each with varying properties.
- The second class of radiation is that which transports only energy. This is the kind we refer to when a pebble is dropped in a still pond and we say that the ripples radiate from the impact. This radiation does real work—each wave erodes the shoreline a tiny bit—so energy is truly being transported.
- Some of the concepts in this lecture are applicable to both kinds, but as we progress we will become more interested in the second class.

Energy, Intensity, and Current



- When it comes to radiation and waves, we are interested in measuring the flow of energy. We have already discussed the fluid flow and even heat flow. Mathematically, analyzing the flow of energy is similar.
- We define the "energy flow rate" as the amount of energy that passes through a given surface area per second. More commonly this is called the **intensity** of the radiation.
- And since power is energy per second it is most common to define intensity simply as

$$I = P/A$$

  - In the context of electromagnetic radiation, the letter S is used to represent this intensity and is sometimes called the "energy flux density".
- For a uniform point source of radiation, the intensity is symmetric so it decreases with the distance squared:

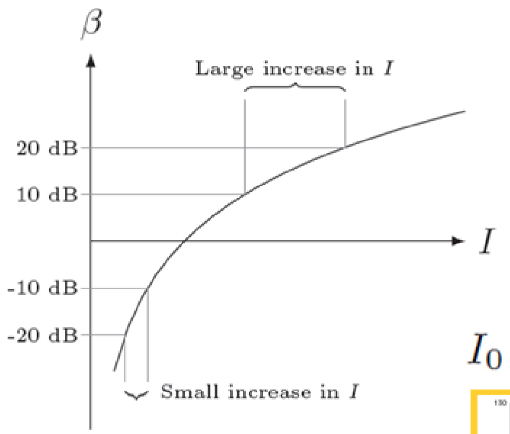
$$I = \frac{P}{4\pi r^2}$$

- For material radiation, the intensity is transported through the kinetic energy of the particles. So the intensity is simply the number of particles through the area per second (n/t, sometimes called the particle current) and the average energy of the particles:

$$I = nE/t$$

- We will have to wait a bit to see how to calculate the energy flow of the second class of radiation.

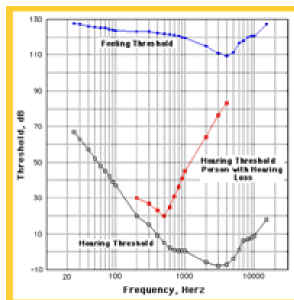
Measuring Intensity

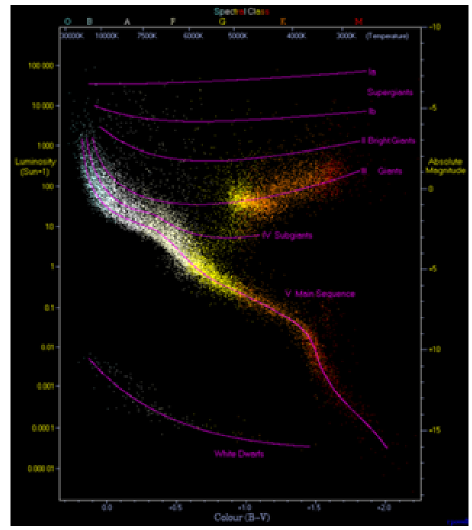


$$\beta = 10\log_{10}(I/I_0)$$

$$I = 10^{(0.1\beta-12)}$$

$$I_0 = 1.0\times 10^{-12}$$

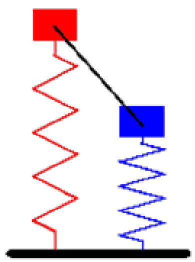


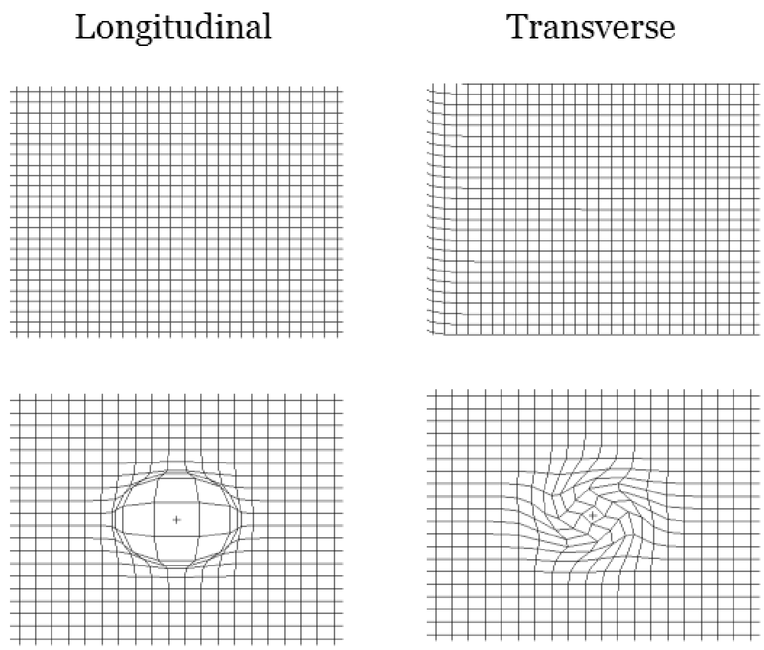


$$m = -2.5\log_{10}(I/I_0)$$

- Our senses of sight and hearing are "instruments" that measure the intensity of light and sound, respectively. However, our senses are more sensitive to small intensity than large intensity—it is not linear. Both sound and light have logarithmic scales to mimic our experience better.
- For sound, we use the **decibel** scale. We have
$$\beta = 10\log_{10}(I/I_0)$$
where the reference intensity is $I_0 = 1.0\times10^{-12}$ watts per square meter—which is the threshold for human hearing.
- To distinguish this from intensity, this β-value is called the "intensity level". But usually it is informally referred to as the decibel level.
- Solving problems with the logarithm sometimes throws some people. There are three facts to remember:
$$y = \log_a(x) \implies x = a^y \qquad \text{and} \qquad \log(ab) = \log(a) + \log(b) \qquad \text{and} \qquad \log(a^n) = n\log(a)$$
- Using these rules, one can show that the inverse to the decibel equation above is
$$I = 10^{(0.1\beta-12)}$$
- For light, the story is more complicated due to the mechanics of the eye: sensitivity of rods and cones to various wavelength, variable pupil and lens, luminosity, candelas, etc.
- But in astronomy, a scale exists which has roots back to antiquity. The **apparent magnitude**:
$$m = -2.5\log_{10}(I/I_0)$$
where I_0 is the brightness of the star Vega. A magnitude 6 star is just barely visible to the naked eye.

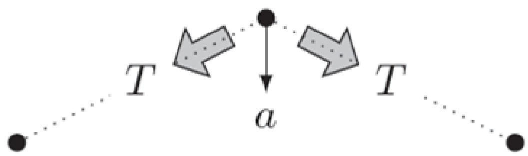
Coupled Oscillators and Waves





- So far I have tried to keep the discussion general to incorporate both kinds of radiation. We are now ready to focus on the second class: the mechanics of wave motion. For a moment, I would like to discuss waves in a very abstract way. It is surprising how wide the application of these ideas are.
- A wave requires a source and a **medium** which does the vibrating. The nature of the medium dominates the character of the wave motion.
  - There are many different types of media: strings are the simplest model, but we will also talk about the way waves move through solids and fluids. The electromagnetic field too, although it is a bit of a special case.
- In order to support the transmission of energy, the medium must be composed of interconnected parts. This is the crucial ingredient and makes the wave motion possible—when one part moves, the others are dragged with it.
- Although the medium itself does not move in bulk, there is a direction involved in this radiation of energy. We say that we have a **transverse wave** if the oscillation in the medium is perpendicular to the movement of the disturbance (like a string). We have a **longitudinal wave** if the oscillation is parallel to the radiation (like a compressed spring, or sound waves).
  - Wave motion is not confined to either/or: water waves have both transverse and longitudinal components.
  - In addition, the disturbance need not be a physical displacement. We could speak of a heat wave, for example. As long as the medium is coupled across this property, radiation will occur.
- A longitudinal wave has two degrees of freedom (the y and the z if it is moving along the x-direction). To completely describe the wave we need to specify the orientation of its displacement. This is the **polarization** of the wave.
  - It is a bit more complicated than this since we could drive the string with a circular type of motion. In that case, the direction of the oscillation rotates in time: this is called **circular polarization**.

The Speed of Wave Radiation



$$v = \sqrt{T/\mu}$$

Solid

$$v = \sqrt{Y/\rho}$$

Liquid

$$v = \sqrt{B/\rho}$$

Gas

$$v = \sqrt{\gamma k T/M}$$

Air

$$v = 331 + 0.6T_C$$

- Most of the essential wave phenomena can be seen in a simple string. The essential characteristic of the wave medium is that it be composed of coupled oscillators all near equilibrium.
  - In a string, the net force on an element acts to eliminate curvature in the string.
  - In general, when the forces of restoration in a medium are proportional to the curvature of the disturbance, we say the medium obeys the **wave equation** and will support the kinds of waves we are discussing now.
- Because our oscillators are connected, when one is displaced it pulls its neighbors with it and therefore does work. This has a dampening effect on the original as the energy is propagated away.
  - The rate at which this energy is lost is called the **impedance** of the medium.
- The speed with which this disturbance propagates depends on two characteristics of the medium: its coupling and its inertia. For a string, the formula is

  $$v = \sqrt{T/\mu}$$

  where T is the tension in the string and μ is its mass per unit length.
- Similar formulas follow for the speed of sound in a solid, liquid and gas. They are

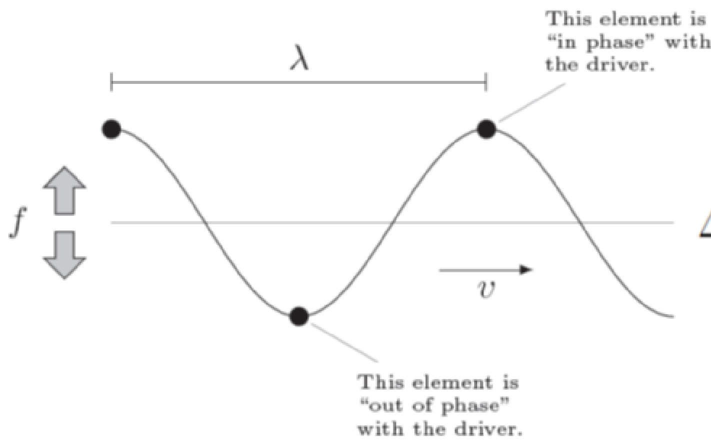
  $$v = \sqrt{Y/\rho} \quad \text{and} \quad v = \sqrt{B/\rho} \quad \text{and} \quad v = \sqrt{\gamma k T/M}$$

- For air we can use the approximation

  $$v = 331 + 0.6T_C$$

  where T_C is the temperature in Celsius.

Frequency, Phase, and Wavelength



$$\Delta\phi = -(2\pi f/v)\Delta x$$

$$\psi = A\cos\left(2\pi f t - \frac{2\pi}{\lambda}x\right)$$

$$v = f\lambda$$

- In order to support a sustained disturbance, a source must drive the oscillation. Energy is put into the system by the source and drained away by radiation through the medium.
- This energy input is driven at some constant frequency f. Energy moves away and the all other elements of the medium vibrate with the same frequency.
  - Remember: this is what happens with driven oscillations. The driver controls the frequency regardless of the natural frequency of the system.
- Since the element of the medium are separated in space, it takes a bit of time for this radiated energy to do its work. As a consequence, the elements suffer a phase shift relative to the source. The formula is

  $$\Delta\phi = -(2\pi f/v)\Delta x$$

- If the source executes simple harmonic motion, the wave profile will also be sinusoidal. The distance between the peaks of the wave is called its **wavelength**. The speed, frequency and wavelength are related via

  $$v = f\lambda$$

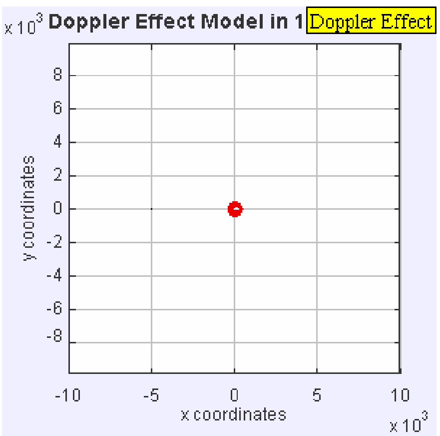
  An equation that is true for all waves whether string, sound or light.
- Using these ideas, we can write the formula for the wave profile as follows:

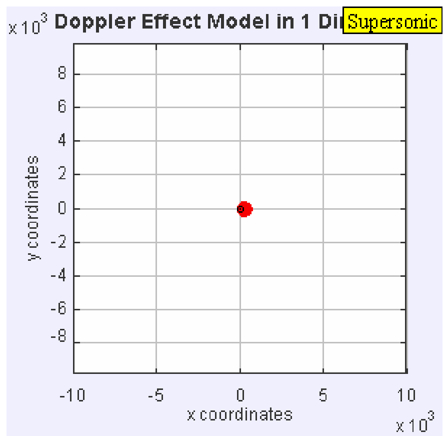
  $$\psi(t) = A\cos(\omega t - kx)$$

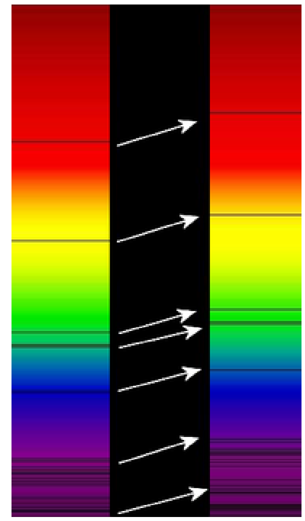
  where $\omega = 2\pi f$ and $k = 2\pi/\lambda$. This k-value is sometimes called the "wave number" since it is the number of full waves per meter.

The Doppler Effect

$$f\lambda = v_r \pm v_s$$







$$f_o = f\left(\frac{v_r}{v_r \pm v_s}\right) \qquad \sin\alpha = v_r/v_s$$

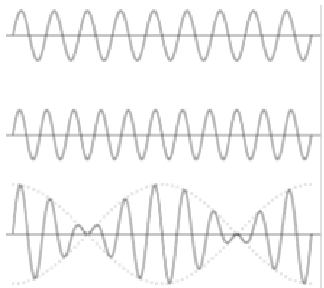
- All of these considerations assume the source is at rest relative to the medium. If the source is moving, the **Doppler effect** will change the observed frequency of the radiation.
- From the viewpoint of the moving source, the medium is flying by. This apparent "wind" affects the speed of the wave transport. Because the speed is different, the wavelength is different according to $f\lambda = v$.
  - The geometry is complicated, so we usually only talk about two special cases: directly in front of the source or directly behind. We have $f\lambda = v_r \pm v_s$ where v_r is the natural speed of the radiation and v_s is the speed of the source. Use the plus sign behind the source and the minus sign in front of the source.
- The observed frequency is therefore
$$f_o = f\left(\frac{v_r}{v_r \pm v_s}\right)$$
- If the speed of the source is close to the radiation speed, a shock wave begins to build where the intensity accumulates in front of the moving source. If the source moves faster than sound we get a **sonic boom**. A high-pressure cone of generated with an angle α given by
$$\sin\alpha = v_r/v_s$$
  - For light, we have to take relativity into account—we cannot get the source to move faster than light.
- Applications of the Doppler effect include
  - Radar guns to measure the speed of vehicles or wind speed in a hurricane.
  - Ultrasound technology can be used as a non-invasive way to measure blood flow.
  - The relativistic Doppler effect is used to estimate the motion of distant stars.

Beats and Fourier Analysis

$$\psi_1 = \cos(2\pi f_1 t)$$

$$\psi_2 = \cos(2\pi f_2 t)$$

$$\psi = \psi_1 + \psi_2$$

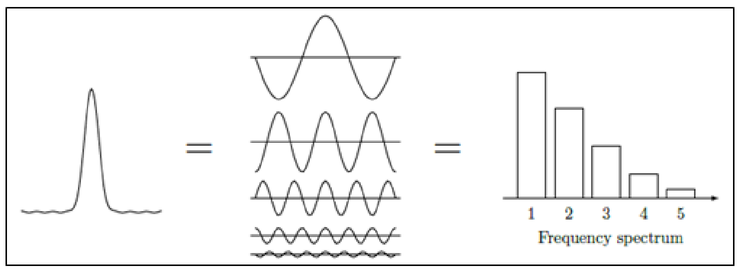


$$\psi(t) = A(t)\cos(2\pi f_{\text{avg}} t)$$

$$A(t) = 2\cos(\pi \Delta f t)$$

$$f_{\text{beat}} = \Delta f = f_1 - f_2$$

Fourier Analysis

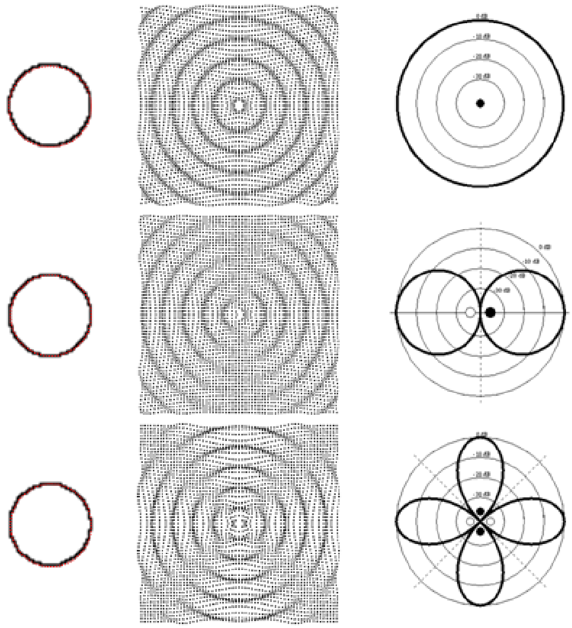


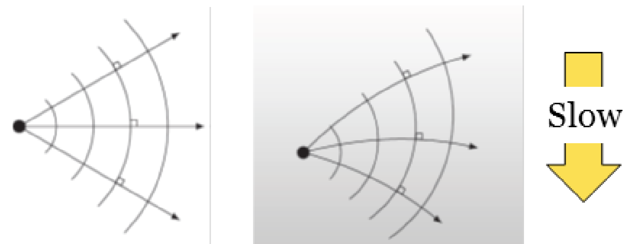
- When two sources act to vibrate a part of the wave medium, in a sense they "compete" for the final motion. This is called **interference**. If the medium obeys the wave equation then the result of this interference is the "superposition" of the vibration from each source individually.
  - This is just a fancy way of saying that the result is the simple sum of the contributions.
- For example: if we have two sources this slightly different frequencies, the combined effect is a simple sum which can be rearranged (using some trig identities) into:

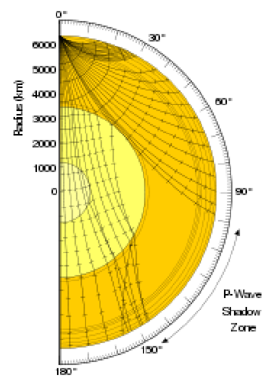
$$\psi(t) = A(t)\cos(2\pi f_{\text{avg}} t) \qquad \text{with} \qquad A(t) = 2\cos(\pi \Delta f t)$$

  where f_{avg} is the average of the two frequencies and Δf is their difference.
- We can reinterpret this as a sinusoidal vibration of frequency f_{avg} with an amplitude that changes over time. The frequency of this change is called the **beat frequency**. You've probably heard this before with sound waves—it gives a characteristic "waa-waa-waa" sound.
- This ability to affect the amplitude is not restricted to two sources. We can generate more complicated patterns by adding more sources with different frequencies and amplitudes. The reverse is also possible: we can separate an arbitrary sound profile into multiple frequency elements—this is called **Fourier analysis**.
- One result of Fourier analysis is that the sharper the profile the more higher frequency contributions you must consider. This is why more data is required to digitize music with high-fidelity.
- A **wave packet** is a Fourier combination of waves that is localized in space. This localization requires a spread of frequencies (the distribution is determined by Fourier analysis). This relationship is the source for the Heisenberg uncertainty principle in quantum mechanics.

Waves and Rays







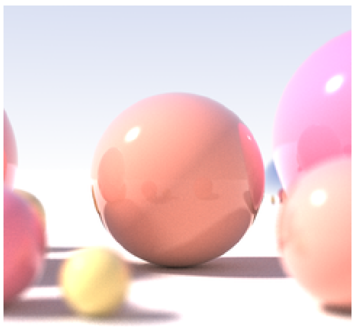
- We have yet to really touch upon wave motion in more than one dimension. Perhaps one of the most characteristic things about waves is their tendency to spread, like ripples in a pond. This is called **diffraction**
- The entire wave profile is a function of both space and time. In the one-dimensional case, we could keep track of the wave using its wavelength. But for more than one dimension there are two complementary ways approaches.
- The first uses the idea of a **wave front**. The wave front is the line or surface that connects all the elements in the medium that have the same phase in the wave. Each ripple in the pond is a wave front.
- These wave fronts can be seen as the "contour map" of the wave. The motion of the wave is perpendicular to these contours. The vector that points against these contours is called a **propagation vector**. Its magnitude is defined as $k = 2\pi/\lambda$.
- If we connect these propagation vectors together we track paths through the wavefronts. These paths are similar (mathematically) to the streamlines we discussed in fluid motion. Especially in optics, but also in general, these paths are called **rays**.
- For a simple point source, the wave fronts are circles (or spheres) and the propagation rays point away uniformly from the source. The fact that the rays spread indicates that the intensity drops with distance.
- When the wave speed varies in a medium, the wavelength changes according to $v = f\lambda$. If, for example, the speed is greater below, the wavelength is longer. This means that the wavefronts are farther apart and the rays diverge up. In general, the rays will bend into the area with slower speed. This is the **refraction** of the wave.

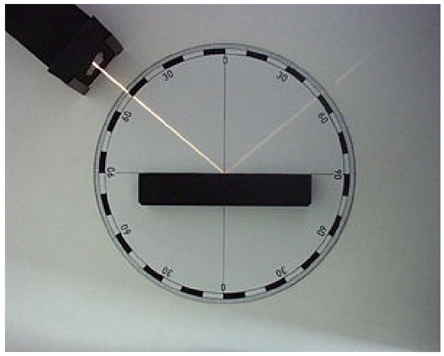
Physics 202 Lecture 8

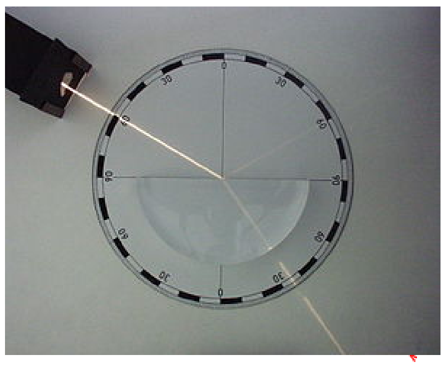
Geometric Optics

- Geometric optics is one of the five branches of physics that goes back to antiquity. Light is assumed to travel along straight lines. The Greeks understood reflection, but could never quite figure out the rules of refraction.
- Since the velocity of light is so large, "travel" may be a bit of an anachronism. Essentially, the large speed of light renders time unimportant, and optical problems reduce to geometry problems.
- However, in Newton's time, as people started to understand the mathematics of wave, a running debate began whether to consider light as a particle or wave. This was definitively decided in the early 1800s by Thomas Young in favor of wave theory.
- This deeper understanding gives us a way to determine the viability of the light ray idea: whenever the objects in the problem are much larger than the wavelength of light, we can safely use the assumption that light travels in rays.
- Although we will be speaking exclusively of light in this lecture, realize that these ideas actually apply to any wave phenomena, like sound.
- But the wavelength of sound is typically on the order of a few meters, so the objects involved have to be on the order of a kilometer to use the idea of a "sound ray". The wavelength of light is on the order of a micron—you can hear around corners, but cannot see around them.

Reflection and Optical Images

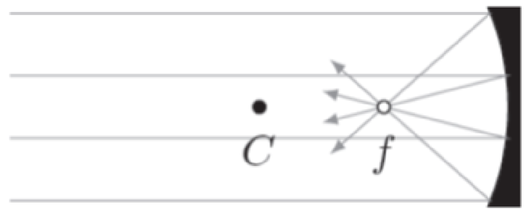




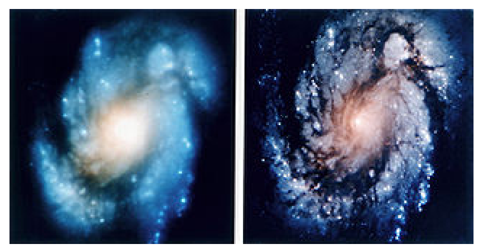


- Like was mentioned, the first rule of geometric optics is that light rays travel along straight lines in a homogeneous transparent medium. Point sources cast sharp shadows.
  - You may look around you and see many diffuse shadows. This is a consequence of the fact that there are multiple light sources around which all cast patterns of light and dark that conflict.
  - Even from a single source you may see a diffuse shadow. This is a consequence of the fact that the source is extended. Each part of the light source casts a different shadow.
- Some obstacles are reflective. If the surface is rough, **diffuse** reflection scatters the incoming light. The kind of reflection from a mirror is called **specular** reflection. From now on we will only consider specular reflection.
- The **law of reflection** states that the reflected angle of a ray of light is equal to the incident angle. In optics, we measure the angles from the surface normal—this doesn't make a difference now, but it will when we talk refraction.
- Frequently in these optical problems, the key will be to locate the **optical image**. This is where the reflected rays of light *appear* to originate. We take the path lines of the reflected rays an extend them backward until they intersect.
  - When you look in a mirror this is the spot behind the mirror where your reflection appears to be. For a straight mirror the distance to the image behind the mirror is equal to the distance to the object in front of the mirror.
- When we have multiple mirrors, we can treat the location of the image from the first mirror as the object for the second mirror, etc.

Focusing Light



$$f = R/2$$

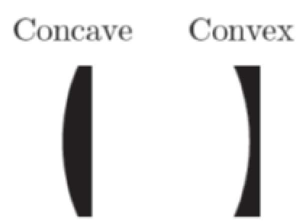


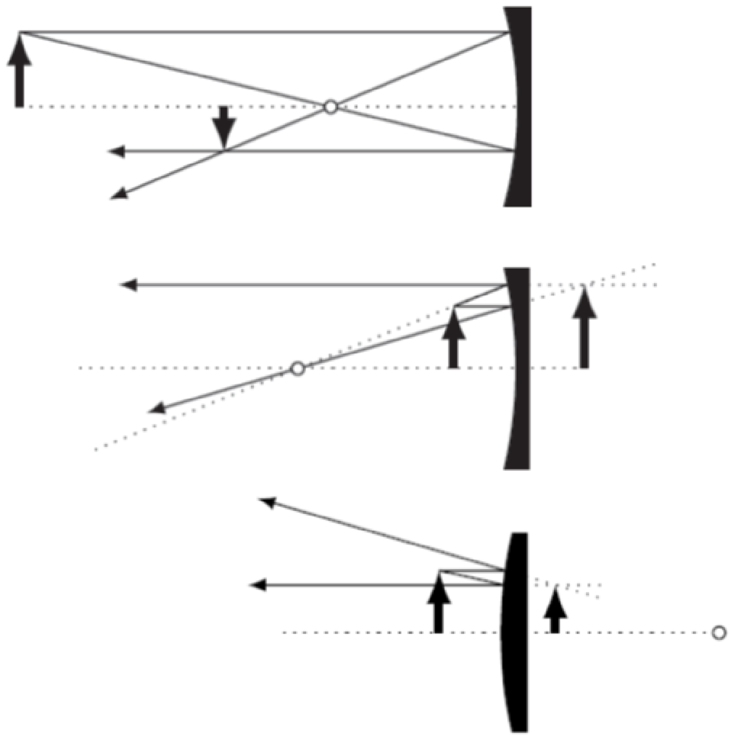
- Things start to get interesting when we talk about curved surfaces. A curved reflective surface will distort its image and when properly constructed, focus light.
  - There is a story about Archimedes doing just that. Supposedly he constructed a series of mirrors designed to focus sunlight on passing ship causing them to burst into flame.
- The ideal shape for the focusing of light is the parabola. This curve has the property that any ray coming in parallel to its axis will bounce and pass through the focus of the parabola.
- This focus is the image of all the incoming light energy and acts like a point source for the reflected light. It is called the **focal point** for the lens.
- It is much easier to produce curved mirrors with a spherical cross-section. For simple applications, a circular lens is sufficient. The focal length of a circular lens is half its radius

$$f = R/2$$

- This focal point can be found by tracing the path of two incoming parallel rays and finding where they intersect. This is called a **ray diagram**. The incoming parallel lines are called **paraxial rays**.
- Each ray diagram is essentially a problem in geometry. There is no time involved. As such, even though we think of the rays as being moving in a direction, if we reverse all the rays the geometry is the same.

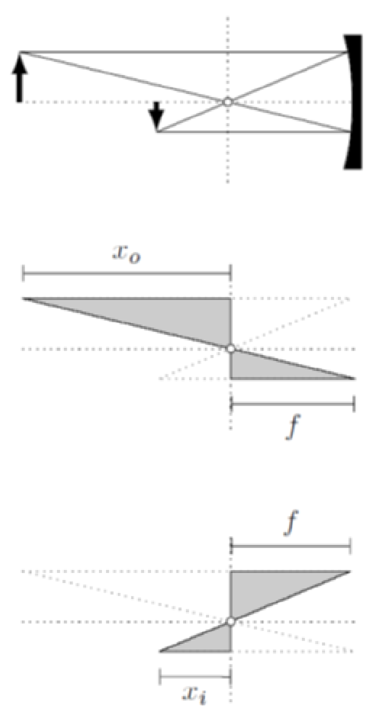
Images and Ray Diagrams





- When we draw out a ray diagram, we only need a few key rays to identify the image location. We are seeking the location where the lens bounces the rays to an intersection point. From this point the light rays diverge—just like a second source. This is the location of the image.
- At minimum, this intersection can be found with two lines. But there are three that are easy to determine. Assume the object is on the left of the curved mirror. The rays that leave the object along the principal axis will bounce right back, so all we need to do is find the image location from the top of the object. Imagine three rays:
  - The first runs parallel to the principal axis. It bounces off the mirror and passes through the focal point (by definition).
  - The second runs through the focal point. It bounces off the mirror and then leaves parallel to the principal axis. Again by definition but in reverse.
  - The third bounces off the center of the mirror. It bounces off with equal angle. This one can be drawn easily by flipping the object upside-down.
- There are two cases to consider: whether the object is outside or inside the focal length.
  - If outside, the image is **inverted** (upside-down) and reduced in size.
  - If inside, the rays diverge and do not intersect. But we can trace where the line appear to be diverging from. This is the image location, but we call this image **virtual** to emphasize that there is no actual light at this point. This virtual image is upright and magnified in size.
- So far we have only been discussing **concave** lenses which focus light inside them. A **convex** lens is the opposite—it diverges light and has a virtual focal point (on the other side of the mirror). The image from a convex lens is always upright and reduced in size.

The Lens Equations



$$x_o x_i = f^2$$

$$\frac{h_i}{h_o} = \frac{f}{x_o}$$

$$\frac{h_i}{h_o} = \frac{x_i}{f}$$

$$\boxed{\begin{gathered}\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \\ m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}\end{gathered}}$$

- Let's label the horizontal distance between the object to the focal point as x_o and similarly for the distance between the image and the focal point as x_i. Using similar triangles, it can be seen that we have the relationships

$$\frac{h_i}{h_o} = \frac{f}{x_o} \quad \text{and} \quad \frac{h_i}{h_o} = \frac{x_i}{f}$$

- We can combine these to get $x_o x_i = f^2$ which can be rearranged to yield the **lens equation**:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

  where d_o is the distance between the object and the lens, d_i is the distance to the image, and f is the focal length.

- Although we have derived this formula for the case of an object outside the focal point of a concave mirror, the equation will also work for an object inside the focal length. The image distance will be negative—which we interpret as to the right of the mirror and virtual.

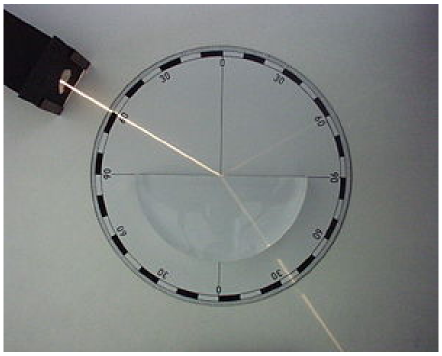
- In fact, the equation also works for a convex lens if we use a negative focal length. Obey these sign conventions and this equation can always be used.

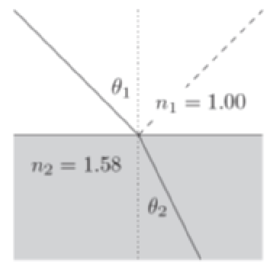
- For the size of image, we can combine the our previous formulas to derive the following formula for magnification:

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}$$

- Notice the negative sign. A negative magnification indicates the image is upside-down.

Snell's Law of Refraction

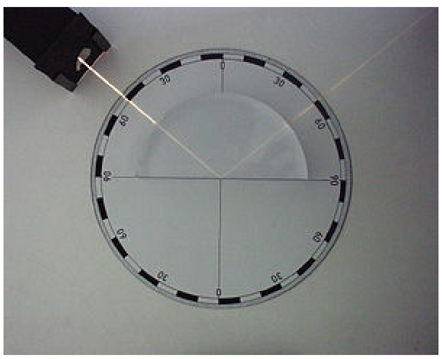


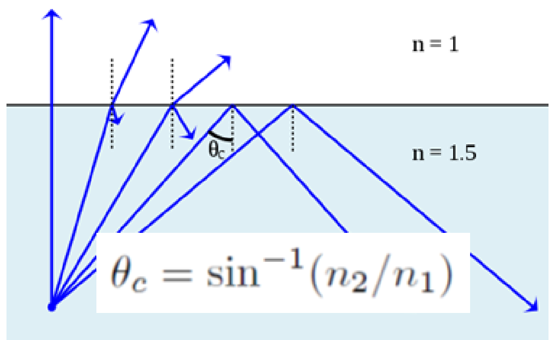


$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

$$v = c/n$$

$$c = 3 \times 10^8 \text{ m/s}$$





- Not all transparent media are the same. Snell's law describes what happens when a ray of light strikes an interface between two media. The ray refracts according to

  $$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

  where the angles are measured from the normal. Each n is called the **index of refraction** for the material.

- Previously we learned that waves refract when the radiation speed in the medium changes. This is true here also. The speed of light in the transparent medium is

  $$v = c/n$$

  where c is the speed of light in vacuum: $c = 3 \times 10^8$ m/s.

- A light ray will bend into a medium with higher index of refraction. Typical indexes of refraction are between 1 and 2. Air is essentially one. Plastics and glass are between 1.3 and 1.6. Diamond is 2.4.

- Of course, a light ray will bend out of high index of refraction also. If the index of refraction is high enough, the light ray will bend to 90°. This is called the **critical angle**:

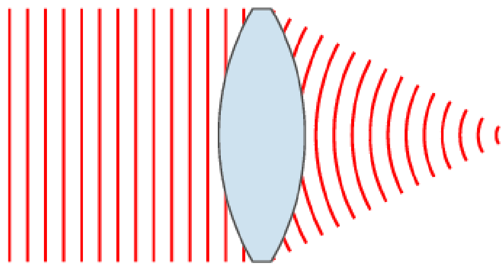
  $$\theta_c = \sin^{-1}(n_2/n_1)$$

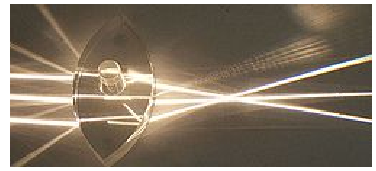
  where n_1 is the material the light is coming out of and n_2 is the material the light is going into.

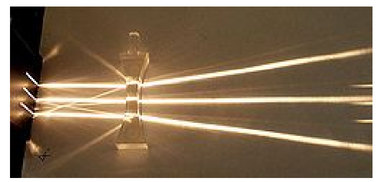
- In general, some incident light is reflected off a transparent surface and some is refracted. But any angle beyond the critical angle will cause **total internal reflection**.

- As we approach the critical angle, the intensity of the refracted beam falls and the reflected beam increases. When the critical angle is reached, the intensity of the refracted ray drops to zero and all the incident energy is reflected.

Transparent Lenses

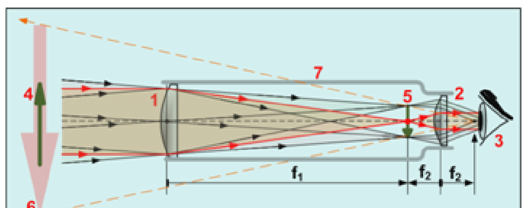






$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}$$

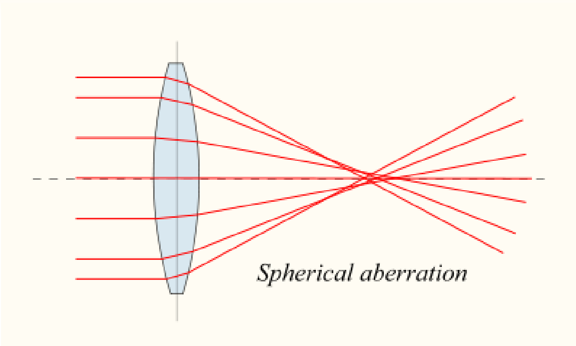


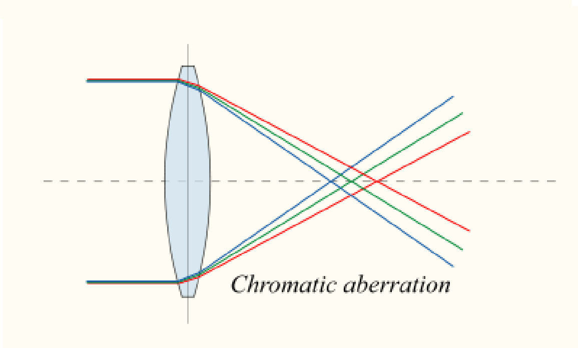
Schematic of **Keplerian** refracting telescope.

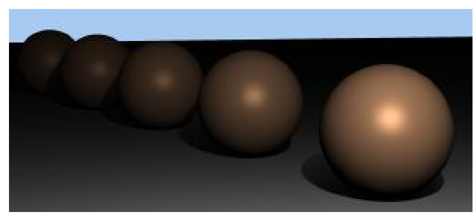
$$m = f_{\text{obj}}/f_{\text{eye}}$$

- Similar to mirrors, we can also use refraction to focus light.
- Also similar to mirrors, circular lenses are much easier to construct than any other shape. If the lens is thin, the geometry is identical to that with a mirror—except everything happens on the other side.
- We have the same lens formulas for transparent lenses also. But in this case, the concave lens is converging and a convex lens is diverging. The focal length of a concave lens is considered positive and image distances to the right are positive also. Thus:
$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$
and
$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}$$
- Combinations of lenses are analyzed by the rule that the image of the first lens is the object of the next, etc.
  - When the distance between lenses is small, the reciprocals of the focal lengths add. In optometry, the reciprocal of the focal length is called the **refractive power** of the lens. So, refractive powers add.
  - In a telescope, the object distances are effectively infinite. This means that the image from the objective lens is at the focal point. If the focal point of the eyepiece is aligned, the secondary rays will exit the eyepiece along parallel lines. The secondary image is effectively infinite but magnified. The magnification for the telescope is
$$m = f_{\text{obj}}/f_{\text{eye}}$$

Aberration, Dispersion, and Attenuation

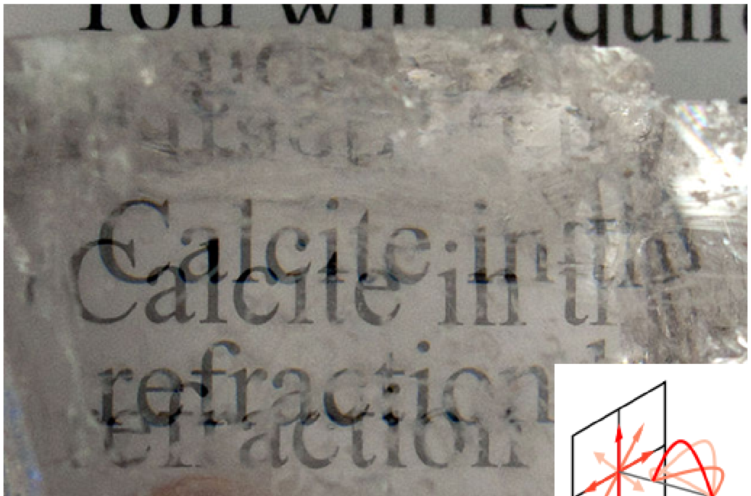


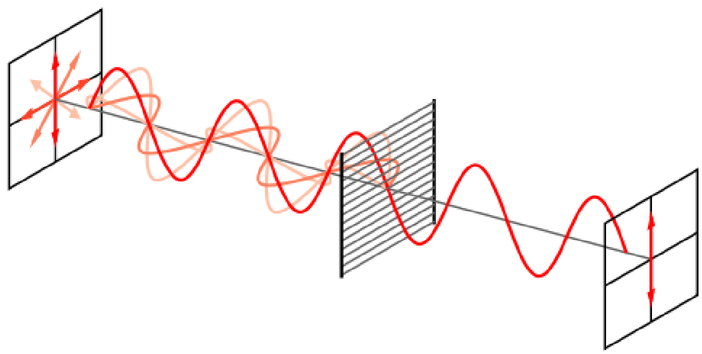




- **Aberration** is the recognition that no lens is perfect. This happens for three distinct reasons.
- The first has to do with the shape of the lens. But this is just an economic issue. It is possible to determine the required shape (e.g., parabola for a mirror), but the construction is more difficult.
- The second reason for aberration is **dispersion**. For a wave, dispersion is the frequency dependence of the wave radiation. The way this manifests in optics is that the index of refraction is dependent on the color of the light.
- Typically differences in color account for about a 2% change in the index of refraction. This leads to some of the most beautiful application in physics—e.g., the rainbow.
- But this also means that different colors of light focus at slightly different points in space. This type of aberration is called “chromatic”.
- It is sometimes possible to correct this kind of aberration by combining lens of different indexes of refraction to bring all the colors of light back together at the final focal point.
- Finally, we discuss **attenuation**. Technically this is not a source of aberration, but a recognition that the intensity of light falls as it travels.
  - Typically we ignore attenuation in geometric optics—though did see it in the context of the critical angle.
- Attenuation is a kind of friction (energy loss) in the movement of light. Physically what happens is that the light is scattered by the transparent medium. The intensity typically drops with an exponential decay.

The Problem of Polarization





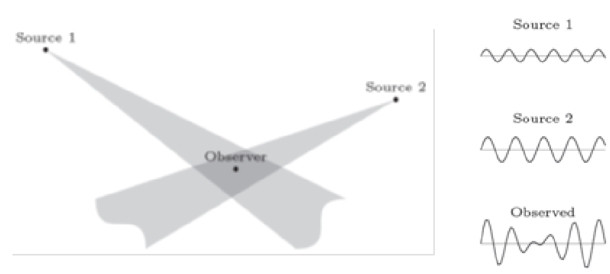
- I'd like to touch on one last topic before we leave geometric optics: the **polarization** of light.
- The polarization of light was first observed in the late 1600s via **birefringence**, also known as double refraction. When a ray of ordinary light passes through certain crystals, the ray splits in two. The ray must be incident at an angle to the so-called **optical axis** of the crystal.
- Essentially, the crystal has two indexes of refraction: one along its optical axis and another perpendicular to the axis. The reasons why the crystal splits the light are one thing—but at minimum it shows there is something more to light than intensity and color.
- The reason why light has this additional degree of freedom (if you will) can be explained with longitudinal wave motion. But we can describe (rather than explain) this double refraction in geometric optics with two different indexes of refraction.
- The index of refraction is said to be **anisotropic**—meaning that it is not rotationally symmetric, there is a "preference" for a particular direction (the optical axis).
- Since these problems involve all three dimensions, the geometry is a bit complicated. It is easier to use more sophisticated matrix algebra, etc. But don't worry: we won't go into any of that.

Physics 202 Lecture 9

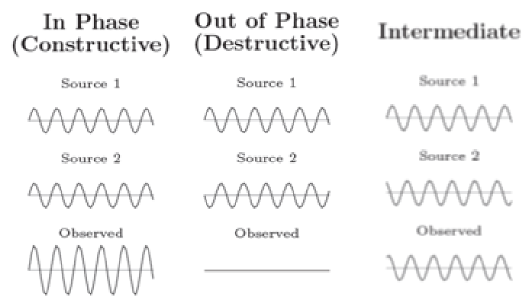
Wave Interference

- We saw in the previous lecture that we can go a long way understanding radiation using the idea of a ray. It is no coincidence that highly energetic electromagnetic waves include the word: X-rays, gamma rays, etc.
- However, we can only go so far. There are two aspects of waves that cannot be explained with mere rays: interference and diffraction.
- Of the two, the first is the offers the highest contrast to the more "particle-like" aspect of rays. Seemingly impossible things can happen through constructive and destructive interference.
- Again, the majority of our examples will involve light. But as I have tried to emphasize throughout the last few lectures, these wave phenomena also apply to strings, sound, and even the "matter waves" of quantum mechanics.

Phase Differentials Create Interference



$$\psi(t) = \psi_1(t) + \psi_2(t)$$



$$\Delta\phi = -(2\pi/\lambda)\Delta x$$

$$\Delta\ell = n\lambda$$

$$\Delta\ell = (n + \tfrac{1}{2})\lambda$$

- We loosely touched on the idea of superposition when we previously discussed beats and Fourier analysis. Let us start there again with a bit more precision.
- When the disturbances from two sources combine the instantaneous amplitudes add:

$$\psi(t) = \psi_1(t) + \psi_2(t)$$

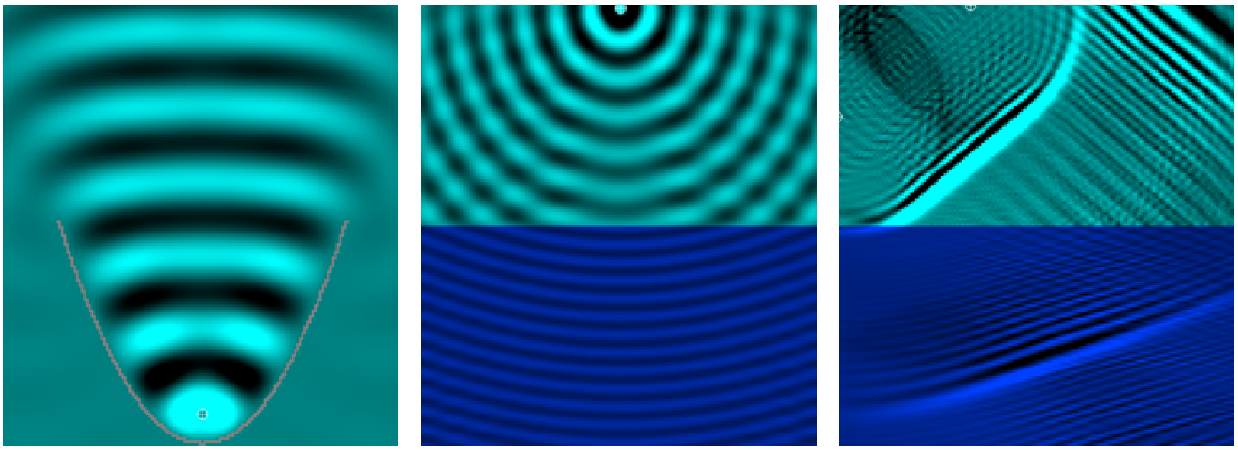
- Two sine waves will produce another sine wave if and only if they have the same frequency. We saw this earlier when we were studying beats and Fourier analysis.
- When two waves combine the interference type depends on the phase difference. They are **in phase** if they both "go up" together. They are **out of phase** when one is "up" and the other "down". Of course, the two waves may be in some intermediate relationship also.
- But if the frequencies are the same, the interference does not change over time. The relative amplitudes of the waves are important too. If they are the same and out of phase, we have **destructive interference** where the vibrations cancel and there is no motion. If they are the same and in phase, we have **constructive interference**.
- The distance between source and observer also adds a phase shift based on wavelength:

$$\Delta\phi = -(2\pi/\lambda)\Delta x$$

- So when the waves from two sources (in phase, with the same amplitude) interfere, the interference is constructive if the path-length difference is a multiple of the wavelength and destructive if there is an extra half-wavelength:

$$\Delta\ell = n\lambda \qquad \text{or} \qquad \Delta\ell = (n + \tfrac{1}{2})\lambda$$

Wave Reflection and Refraction

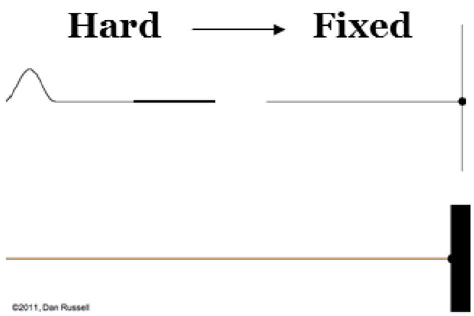


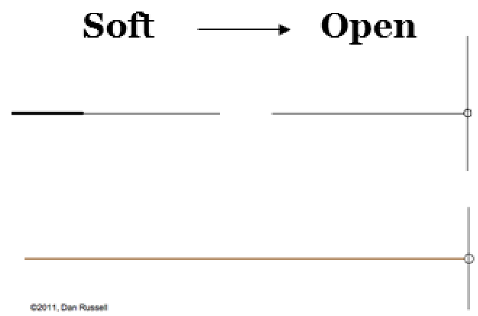
http://www.falstad.com/ripple/

http://www.falstad.com/mathphysics.html

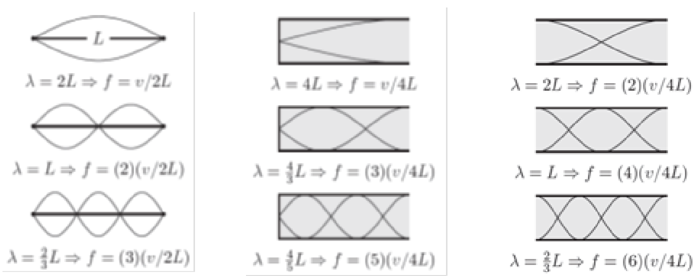
- Once again consider our archetypal string. Consider a situation in which two strings with different mass densities are connected together with the heavier one on the right. A disturbance propagates from the left (called the **incident wave**) and encounters the connection. What happens?
- The tension is the same in both sections, but the masses are different—so the propagation speed is different: it slows down. Also due to the greater mass, the amplitude of this **transmitted wave** is reduced.
- But the wave energy is proportional to the square of the amplitude. Where does the residual energy go? It bounces back as a **reflected wave**. But since the energy is moving in the opposite direction, the amplitude is reversed: the profile is upside-down.
- If we swap the sections of string, we also get reflection but in this case the transmitted wave is larger than the incident wave because of the lower mass density. The profile of the reflection is upright in this case.
- We can push both of these cases to the extreme. In the first case, the connection becomes fixed with infinite mass. There is no transmission but the reflection has the opposite polarity of the incident disturbance. This is called **hard reflection**.
- In the second case, the connection becomes completely open with no mass attached. Again, there is no transmission and the reflection has the same polarity as the incident disturbance. This is **soft reflection**.

1D Standing Waves: Harmonics





http://www.acs.psu.edu/drussell/Demos/



$$f_n = n\frac{v}{2L}$$

$$f_n = (2n-1)\frac{v}{4L}$$

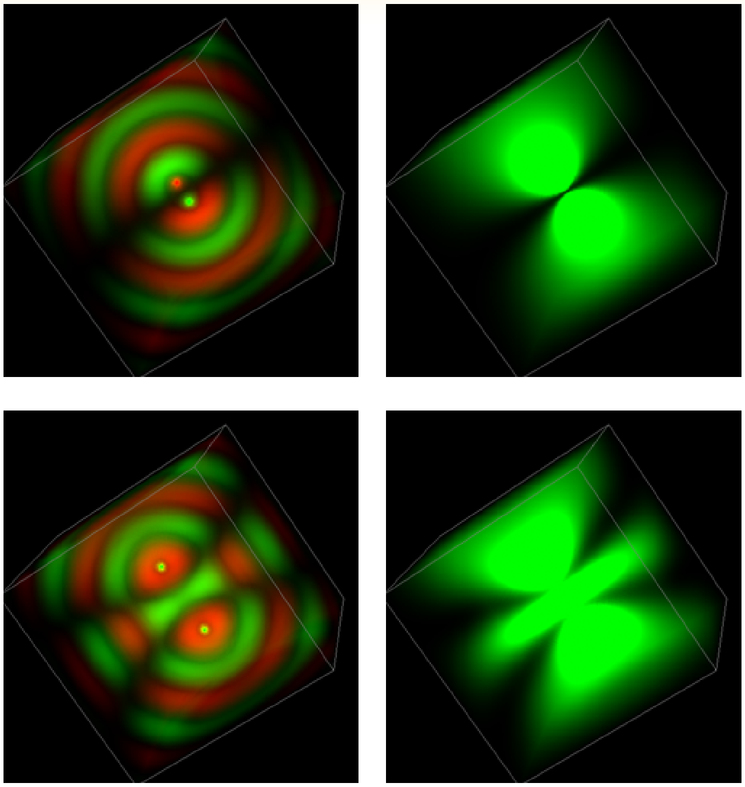
- For an extended wave, the reflection can run back into the incident wave and they interfere. In effect, the boundary acts as a second source. And since the wave energy is coming from both directions equally, a standing wave develops. The amplitude oscillates up and down but the profile does not move at all.
- The points of destructive interference are the **nodes** (no motion) and the points of constructive interference of the standing wave are the **anti-nodes** (maximum motion).
- Now consider a string of length L with one end fixed and at the other end a source of vibration (with frequency f). As the initial disturbance propagates through the string, it will reflect off the fixed end and form a standing wave. But when the reflected wave reaches the other end (with the original source) what happens?
- It will reflect again. But this time there is a difference—this "third" source is right on top of the first. Unless the situation is just right, the reflections will work against the vibration source. But if the secondary reflection is constructive, the standing wave will resonate.
- And this is how every stringed musical instrument works. The required condition is simply that the wave repeat after it has traveled the length of the string and back: $n\lambda = 2L$. In terms of the vibration frequency we have:

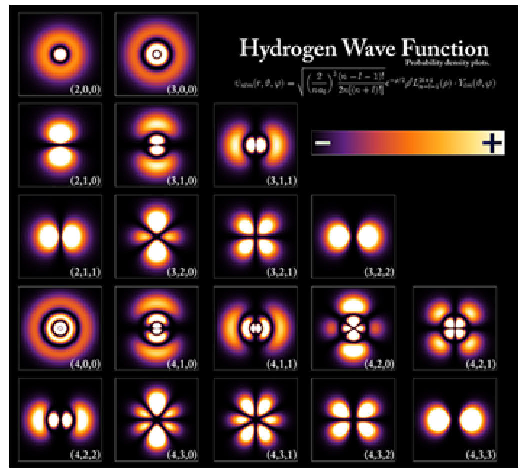
$$f_n = n\frac{v}{2L}$$

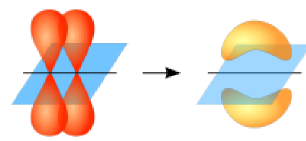
- These are the **harmonics** of the string. The first harmonic, f_1 is called the **fundamental frequency**.
- Another one-dimensional class of musical instrument are the wind instruments. Frequently these instruments are open at one end. This means that the reflected compression wave has the same polarity as the incident wave and we are off by half of a wavelength in order to get the standing waves to resonate. The harmonics of a half-open column of air is

$$f_n = (2n-1)\frac{v}{4L}$$

3D Standing Waves: Orbitals

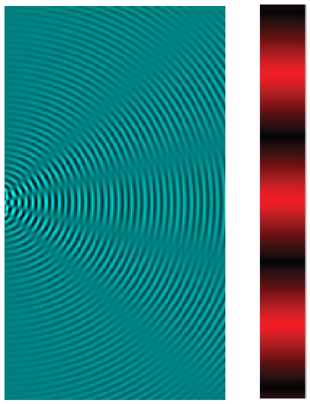


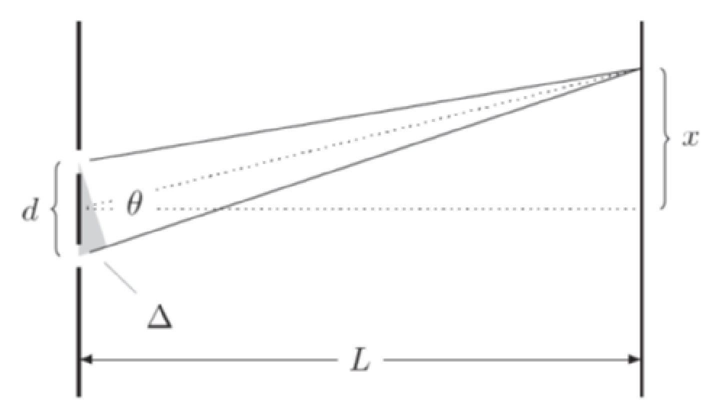




- The third major class of musical instruments is the percussion instruments, which involve banging things together. Although we could consider the triangle as a one-dimensional percussion instrument, usually we think of the drum.
- The vibration of a two-dimensional membrane is complex and very interesting. The geometry of the boundary will determine the way the waves are reflected back into the membrane.
- The main constraint is that the membrane cannot move along its boundary (this is called a **nodal line**). The notion of a one-dimensional harmonic is replaced with the idea of a vibrational **mode**. Each mode has its own frequency and is primarily characterized by its nodal lines.
- Since these membranes are two-dimensional, each mode is characterized by two integers (for example, the number of circular and radial nodal lines).
- We can talk about standing waves in three dimensions (applications to sound production and echoes are relevant here). In this case we need the mathematics of spherical harmonics which are classified by three integers.
- It is even possible to consider the four quantum numbers used in defining the structure of the atomic orbitals as defined by standing matter waves. There are four numbers in this case that correspond to the four dimensions of space-time. But we will leave that discussion for quantum mechanics.

Young's Double Slit



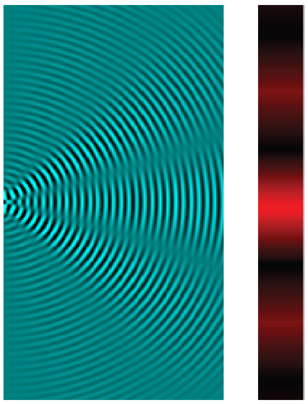


$$\sin\theta = n\lambda/d$$

$$\Delta\ell = d\sin\theta$$

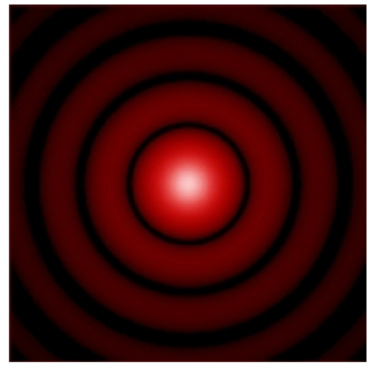
- The **double slit** experiment is simply that: two parallel slits close together. If we send a single coherent wave through these slits, they both act like wave sources that are in phase with one another and interfere. If we place a screen in the distance, we can capture the pattern of constructive and destructive interference.
- The path-length difference from the slits separated by a (small) distance d is
$$\Delta\ell = d\sin\theta$$
- We know we will have constructive interference whenever this distance is an integer number of wavelengths. Thus, we expect to see a bright band on the screen if the angle is given by
$$\sin\theta = n\lambda/d$$
- The main obstacle to overcome when observing the interference of light is the small wavelength. The range for the human eye is about 390 to 750 nanometers for blue and red light respectively. The interference pattern depends on this wavelength, so it is typically very hard to see unless the geometry is of the same order.
- A **diffraction grating** is effectively a double slit analysis on steroids. A diffraction grating is good at displaying the wave nature of light because it separates the colors through interference rather than dispersion. A standard DVD produces this effect because of the closely spaced tracks on the disk.
- For angles where the double-slit produced constructive interference, the diffraction grating does too. However, when the top and bottom slits of the grating are destructive, others will propagate through. The diffraction pattern still peaks at the constructive interference previous points, but the peaks are much shaper (narrower and taller).
- A typical diffraction grating will have many slits. Typically we are told the number of slits per centimeter or some similar metric. The angle of diffraction is proportional to this number.

Single Slit Diffraction



$$\theta = n\lambda/a$$

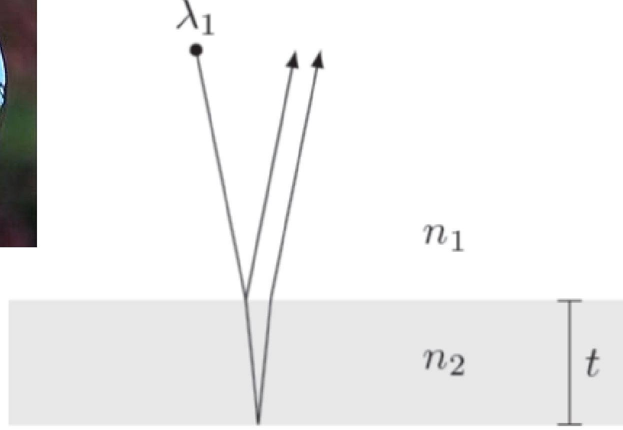
$$I = \left[\left(\frac{\lambda}{\pi a\theta}\right)\sin\left(\frac{\pi a\theta}{\lambda}\right)\right]^2$$

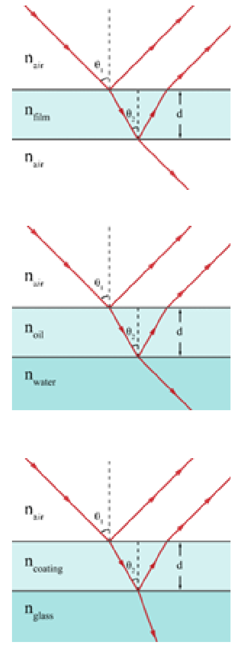


$$\theta_{\mathbf{min}} = 1.22\lambda/D$$

- Young's original double slit experiment depends upon the **diffraction** of light. This is the tendency for waves to bend around obstacles and why you can hear around corners. This is why the light bends and interferes in his double slit experiment rather than simply casting a very sharp shadow of the two slits.
- If we were to remove a slit from Young's arrangement, the remaining slit will cast its shadow on the screen. But since the slit is small, the shadow will be fuzzy. The intensity of the diffraction pattern is given by
$$I = \left[\left(\frac{\lambda}{\pi a\theta}\right)\sin\left(\frac{\pi a\theta}{\lambda}\right)\right]^2$$
where a is the width of the slit. This function assumes that the distance to the screen is much larger than the slit width so that the rays of light are effectively parallel.
- The diffraction given drops to zero whenever
$$\theta = n\lambda/a$$
except when $n = 0$ which is the brightest point of all. So even with a single slit we can see the wave nature of light. But the side lobes are usually very faint, so effectively the image is spread over the range defined by the first fringe.
- We can use this fact as an objective measure of optical resolution. We say that we can distinguish two sources if the diffraction peaks do not overlap. This is called the **Rayleigh criterion**.
- However, the previous formulas assume the diffraction aperture is a thin slit. Since the typical aperture in a telescope is circular (and the pupil in your eye), the actual criterion is different. For a circular aperture of diameter D the angular separation we can distinguish is
$$\theta_{\text{min}} = 1.22\lambda/D$$
- Thus, the human eye can resolve images with an angular separation of 1.4×10^{-5} radians or about 100 arc-seconds.

Thin Film Interference





$$2t = (m + \tfrac{1}{2})(n_1/n_2)\lambda_1$$

- Iridescence in nature is usually not from a diffraction grating. It is much more common for this effect to be produced by **thin-film interference**. Simple examples include soap bubbles and the colorful sheen of oil on water.
- Physically this occurs because both the inner and outer boundary act as reflection points. The light from the front of the layer will constructively interfere with the light from the back of the layer if the distances are just right.
- If the thickness of the film is t, the path-length difference is $2t$, so you might think we require

  $$2t = m\lambda$$

  (we use m to label the integers in order to avoid confusion with the index of refraction). However, the reflection from the front is a hard reflection because the refractive index of the film is larger than air.
- The reflection off the back is a soft reflection for similar reasons, so the interference acts something like the open tube. This means we require

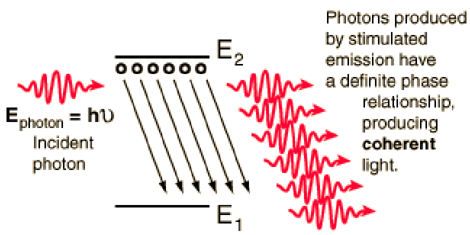
  $$2t = (m + \tfrac{1}{2})\lambda_2$$

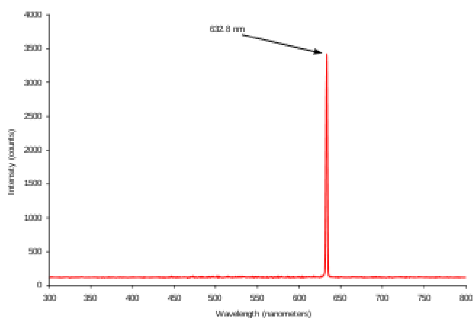
  where λ_2 is the wavelength of the light in the medium.
- But we know that the wavelength of the light changes with any change in the index of refraction because its speed changes. The thin-film interference formula is:

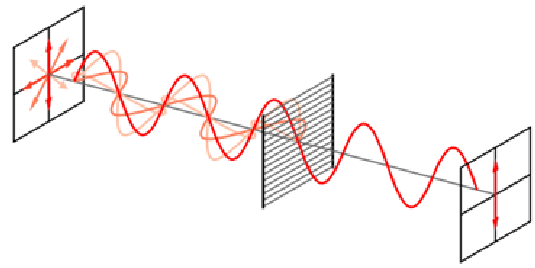
  $$2t = (m + \tfrac{1}{2})(n_1/n_2)\lambda_1$$

  where we give the wavelength the subscript 1 to emphasize that this is the source wavelength.
- You might wonder why the film must be thin. Typically as light moves through a transparent medium, it loses intensity. This attenuation of light causes the intensity to fall exponentially. But the two rays must be of comparable amplitude in order to interfere. If the film must be thick, we get no interference.

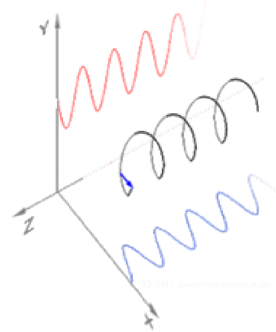
Laser Light and Polaroids







$$I = I_0 \sin^2 \theta$$



- There are two more illustrations of the wave nature of light. The first is **coherence**. This is the difference between incandescent light and laser light.
- In laser light, each pulse of light is aligned by both phase and polarization. This is what we mean by coherence.
  - This occurs because a type of standing wave is built in the lasing material which excites more coherent light via a quantum mechanical process. Although the production is based on quantum mechanics, the coherence of the light is understandable with classical ideas.
- With an incandescent light, the alignment of phase and polarization is random: sometimes its constructive, sometimes its destructive. But laser light is coherent so the light rays constructively interfere all the time.
- Since the intensity of the light is proportional to the square of the amplitude, when the intensity of the incandescent bulb increases two-fold, the intensity of laser light increases four-fold. So, the high intensity of laser light is due to its coherence—a wave phenomena.
  - The second illustration is from the polarization of light. When we pass light through a polarizing filter, the component of the light wave aligned with the polarizer axis passes through.
  - If we place a second polarizer at perpendicular to the original, light is blocked because there is no component left to pass through. But insert a third polarizer between them and light will begin to pass through again. The simplest explain of this is with a transverse wave.
  - When two polarizers are at an angle θ, the intensity of light that passes through is given by **Malus' law**:

$$I = I_0 \sin^2 \theta$$

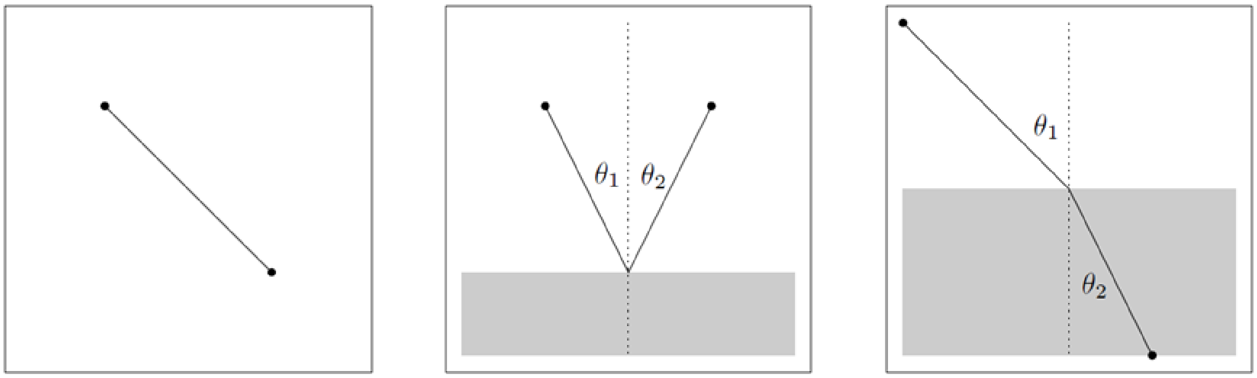
  - This follows because the intensity is proportional to the square of the amplitude component allowed through.

Physics 202 Lecture 10

Quantum Optics and Least Action

- As we alluded to when discussing the ultraviolet catastrophe, the wave nature of light is not the end of the story.
- There are four facets to understanding the nature of light:
  - Light rays
  - Light waves
  - Electromagnetic waves
  - Quantum photons
- Surprisingly, unlocking the mysteries of light involves nearly every branch of mechanics. The final theory that wraps it all up is called **quantum electrodynamics**, or QED.
- Feynman was awarded one-third a Nobel prize for figuring out how to make this theory work. He also wrote a book targeted to non-physics people on the subject called *QED: The strange theory of light and matter*.
- This lecture is inspired by that book. But I have also taken the liberty of manipulating the story line in order to integrate more closely with previous lectures.

Fermat's Principle: Least Time

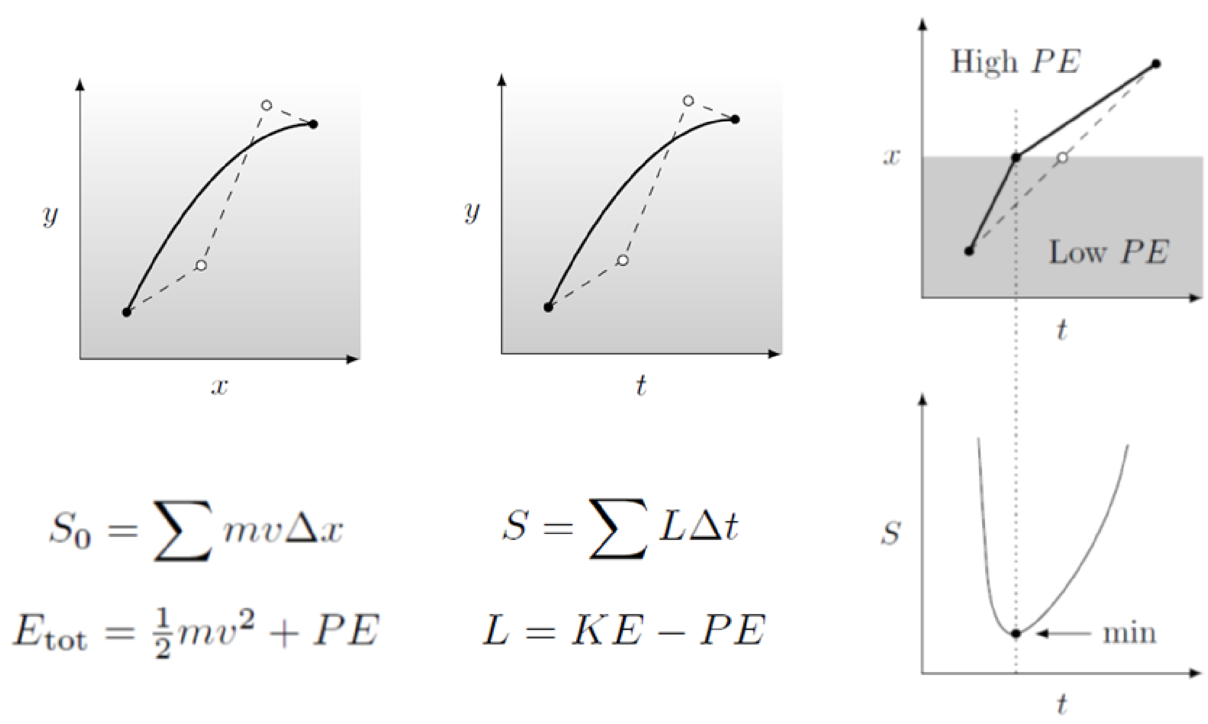


$$v = c/n$$

$$\ell = \sum n \Delta x$$

- Our story starts at the dawn of modern physics back in the 1660s (the first edition of Newton's *Principia* was published in 1687).
- **Fermat's principle** is the statement that the path a ray of light takes is the one that requires the least amount of time to travel. The three laws of geometric optics follow from this principle:
  - Rectilinear propagation: when the speed of light is constant, the least time is provided by the path with the shortest distance. The shortest distance between two points is a straight line.
  - Law of reflection. Again, the shortest distance between two points that bounces off a mirror is the path in which the angle of reflection equals the angle of incidence.
  - Snell's law of refraction. This is a bit more difficult to prove and you must remember the relationship for the speed of light in a transparent material is $v = c/n$.
- Another way to state the principle of least time is through the **optical path length**, which is simply the distance multiplied by the index of refraction:
$$\ell = \sum n\Delta x$$
- This is equivalent to the time multiplied by the speed of light in vacuum, so Fermat's principle also minimizes the optical path length for the light ray.
- Initially there was a lot of resistance to this idea for metaphysical reasons. Nonetheless, the math works.

Hamilton's Principle: Least Action



- In the mid-1700s, Euler discovered something similar to Fermat's principle. The sum of the displacement of a particle multiplied by its momentum for the actual particle trajectory is smaller than any other imagined trajectory.
- This quantity is called the **abbreviated action**:

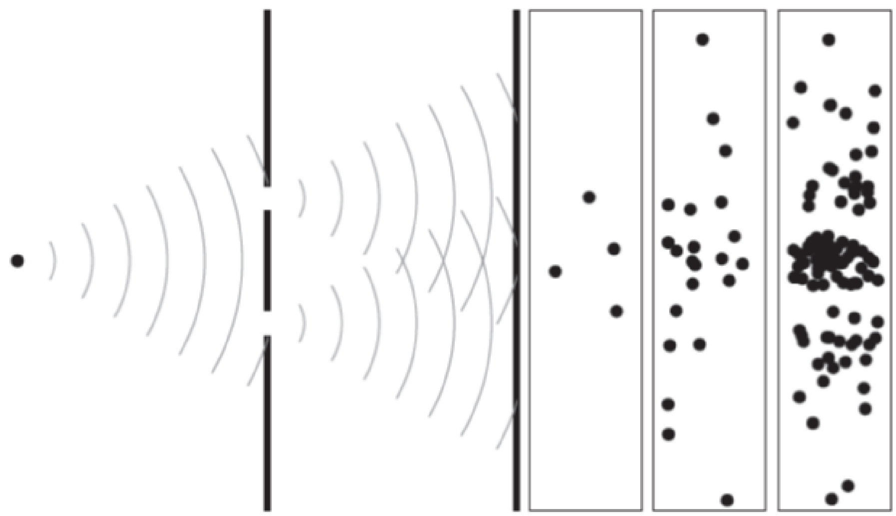
$$S_0 = \sum mv\Delta x$$

  where the speed is most easily determined from the conservation of energy: $E_{\text{tot}} = \frac{1}{2}mv^2 + PE$
- As Edwin Taylor says (see http://www.eftaylor.com/pub/variational1D.html),
  - Newton's laws answer "What happens next?"
  - Euler's abbreviated action answers "How do I get from here to there?"
  - But Hamilton's principle answers the big queston: "From here now how do I get to there then?"
- The **action** (no modifier) is defined as

$$S = \sum L\Delta t \qquad \text{with} \qquad L = KE - PE$$

  where the sum is between each point in space and time for the motion. Hamilton's principle states that the actual motion of the particle follows the path that makes this quantity smallest.
- The principle of least action offers two advantages over the abbreviated action:
  - It gives us the motion in space and time.
  - The conservation of energy follows as a consequence rather than an assumption.
- One application is exactly determining the time and trajectory required to get to the Moon.

Feynman's Photons



$$\psi = \exp(iS/\hbar)$$

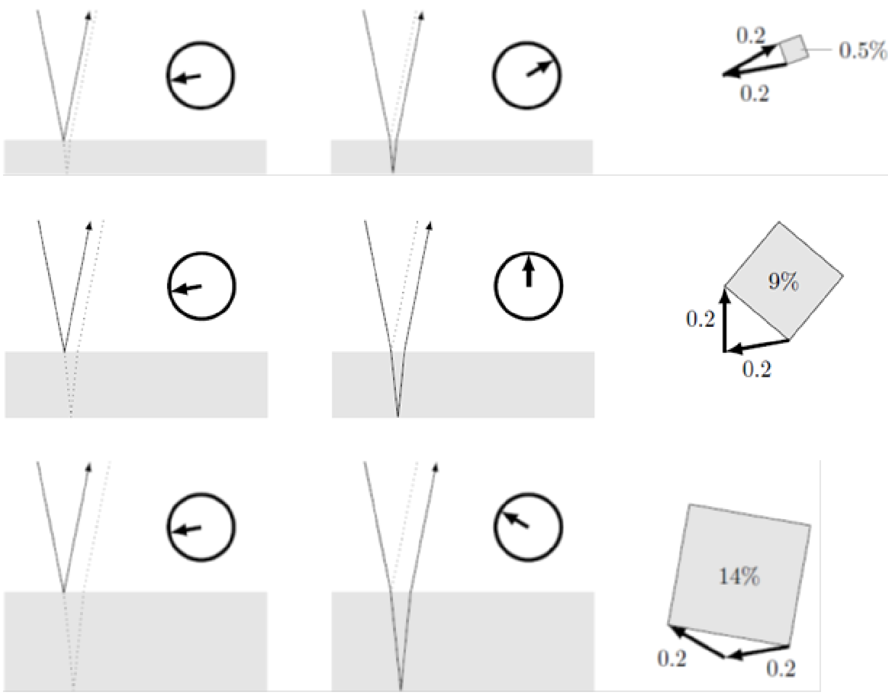
- So optics inspired a new approach to mechanics through Fermat's principle. In an interesting round-about way the principle of least action returns the favor via Feynman's approach to quantum electrodynamics.
- In quantum mechanics there is a concept called "wave-particle duality". The idea being that every quantum object has both wave and particle characteristics. Including light.
- But Feynman is adamant: at its root, light is a particle. Why? As the intensity of light is reduced we literally see the independent impact of each photon. In fact, we saw evidence of this particle-like nature earlier when discussing the ultraviolet catastrophe. Einstein called these light-particles **photons**.
- What then of all the evidence for the wave nature of light? Here is where the principle of least action enters. Feynman says that the photon actually "explores" every imaginable path from point A to point B. The probability that it gets there is proportional to the sum over all these possible histories.
- The contribution from each possible path is proportional to

  $$\psi = \exp(iS/\hbar)$$

  where S is the action for the path under consideration. Note that this is a complex number. I am going to call this the **contribution amplitude** of the path or process.
- In this way, Feynman gets to have his cake and eat it too. Because the photon is affected by all the possible paths, a double slit (for example) will impact the overall probability of the motion.
- When the frequency is constant, the action for a photon is simply $S = hf\Delta t$. So the contribution amplitude depends only on time:

  $$\psi = \exp(2\pi i f \Delta t)$$

The Grand Principle: Spinning Arrows



$$S = hf\Delta t$$

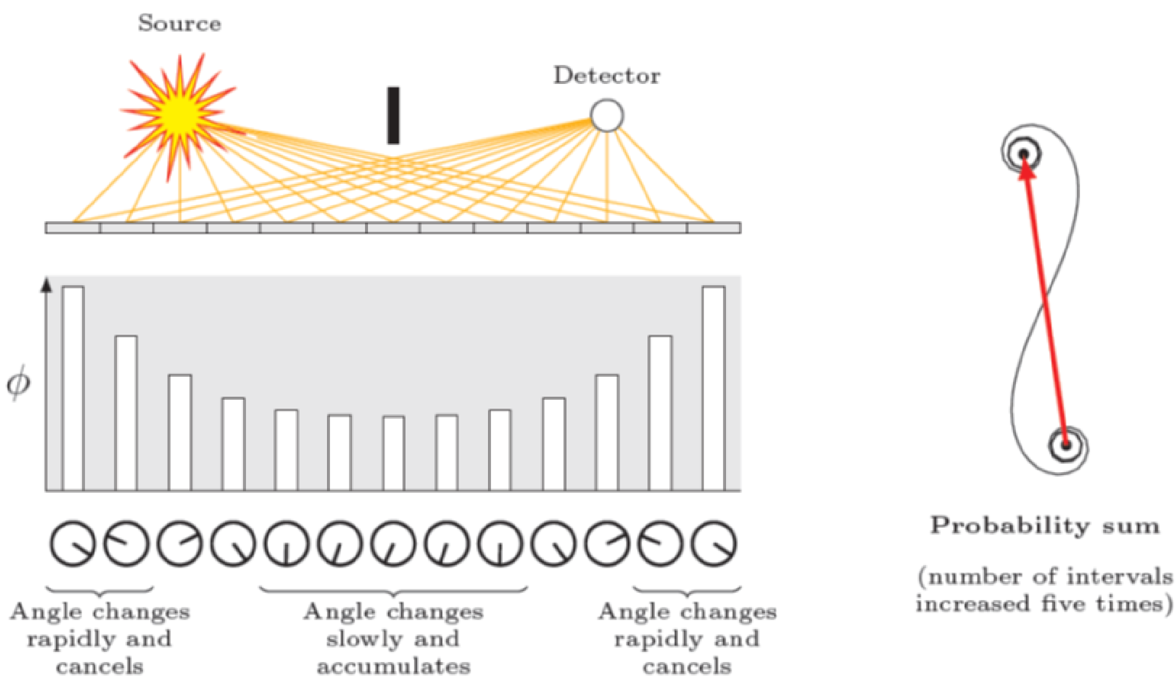
$$\phi = S/h = f\Delta t$$

- In geometric optics a ray of light that encounters a transparent block will be partially reflected and partially transmitted. For normal glass about 4% of the light is reflected.
- Feynman emphasizes how this forces us to consider probability in our understanding of nature. Send 100 photons toward a glass slab and four bounce back. Why, if they are all the same? Most physicists argue there is no answer.
- Setting aside these metaphysical concerns, the probability or intensity of a process is proportional to the contribution amplitude squared. Feynman calls this the "grand principle" in his book. In his words:

  > The probability of an event is equal to the square of the length of an arrow called the "probability amplitude." An arrow of length 0.40, for example, represents a probability of 0.16, or 16%.

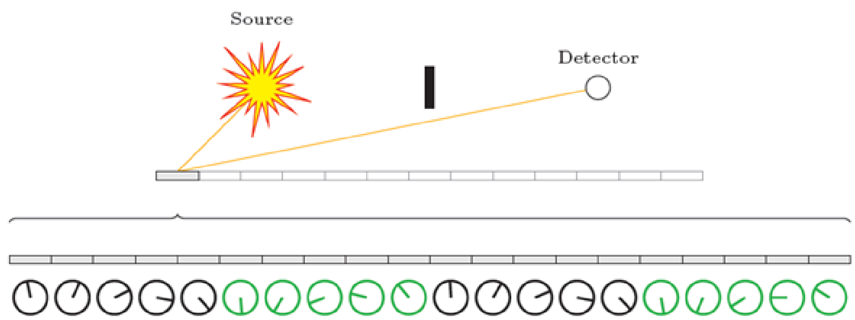
- The reason he calls the amplitude an arrow is that it is a complex number, and a complex number can be thought of as a two-dimensional vector (with real and imaginary components).
  - The angle of this arrow is S/h in general and $f\Delta t$ for our photons.
  - The magnitude of this arrow squared yields the probability or intensity for the path or process.
- For reflected light the length of the arrow is 0.20 to yield the 4% intensity we see. The transmission arrow must have a length of 0.98 to yield the remaining 96% of the intensity.
- When we have more than one path, we add these spinning arrows like vectors. Example: thin film interference.
  - There are two rays of light each with an amplitude of 0.20 units. The direct reflection is hard and flips the direction of the arrow. The internal reflection is soft and leaves the direction intact.
  - The direction of these arrows is determined by the path lengths through $f\Delta t$. The vector sum of these two arrows will vary in size between 0 and 0.40 units depending on the orientation of these two arrows. We have described an interference pattern without using waves!

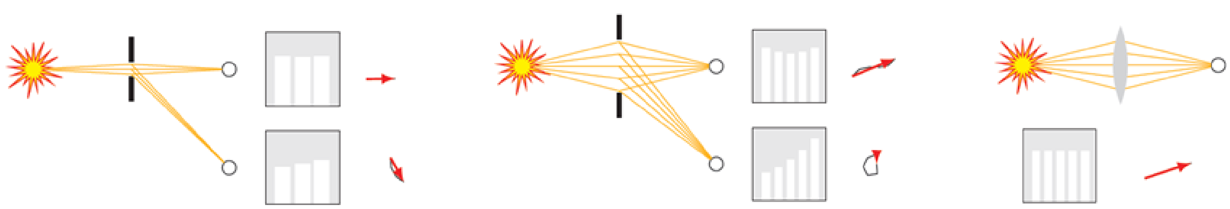
The General Rule: Explore All Paths



- Feynman calls this amplitude sum the "general rule" of QED. We have spoken of adding two arrows. The reality is that one must add *all* the possible arrows to get the final probability or intensity. Mathematically this is called a **path integral**, but it is also more picturesquely a called **sum over histories**.
- Even moving in a straight line from point A to point B involves summing an infinite number of possibilities. A photon may go in a straight line, but it may also take a detour to Jupiter. As long as it can get back, it contributes.
- But something incredible happens. Any path with an action far from the minimum gets canceled out by neighboring paths. The action only need differ by half of Planck's constant $h/2$ and the arrow points the other way.
- What really happens is that only paths with an action close to the minimum will contribute constructively to the total. Look what we get from these little spinning arrows:
  - An explanation for the principle of least action—and therefore all of classical mechanics.
  - An explanation for Fermat's principle—and therefore all of geometric optics.
  - An explanation for the interference effects in optics, i.e. all of physical optics without waves.
  - Quantum mechanics and relativity are built into the theory from the beginning.
  - Electromagnetism (and the two nuclear forces) can be explained in this framework also.
- Of course, the theory is not all dancing unicorns and sparkling rainbows.
  - First off, you have to buy the magic beans. Probability amplitudes based on action with a sum over histories.
  - Second, there are giants at the top of the beanstalk. Obviously the math is hard: some uber-calculus stuff is required here. But even if you can do that, the theory encounters issues similar to the ultraviolet catastrophe (an infinite electron mass, for example). In the 1940's QED was DOA. Feynman, Schwinger and Tomonaga shared a Nobel Prize for slaying these giants.

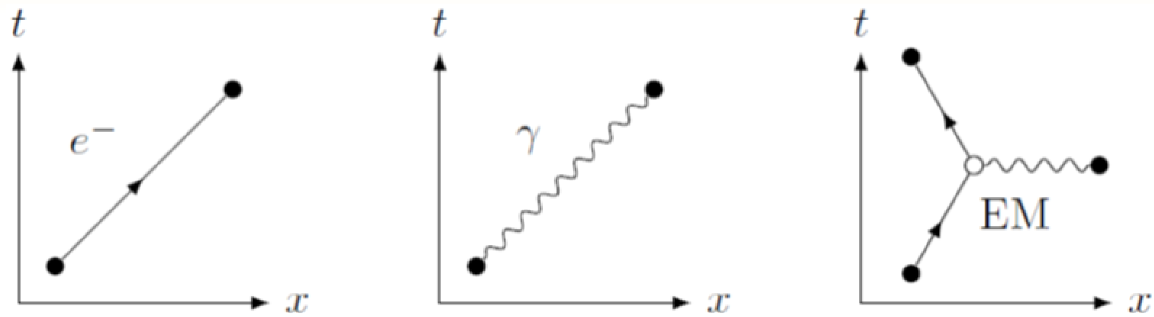
All Optics Explained (and QM)

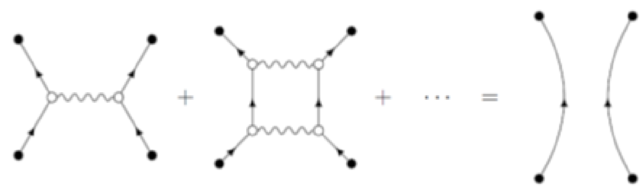


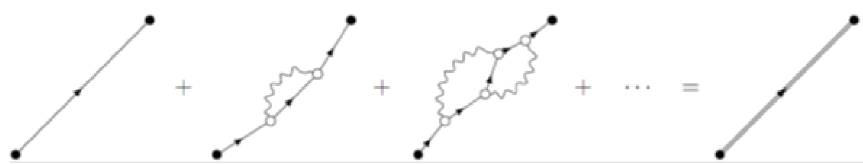


- Do we really need every possible path if they all cancel out in the end?
  - Consider a light path that "reflects" far from where it should (i.e., $\theta_r \neq \theta_i$). The action slopes with the distance from the classical reflection point. This change in action spins the arrows causing destructive interference.
  - However, we can scrape off parts of the mirror and pick which paths to "reflect" and which to absorb. If we absorb all the destructive paths, the remainder will constructively contribute to the sum.
  - These scrapes must be based on the rate the arrows spin, i.e. the frequency of the photon. We have just described a standard diffraction grating.
- We can also explain diffraction effects.
  - Consider light traveling through a slit and an arbitrary point on the screen behind the slit. All the possible paths from within the slit to the point on the screen will vary in distance. This variation in distance will manifest as varying angle for the contribution amplitudes from these paths.
  - If the slit is very small we will block the destructive influence of the paths not coming from the center of the slit. By very small we need small time we need $f\Delta t \sim 2\pi$ which implies $d \sim 2\pi\lambda$. So if the slit is on the order of the wavelength of the light, we will get some intensity off center—which is diffraction.
  - In QED, the same logic applies to electrons and is the source for the "Heisenberg uncertainty principle".
- We can even explain how lenses work.
  - As was said earlier, the paths around the straight line for a photon cancel. However, if we can speed up the photon along the non-central paths, we should be able to get them to contribute at the observation point.
  - We cannot speed up the light, but we can slow it down. If we force the central path through glass the speed of light will drop according to $v = c/n$. The shape of a lens is such that all our spinning arrows end up aligned and create a much more intense light than that without the lens.

QED: Electrons Exchange Photons

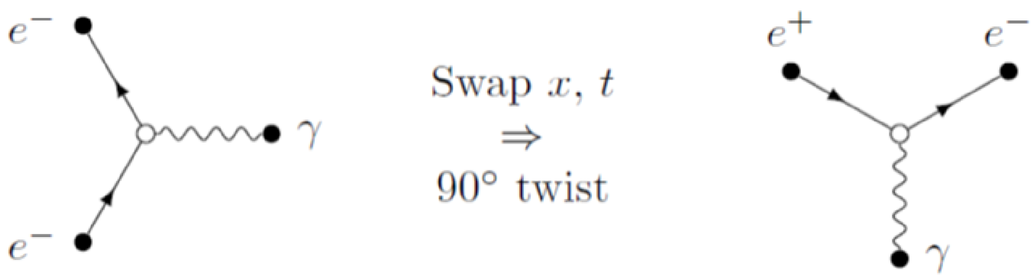


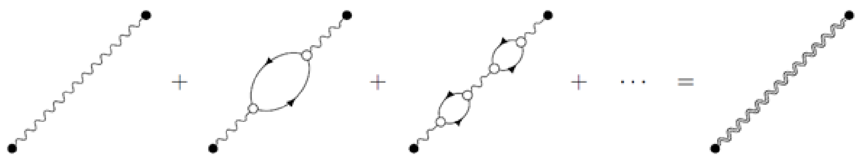


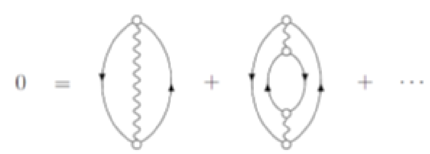


- We are now ready to take the next step: the electromagnetic interaction. This is done through **Feynman diagrams**. Each diagram is a combination of three fundamental components:
  - The motion of an electron symbolized by a straight line.
  - The motion of a photon symbolized by a squiggly line.
  - The interaction: the electron emits/absorbs a photon.
- The electromagnetic interaction is represented by the unobserved exchange of a photon between electrons. This exchange cannot be observed because to do so would destroy the very photon being exchanged—it's all or nothing.
- Although the virtual photon cannot be seen, it does affect the state of the electrons. But because we do not know how the photon is exchanged, we must consider all possible histories.
- A Feynman diagram is so suggestive it is important to emphasize what it is not.
  - It does not represent the actual interaction—it is a book-keeping tool used to calculate the interaction.
  - The diagram represents a whole class of similar diagrams with the unobserved virtual events located in different places. Anything goes as long as the initial and final states are the same.
- So the interaction involves contributions in which either the first or the second electron emits a photon. The sum total of all the virtual interactions is symmetric even if the individual ones are not.
- Why not more than one photon? Well, you've got to include all of those cases also. When you do, you recover Maxwell's laws of classical electromagnetism.
- One interesting example is an electron interacting with itself. In QED, the straight line path of an electron is covered by a swarm of virtual photons being constantly emitted and reabsorbed.

QED: Matter and Anti-Matter







- With Feynman diagrams, truly anything goes. Including moving backward in time.
- As one plays around with these diagrams, one gets used to imagining them in a malleable way because each of these internal changes are indistinguishable.
- QED is also relativistically valid and in relativity time and space are joined in space-time. Understanding this, Feynman asked the question: what happens if we rotate these diagrams 90°? Essentially this swaps the role of space and time. The two diagrams are no longer equivalent, but it does beg the question what is it?
- In particular, our basic interaction diagram: an electron absorbs a photon. On its side, this diagram starts with just a photon and ends with two electrons, one moving up in time (normal) and another moving backward in time.
- Feynman recognized this process as the creation of anti-matter. Energy goes in (as a photon) and is converted into an electron (the normal one) and its anti-matter twin the positron (an electron with positive charge).
- In this way the theory provides a simple explanation for the existence of anti-matter and why every particle must have an anti-particle twin.
- Whether or not we interpret anti-matter as time-reversed matter, the fact is that a photon can dissolve into matter/anti-matter pairs and vice versa. This means that we must consider so-called **bubble diagrams** as possible contributions to our overall sum over history.
- And it is precisely these bubble diagrams that explain the slight deviations from classical field theory such as the gyromagnetic moment of the electron to astonishing accuracy.
- We can even draw a self-contained bubble diagram without any incoming or outgoing particles. The vacuum is a sea full of virtual matter-antimatter pairs. In QED even the vacuum is a many-body problem!

Index