Chapter 1: **Introductory / background material**

- some history / background
- celestial mechanics; Lagrange points, tides
- relativity
- forces and elementary particles
- light & spectra; atomic spectra, molecular spectra, excitation energies
- nucleosynthesis
- sample problems

Introduction

This text is intended to supplement a 3-semester undergraduate introduction to astronomy for students with a decent high school preparation in math and physics. In other words, I presume that your algebra skills are solid, that you understand how to handle units, and that you are not unduly frightened by things like logs and exponents. A bit of calculus will be helpful in some places, but is not necessary. I expect that you've seen Newton's Law of Gravitation, that you know that light can act as either a wave or a particle, and that you won't be shocked to discover that the Universe is expanding. But we'll go over all of that.

The first section covers most of the basic physics and geometry that we'll need to make sense of planets, stars, and galaxies. The following three sections are planetary, solar / stellar, and galactic / extragalactic / cosmology. You can tackle these three in any order. In the process of learning the basic astronomy you should also learn a bit about how we've come to know what we know—underlying assumptions, how astronomers ask questions, take data, develop and evaluate models to explain the data, and present those models for scrutiny by others.

Some history / background

Students coming out of high school today have never known a world without moon rocks, galaxies, gravity, telescopes; as I write in 2015, we have rovers on Mars, spacecraft encountering interstellar space, and telescopes, ground-based and / or in space, capable of observing celestial objects with good angular resolution across the entire electromagnetic spectrum. We are debating whether Pluto should really be called a planet and how old the universe was when the first stars were formed. But think back. Spacecraft didn't exist in 1950. Galaxies didn't "exist" in 1900—we weren't sure yet what those fuzzy things were. By 1800 we knew of the existence of Uranus but not Neptune or Pluto. Newton, whose explanation of gravity permits us to send spacecraft to intersect planets in their orbits, was born in 1642 (or 1643, depending on which calendar you use), the year Galileo died. In 1600 no human being had ever looked through a telescope, ever seen the moons of Jupiter or the phases of Venus; no one had imagined the galaxies in the Virgo cluster or the cosmic microwave background (CMB). We didn't know radioactivity or calculus or thermometers. In 1600 Giordano Bruno was burned at the stake, in part for his adherence to the theory that the Earth revolved around the Sun, rather than the other way around. We can, today, with ease view high-resolution images of the moons of Jupiter taken by the Galileo spacecraft, Hubble Space Telescope (HST) images of galaxies in Virgo and beyond, or the latest observations of the CMB.

Pause here to note a few currently excellent web sites for doing precisely that:

- http://apod.nasa.gov/apod/ Astronomy Picture of the Day
- http://www.nineplanets.org/ Nine Planets
- http://outreachoffice.stsci.edu/ Space Tel. Science Institute (Hubble)
- http://map.gsfc.nasa.gov/ the Wilkinson Microwave Anisotropy Probe
- http://www.spaceweather.com/ Spaceweather

On the other hand, in a world lit 24/7 many of you have also not looked with your own eyes at the moons of Jupiter or at Virgo and with cable or satellite you tend not to notice the static as your TV sees some of the CMB. And on the other other hand, we learned a lot pre-1600, in the world of eyeballs and often crude measuring devices. Astronomy has a history. Let's take a look at a few pieces of it.

Think first about what you *can* see with your eyes: the Sun rises and sets each day; the Moon goes through its phases; the planets don't stay in the same place among the stars, and sometimes seem to go "backwards" against the background stars; eclipses happen. And the questions you could ask: What does the Sun do over the course of the year? Why? What about the planets? *where* among the stars are they found? How often do eclipses happen? and *why* do they happen? are they all the same? What do the stars do over the course of one night, one year, one thousand years? What happens to any of your observations if you go to some other place on Earth? What would the Earth look like if you could imagine seeing it from the Moon? And yes, your mind's eye is also a useful "tool", although a few other tools would clearly be handy! What do you need? A calendar. Some means of indicating directions, e.g., landmarks along the horizon and something to give you elevation above the horizon. A map of the sky would be nice. Some means of indicating how bright a star or planet is relative to others. A system to keep track of where things were last year. And a conceptual framework, a cosmology, to put it all in.

The tools and the questions are interrelated. Think, for instance, about the calendar. It may have religious significance; it definitely has economic significance. Is it time for the Nile to flood yet? time to plant or hunt because you can expect the weather or the migration or whatever because you know the day? Landmarks along the horizon are a handy way to mark the passing of the year by the rising or setting location of the Sun. If you don't have any handy landmarks, build a structure; in other words, measure the direction toward the rising or setting Sun (or the Moon, or Venus) and record it for future reference. The Native American Medicine Wheels, found in many places throughout the Great Plains, served as astronomical / calendrical observatories. So did the Henge monuments of the British Isles.

Figure 1.0: Big Horn Medicine Wheel, Medicine Mountain National Historic Landmark

https://www.fs.usda.gov/detail/bighorn/home/?cid=fseprd521531

As the observations accumulate, the conceptual framework has to morph to accommodate the data. For instance, is the Earth round or flat? (or disc-shaped, for the Terry Pratchett fans.) Several ancient Greek astronomers could appeal to two pieces of observational evidence to demonstrate that the Earth is round: first, at the beginning or end of a lunar eclipse, as the Moon is entering or leaving the Earth's shadow, the shadow cast by the Earth on the Moon is always curved; second, travellers returning from different latitudes on Earth reported being able to see different stars, and familiar stars at different elevations, than they saw from Greece. Only a round object could always cast a curved shadow; only a tipped horizon could result in a slightly different set of stars being visible at different latitudes. Take a look at that Earth shadow again: the radius of curvature of the shadow is greater than the radius of curvature of the limb (edge) of the Moon. The Earth must be larger than the Moon. How much larger? That requires knowing at least roughly how far away the Moon is.

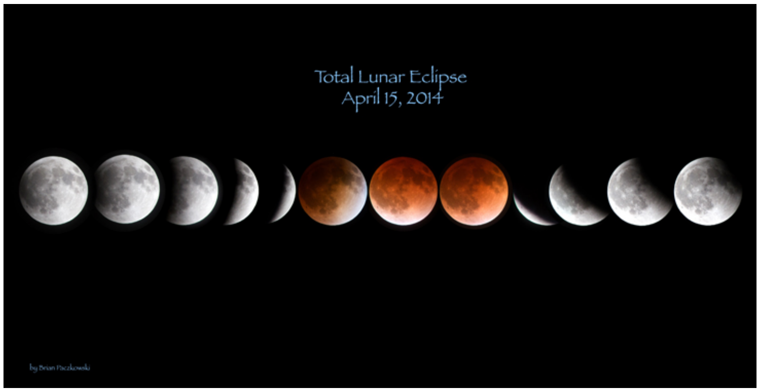


Figure 1.1: Total Lunar Eclipse

Credit: Brian Paczkowski

https://solarsystem.nasa.gov/multimedia/display.cfm?Category=Planets&IM_ID=19209

Pause here briefly to review units for angular measurements: There are 360° (degrees) in a full circle; you may also have used *radians* and recall that there are 2π rad in a full circle. The degree is subdivided into minutes and seconds of arc (′ and ″, respectively). There are 60″ / 1′ and 60′ / 1°.

Aristarchus of Samos (c. 310 – 230 BCE) was one of the first to beat his head against that question of the relative size and distance of the Moon and Sun. He reasoned as follows: at 1st and 3rd quarter, the Earth – Moon – Sun angle is 90°. If the Sun were infinitely far away, the Moon – Earth – Sun angle would also be 90°. But it isn't, so it isn't. If the Moon's orbit around the Earth were a circle, then there should be a difference in the lengths of the times between New Moon and 1st Quarter and between 1st Quarter and Full Moon. If the difference were, say, 5 days vs. 10 days, then the Moon – Earth – Sun angle would be 5/15 of 180° or 60°. Knowing the angles, you can at least estimate the relative distances to the Moon and Sun. The Moon's orbit isn't a circle, so its velocity isn't constant, and it's hard to measure the exact instant of the New or Quarter Moon, and the angle is about 89°50′, i.e., not all that different from 90° when you're trying to measure by eye. Still, Aristarchus did convince himself that the Sun was about 18-20 times farther away than the Moon. That's about a factor of 400 too small, which isn't too wildly off given his tools. More to the point, the resulting model was derived by carefully thinking about the problem and making observations. The following sketch shows what the geometry would look like if the Moon – Earth – Sun angle were 80°; you can see that the number of days from First Quarter to Third Quarter Moon would be distinctly different than the number of days from Third to First.

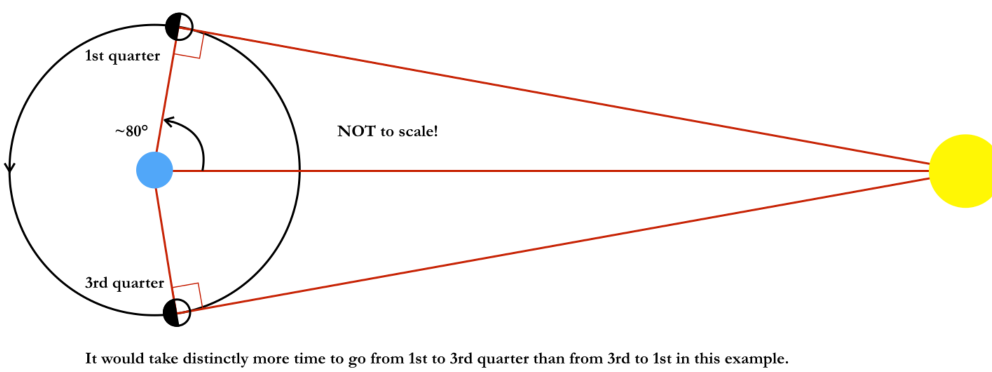


Figure 1.2: Quarter Moon geometry and the relative distance to the Sun

Now to get back to the relative sizes: solar eclipses show that the Moon and the Sun subtend about the same angular size on the sky (~ ½°), so if the Sun were 20 times farther away, it must be 20 times larger than the Moon. You can get the relative size of the Earth by measuring how long it takes for the Moon to move through the Earth's shadow during a lunar eclipse. That's a lot of geometry. Stop and see if you can sketch it.

Relative sizes are nice but we'd like to know how large the Earth is. Eratosthenes (c. 276 – 195/6 BCE) used the assumption that the Sun is *almost* infinitely far away; it's so far away that the rays of sunlight are basically parallel to each other when they get to the Earth. Eratosthenes worked in Alexandria, where the Sun never goes directly overhead. He had heard that on the day of the summer solstice the sunlight fell directly down a well at noon, i.e., it went directly overhead, in the city of Syene, to the south of Alexandria (near modern Aswan). Using a gnomon to cast a shadow Eratosthenes measured the angle of the Sun from the zenith at noon on the day of the summer solstice in Alexandria and found that it was 7°. (You can measure the Sun's elevation by using a vertical stick of known length, measuring the length of the shadow it casts, and doing a little trigonometry.) Eratosthenes hired someone to pace off the distance from Alexandria to Syene. We know that he got a result of 5000 stadia, but there were a couple of different sizes for the Greek stadium, meaning that we aren't exactly sure how to translate that distance into modern units. In stadia, the circumference of the Earth would be 5000 x 360/7; depending on which stadium Eratosthenes was using (your run-of-the-mill stadium vs. the Olympic-sized version), his result is off by something between 1 and 20%. Once again, stop and consider a sketch and make sure the geometry makes sense.

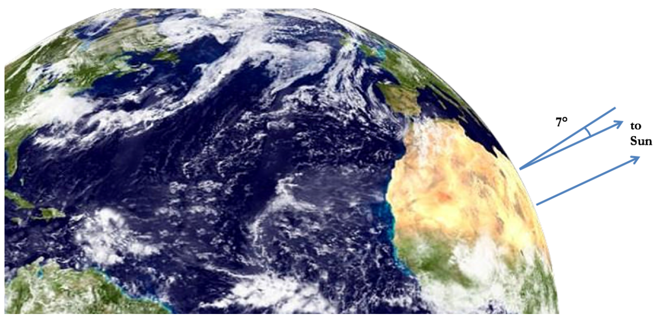
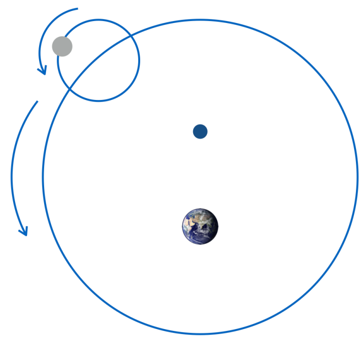


Figure 1.3: Eratosthenes' geometry for measuring the Earth

Earth image: http://oceanservice.noaa.gov/facts/atlantic.html

If the model you are constructing works, then you can *predict* what the Sun, Moon, etc., will do in the next cycle. If they don't behave as expected, you have a problem. Once you have a model that almost works, though, when the data don't fit it's human nature to tweak the model rather than trash it. Astronomy in the Greek tradition got hung up on two points, in particular. One was observational: Aristotle reasoned that if the Earth went around the Sun then we should see the stars shift back and forth over the year (this is called *parallax*; hold your finger up in front of your face and look at it with first one eye and then the other and observe your finger shift back and forth relative to more distant objects); we don't see this, therefore we must be stationary. The only problem with Aristotle's logic is that he could not conceive of the incredible distances to the stars (and, therefore, the incredibly small parallaxes). The second point was aesthetic: circles are more symmetric than other geometric figures and therefore more aesthetically pleasing and therefore motions in the heavens should be circular.

The most famous astronomer of the Hellenistic world was Claudius Ptolemy (ca. 140 CE), also of Alexandria. His compilation, of his work and others', was translated into Arabic and is known to us today as the *Almagest*. The elaborate geocentric model of the cosmos that flourished in Europe until the 17th century is called the Ptolemaic model in his honor. The basic structure of this model is that the Sun and planets orbit the Earth, riding on small circles called epicycles which themselves orbit the Earth on a larger circle called a deferent. It was apparent by Ptolemy's time that the individual planets do not move around the sky at a constant rate nor do they each reach the same maximum brightness each time around the sky (each *synodic* period). Ptolemy's model accounted for this by having the Earth offset from the center of the deferent and having the epicycle move at a constant rate not around the Earth or the center of the deferent but around an equant point on the opposite side of the center of the deferent from the Earth.



The basic elements of the geocentric cosmos: Earth offset from the center, planets on epicycles orbiting the Earth on a deferent.

Figure 1.4

Models can be mechanical as well as geometrical. (Ptolemy's is similar to the way the gearing in a mechanical planetarium projector works today.) About the same time as Ptolemy, the astronomer Zhang Heng of the Eastern Han dynasty designed an elaborate, geared, water-powered armillary sphere to mimic the motions to the planets among the stars (Zhang is also credited with inventing the world's first seismometer, improving calculations for the value of pi, and writing poetry).

Again, for a model to be useable, it should be able to predict where among the stars the planets will be at any given time. That's not something that may be obvious from one year to the next. It's a long-term project. It would be extremely difficult, for instance, to discern patterns in eclipses (location, duration, etc.) without a couple of hundred years' worth of eclipse records. Fortunately humans tend to keep records. As early as 1600 BCE the Babylonians had star catalogs. By the 8th century BCE both the Chinese and the Babylonians were keeping records of solar eclipses. Chinese astronomers recorded a comet in 613 BCE that may have been the first record of an apparition of Halley's Comet. In 44 CE the Chinese and the Romans recorded a blood-red comet – Mt. Etna was erupting at the time and putting large amounts of dust into the air world-wide. Comets as portents? 44 *was* the year Julius Caesar was assassinated.

Eventually Hellenistic natural philosophy ground to a halt. The death of Hypatia of Alexandria, in 415 CE, is emblematic. Hypatia was a mathematician, philosopher, teacher, inventor, writer; she's credited with inventing an astrolabe (an inclinometer appropriately marked to point to prominent astronomical objects) and with a device for distilling water. Her writings were destroyed along with the library of Alexandria, but we know of some of what she wrote. She's credited, for instance, with writing that you should "reserve your right to think, because even to think wrongly is better than not to think at all." She got sideways with Cyril, the local bishop, or got between Cyril and the civil authorities, although it's unclear whether it was at Cyril's urging or not that a mob pulled Hypatia from her classroom and skinned her to death with oyster shells.

Ptolemy's mathematical model was not always in accord with its underlying Aristotelian cosmological principles, for instance in the matter of uniform circular motion. Islamic astronomers of the Middle Ages were highly critical of this intellectual disconnect and worked to improve upon the Greek models. One of the most elaborate astronomical observatories of its time was founded by Nasīr al-Dīn al-Tūsī in Marāgha, in what is now the East Azerbaijan province of Iran, in 1259. It included an extensive library and high-quality instruments built by Tūsī's colleague Mu'ayyad al-Dīn al-'Urdī. Tūsī and 'Urdī devised sophisticated mathematical theorems, which today bear their names, in their attempts to construct an accurate, intellectually coherent model of the celestial sphere. Historian George Saliba, among others, argues that this work was available to Copernicus three centuries later as he, also, tackled the challenge of creating a viable model of the cosmos. There is an excellent article about medieval Arab astronomy by Saliba in the July/Aug, 2002, issue of *American Scientist* (p. 360).

The Copernican Revolution is, in some sense, a bit of a pun: In Latin, the title of his treatise is called *De revolutionibus orbium coelestium* (On the revolutions of celestial bodies). It was published in

1543, while Copernicus was literally on his deathbed. He'd worked out his heliocentric model over several decades but was reluctant to publish it given the intellectual climate in Europe (and yes, Copernicus' work spent about 200 years on the *Index* of books prohibited by the Catholic Church). His model *is* revolutionary in that the Earth orbits the Sun. It still involved circles and epicycles and was not wildly accurate; on the other hand, placing the Earth itself among the heavens is a monumental shift in the conception of our place in the cosmos. Think, for instance, of the difference between Copernicus' view of the Earth as one planet among many and Dante's view of Earth in the *Inferno*. It makes a difference whether Earth is imagined as the location of hell or in the heavens.

We are going to be discussing objects that are large (and larger) distances away. Let's look at some of the relevant units. First, a note about metric distances: the standard unit of measurement is the meter. There are some prefixes with which you should be familiar (and a few more that are fun simply because they are so extreme). For example, the centimeter (cm) is 10^{-2} m. Most of the interesting units come in multiples of 10^3. Here's a table of units:

millimeter (mm)	10^{-3} m	kilometer (km)	10^3 m
micrometer or micron (μm)	10^{-6} m	megameter (Mm)	10^6 m
nanometer (nm)	10^{-9} m	gigameter (Gm)	10^9 m
picometer (pm)	10^{-12} m	terameter (Tm)	10^{12} m
femtometer (fm)	10^{-15} m	petameter (Pm)	10^{15} m
attometer (am)	10^{-18} m	exameter (Em)	10^{18} m
zeptometer (zm)	10^{-21} m	zettameter (Zm)	10^{21} m
yoctometer (ym)	10^{-24} m	yottameter (Ym)	10^{24} m

The average distance between the Earth and Sun, known as the astronomical unit (AU) is ~150 million km ($1.495 \cdot 10^{11}$ m). Onward to what we observe in a heliocentric cosmos.

We are also going to be using the terms *mass* and *density* frequently. Mass is the amount of something, usually expressed in units of kilograms or solar masses; density describes how tightly packed that mass is, usually expressed in units of kg/m^3 (or g/cm^3). Temperatures are going to be in kelvins, which are like degrees Celsius, but starting from absolute zero; −273 °C = 0 K.

Celestial Mechanics – gravity, orbits, that sort of thing

The segue from history to physics is Tycho Brahe (1546 – 1601). Tycho was the last great astronomer of the pre-telescopic era. Tycho was born into a noble Danish family; he was raised by his uncle, who died of pneumonia after pulling King Frederick II out of the ocean, saving the good king from drowning. Tycho was interested in astronomy, particularly after the supernova of 1572, and made meticulous observations; Frederick was impressed, and financed the building of two observatories for Tycho on the island of Hven. Tycho didn't get on with the next king of Denmark, and in 1599 he moved to Prague and with the patronage of the Holy Roman Emperor Rudolf II, built a new observatory. For the last two years of his life, in Prague, Tycho was assisted by Johannes Kepler (1571 – 1630). Kepler was a committed Copernican and a pious mystic. Tycho was conservative in his cosmology—he was definitely not a Copernican—and a bit rowdy in his personal life, to put it mildly. While in Denmark, Tycho is reputed to have had a pet elk who died after getting drunk and falling down the castle stairs. Tycho was also an extremely meticulous observer, and the extensive and precise records of planetary positions that

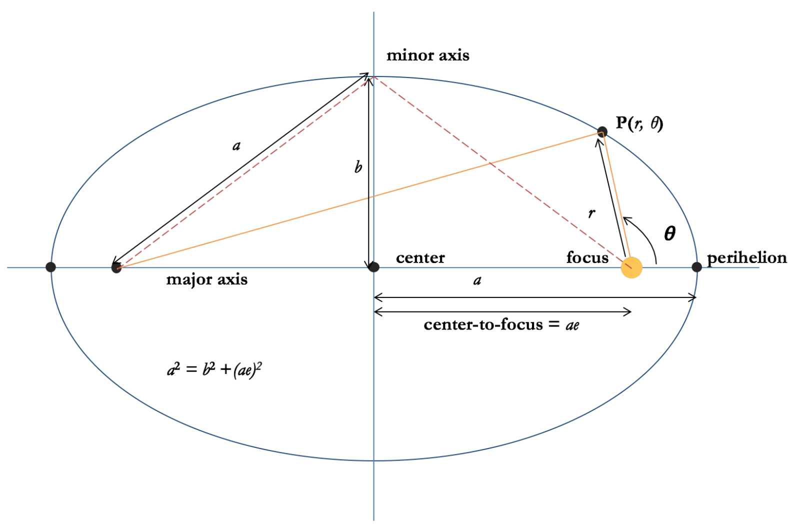
Kepler inherited were critically important. Studying the data for the positions of Mars, Kepler realized that the perfect circles had to go: Orbits are elliptical.

Recall conic sections: slice a cone at different angles and you get different types of figures. The eccentricity defines how much the figure deviates from a circle. An eccentricity of 0 = circle; $0 < e < 1$ is an ellipse; 1 = a parabola; $e > 1$ is hyperbolic. Ellipses have two foci (singular: focus) equidistant from the center; the sum of the distances from the foci to the ellipse is a constant. Align the ellipse with the long axis horizontally, select the right-hand focus, define r as the distance from that focus to the ellipse, and θ as the angle counterclockwise from the right. The semi-major axis, a, is half the length of the long axis. The eccentricity is defined as the ratio of the center – focus distance divided by the semi-major axis. The equation describing the ellipse is:

$$r = a(1-e^2)/(1+e\cos\theta).$$

The semi-minor axis gives us another way to consider the ellipse. The statement that the sum of the distances from the foci to the ellipse is constant can be expressed, thanks to Pythagoras, as $a^2 = b^2 + (ae)^2$. The area of an ellipse is given by $A = \pi ab$. The distances between one focus and the ends of the major axis are given by $r = a(1 \pm e)$. At the ends of the major axis, θ is either 0° or 180°, so cos θ is ±1 and the denominator becomes either $(1 + e)$ or $(1 – e)$. Expand the term in the numerator as $(1 + e)(1 – e)$; one of these factors is going to have to cancel with the denominator, leaving $r = a \cdot (1\pm e)$. Write it out if it's not obvious.

Figure 1.5



Back to Kepler. Kepler was considering the orbits of planets around the Sun. He did not have a model for *why* orbits should be elliptical and no knowledge of binary stars in orbit around each other or galaxies in orbit. The "laws" that Kepler formulated for orbits are empirical, i.e., they are fits to the data, but they make sense, whether for planets, stars, or galaxies, in terms of the formulation of gravity later developed by Newton. What Kepler said:

- orbits of planets are ellipses with the Sun at one focus (that'll morph into having the center of mass at one focus).
- the area swept out by r (i.e., by the line connecting the Sun to the ellipse) in a given time is constant (this turns out to be an expression of conservation of angular momentum); calculus: dA/dt = constant.
- the orbit period is related to the orbit size: $P^2 = \dfrac{4\pi^2}{G(m_1+m_2)}a^3$.

A note on this one: in solar system units, this reduces to $P^2 = a^3$; periods are in years, m_1 is one solar mass, the masses of the planets are effectively zero, and the distances are in astronomical units. One AU is the average Earth – Sun distance. We go around the Sun at one AU with a period of one year. Jupiter, at 5.2 AU, takes about 12 years. In those solar system units, with $M_\odot$ representing one solar mass, G, the gravitational constant, has a value of $4\pi^2$ AU^3/($M_\odot$ yr^2). This expression turns out to be related to the energy of the orbit, which should make sense: think about launching a rocket—it takes more energy to put it into a higher orbit.

Recall Newton's laws of motion:

- inertia exists
- $\vec{F} = m\vec{a}$ (or, more properly, $\vec{F} = \frac{d\vec{p}}{dt}$.)
- for each action there's an equal and oppositely directed reaction

For gravity, Newton's second law becomes $F = \frac{Gm_1m_2}{r^2} = gm_2$. When you work through the math you find that Kepler's rule for orbits describe accurately how orbits behave when subject to $1/r^2$ forces.

Often we can ignore the mass of one object; e.g., relative to the Sun, the mass of the Earth is approximately 0. When we can't do that, we'll need the relationship for center of mass: $m_1r_1 = m_2r_2$, where the total separation between the masses is $r = r_1 + r_2$.

How fast is an object going at any point in its orbit? Calculus alert: this requires noting that the velocity is the sum of its radial and angular parts and taking the derivative of the ellipse equation with respect to both r and θ and adding them together. The result is called the *vis viva* equation:

$$\mathrm{v}^2 = G(m_1 + m_2)\cdot\left(\frac{2}{r} - \frac{1}{a}\right).$$

(You should memorize this one.) Stop. Make sure you understand what all the symbols stand for. What is the expression for circular velocity? What about escape velocity (i.e., when your orbit is basically infinitely long)? What are the maximum and minimum speeds for an object in an elliptical orbit? Where in the orbit does an object have its max and min speeds?

Vocabulary: those max and min points have names. For a planet in orbit around the Sun, the closest distance to the Sun occurs at the near end of the major axis (i.e., where $\theta = 0$). It's called perihelion, usually denoted q. The opposite end of the major axis is called aphelion (Q). A planet has its highest speed at perihelion and its lowest at aphelion. The prefixes stay (almost) the same; the rest of the word tells you what you are orbiting:

Sun:	perihelion	aphelion
Earth:	perigee	apogee
Jupiter:	perijove	apojove
a star:	periastron	apastron
a galaxy:	perigalacticon	apogalacticon
generic:	periapse	apoapse

Those speed questions, above: note that for a circular orbit, r is always the same and $= a$; for escape velocity, a is infinite. This gives

$$\mathrm{v}_{\mathrm{escape}} = \sqrt{2GM/r}, \text{ and } \mathrm{v}_{\mathrm{circular}} = \sqrt{GM/r}.$$

Sometimes it's convenient to think of the interaction of two (or more) objects in terms of forces. Other times it will be more convenient to think about energies. And of course they are related: the work done by a force on an object results in a change in the object's kinetic energy. For our purposes kinetic energy can usually be expressed as $\frac{1}{2}m\mathrm{v}^2$. If we have two objects in orbit, the total kinetic energy of the system will be sum of the two individual kinetic energies. In terms of gravity, the mutual potential energy of the objects is $-\frac{Gm_1m_2}{r}$. (Physics: $\vec{F} = \nabla\varphi$.) The total energy for the system of two objects is the sum of the energies for both objects: $TE = \frac{1}{2}m_1\mathrm{v}_1^2 + \frac{1}{2}m_2\mathrm{v}_2^2 - \frac{Gm_1m_2}{r}$.

The problem of two bodies in orbit around each other can be reduced to an equivalent problem of one "reduced" mass object orbiting the center of mass. Here's how that works: the kinetic energies can be expressed in terms of momentum as $\frac{p^2}{2m_1} + \frac{p^2}{2m_2}$. The momenta have to be equal and opposite and then we square them, meaning that we don't need to distinguish p_1 from p_2. Cross multiply and the total kinetic energy becomes $\frac{p^2}{2}\left(\frac{m_1+m_2}{m_1m_2}\right) = \frac{p^2}{2\mu}$, where μ is called the *reduced mass*. In these terms, and setting $M = m_1 + m_2$, $TE = \frac{1}{2}\mu\mathrm{v}^2 - \frac{GM\mu}{r}$. In other words, the problem of m_1 and m_2 in orbit around each other is equivalent to the problem of an object of mass μ in orbit around a mass M at a distance $r = r_1 + r_2$.

In a closed system, the total energy is conserved. If we had a circular orbit the velocity would be $\sqrt{GM/r}$. Square this and you can see that the kinetic energy is half the gravitational potential energy. That statement is known as the *virial theorem*. Note also that the total energy could be expressed as $-\frac{GM\mu}{2a}$. Negative total energies imply that the system is bound; positive are unbound; 0 total energy would be a parabolic orbit, i.e., one in which the velocity just = the escape velocity. Plug this expression for total energy to the one in the previous paragraph, solve for v and cancel μ:

$$-\frac{GM\mu}{2a} = \frac{1}{2}\mu\mathrm{v}^2 - \frac{GM\mu}{r} \rightarrow \mathrm{v}^2 = G(m_1+m_2)\left(\frac{2}{r} - \frac{1}{a}\right).$$

The *vis viva* equation is a statement of energy conservation.

About the virial theorem, that bit that says that the kinetic energy equals ½ the potential energy: it's handy. Remember it. It will crop up in several places. A prime example is that of modeling the formation of a spherical system by imagining that the pieces fall together from infinity. The "system" could be a giant planet, a star, or a cluster of galaxies, whatever. The particles have some potential energy; to come to a stable final configuration they must radiate away ½ of that energy. The other half will go into their final kinetic energy. This principle will govern the time it takes for a star to form – how long will it take to radiate away half the potential energy? It will permit a means of estimating the mass of a cluster of galaxies – the average kinetic energy can be estimated by measuring the galaxies' velocities; since that kinetic energy came from the half of the gravitational potential energy that didn't get radiated away and since the potential energy depended on the total mass, we can estimate the mass of the cluster. Very nifty.

Examples:

1. The semi-major axis of the orbit of Pluto is 39.5 AU; it has an orbital eccentricity of 0.25 (i.e., a = 39.5 AU, e = 0.25).

   a) How long does it take Pluto to orbit the Sun? We can use the simple version of Kepler's third law, $P^2 = a^3 \rightarrow \sqrt{(39.5\ \text{AU})^3}$ = 248 years

   b) How close does it get to the Sun, i.e., what is its perihelion distance?
   $q = a(1-e)$, or q = 39.5 AU (1–0.25) = 29.6 AU

   c) What is Pluto's perihelion velocity? For this we need $\text{v}^2 = G(m_1+m_2)\left(\frac{2}{r}-\frac{1}{a}\right)$.

   The mass is basically the mass of the Sun (2 x 10^{30} kg or 1 solar mass), since Pluto is so small; r is the perihelion distance and a is the semi-major axis. Let's do this in two different sets of units:

$$\text{v} = \sqrt{6.67\cdot 10^{-11}\frac{\text{m}^3}{\text{s}^2\text{kg}}\cdot 2\cdot 10^{30}\text{kg}\left(\frac{2}{29.6\text{AU}}-\frac{1}{39.5\text{AU}}\right)\cdot\frac{1\text{AU}}{1.5\cdot 10^{11}\text{m}}} = 6100\ \text{m/s}$$

$$\text{v}=\sqrt{4\pi^2\frac{\text{AU}^3}{\text{yr}^2\text{M}_\odot}\cdot 1\text{M}_\odot\left(\frac{2}{29.6\text{AU}}-\frac{1}{39.5\text{AU}}\right)} = 1.3\ \text{AU/yr.}$$

   Are these results the same? 4.74 ((km/s) / (AU/yr))·1.3 AU/yr = 6.1 km/sec. Yes.

2) Consider a binary star w/ components having 0.5 and 3.0 solar masses, separated by 280 AU, in a circular orbit.

   a) What is their orbit period? Let's use $P^2 = a^3/M$ in solar system units:

$$\sqrt{\frac{(280\text{AU})^3}{(0.5+3.0)\text{M}_\odot}} = 2500\ \text{yr.}$$

   The mass isn't = 1, so we have to include it explicitly.

   b) Where is the center of mass located? Recall that $m_1r_1 = m_2r_2$. "Separated by 280 AU" means that $r_1 + r_2 = r$ = 280 AU. Solve for r_1 and then r_2. Because m_1 is 6 times more massive than m_2, m_2 must be 6 times farther from the center of mass.

$$r_1 = \frac{m_2}{m_1}r_2 = \frac{m_2}{m_1}(r-r_1) \rightarrow r_1\left(1+\frac{m_2}{m_1}\right) = \frac{m_2}{m_1}r \rightarrow r_1\left(1+\frac{0.5}{3.0}\right) = \frac{0.5}{3.0}280\ \text{AU} \rightarrow r_1 = \frac{46.67}{1.167} = 40\ \text{AU}$$

$$r_2 = r - r_1 = 280\ \text{AU} - 40\text{AU} = 240\ \text{AU}$$

   The stars are 40 and 240 AU from the center of mass. Factor of 6? Yes, check.

   c) What's the total velocity difference between the two? Let's do this two ways:

$$\text{v} = \text{v}_1 + \text{v}_2 = \frac{2\pi r_1}{P} + \frac{2\pi r_2}{P} = \frac{2\pi}{2500\text{yr}}\cdot 280\ \text{AU} = 0.7\ \text{AU / yr;}$$

$$\text{v} = \sqrt{\frac{GM}{r}} \rightarrow \text{v} = \sqrt{\frac{4\pi^2(3.5\text{M}_\odot)}{280\text{AU}}} = 0.7\ \text{AU / yr.}$$

   Note that the velocities are divided in the same ratio as the masses: 0.1 : 0.6.

3) The Andromeda galaxy (a.k.a. M31) is about 2.2 million light years away. Within the inner 0.2″ we see material with a maximum velocity of about 240 km/s. Approximately how many solar masses worth of material must be in the core of M31?

a) We need to convert an angle of 0.2″ at the distance of M31 into a linear distance within M31. $r = d \tan \theta$. Our angle is very small, though, so you can ignore the tan if you wish. Recall that there are 60 arcsec in 1 arcmin and 60 arcmin in 1 degree. For converting d, look up a light year and find that it is about 9.5 10^{15} m.

$$r = (2.2 \cdot 10^6 \text{ly} \cdot 9.5 \cdot 10^{15} \text{m / ly}) \cdot \tan\left(\frac{0.2''}{3600''/\circ}\right) = 2.0 \cdot 10^{16} \text{m}.$$

b) Now apply the circular velocity equation:

$$M = \frac{v^2 r}{G} = \frac{(240 \cdot 10^3 \text{m / s})^2 \cdot 2 \cdot 10^{16} \text{ m}}{6.67 \cdot 10^{-11} \text{ m}^3 \text{ / s}^2\text{kg}} = 1.25 \cdot 10^{37} \text{ kg}.$$

Next, convert this to solar masses:

$$1.25 \cdot 10^{37} \text{ kg} \cdot \frac{1 \text{ M}_\odot}{2 \cdot 10^{30} \text{ kg}} = 6 \cdot 10^6 \text{ M}_\odot.$$

This is a lot of mass in a very small volume; it's highly likely that there's a black hole at the center of M31.

c) How long would it take material at this distance to orbit the center of M31?

That's a Kepler's third law question. Let's do it in solar system units; first, we'll convert the distance from meters to AU:

$2.0 \cdot 10^{16}$ m (1 AU / $1.5 \cdot 10^{11}$ m) = $1.3 \cdot 10^5$ AU;

$$P^2 = \frac{4\pi^2}{G(m_1 + m_2)} a^3 \rightarrow P = \sqrt{\frac{(1.3 \cdot 10^5 \text{ AU})^3}{6 \cdot 10^6 \text{ M}\odot}} = 19 \cdot 10^6 \text{ yr}.$$

Lagrange points

As soon as you move beyond two masses the math gets much messier. The next most straightforward problem is to look at two significant masses in circular orbits around each other and ask what would happen to a small test mass in the orbital plane of the other two. In other words, the test mass is being acted on by the gravitational attraction of both large masses and it is in a rotating system. There are two extremes, where the problem reduces pretty well to a two-body problem: if the test mass is very close to one large mass, then we can ignore the second mass (that's like considering the Moon's orbit around us while ignoring the Sun); if the test mass is very far away from both large masses, then they behave as if they were one, located at their center of gravity. In other words, we consider centripetal plus one gravitational force. Anywhere in between, you have to balance two gravitational forces plus the centripetal. There are some interesting cases, studied by, and hence named for, Lagrange. He found that there are five stable points, in the orbital plane of the two large masses, where we could put our test mass and have it move in a stable circular orbit.

Points L4 and L5 make equilateral triangles with the two large masses. In the Sun – Jupiter system, for instance, there are stable points in Jupiter's orbit, ±60° from Jupiter; the Trojan asteroids hang out at / near these points (right, there's only room for one asteroid really at either L4 or L5; what we see is that the asteroids librate, oscillate, around L4 and L5, kind of as if it were an attractor).

Points L1, L2, and L3 are a bit less stable and are along the line joining the two large masses. L1 and L2 are inside and outside the smaller mass, respectively, and L3 is on the far side of the large mass. Physics note: here's what an equipotential plot looks like for the Sun-Earth system:

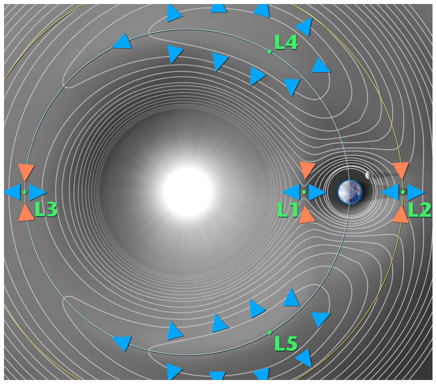
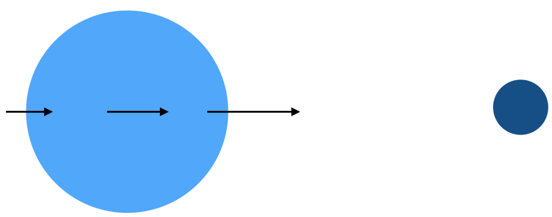


Figure 1.7:
Sun-Earth equipotential plot; credit: NASA / WMAP Science Team
https://map.gsfc.nasa.gov/media/990529/990529.jpg

Where we will see this: We have spacecraft placed at both L1 and L2 in the Sun-Earth system, safely out of the way of infrared emission from Earth but still orbiting the Sun with a period of one year; when we have two stars in a close binary system, we may get mass transfer and it will be through the L1 point; and, as noted above, asteroids settle into the L4 – L5 spots.

Tides

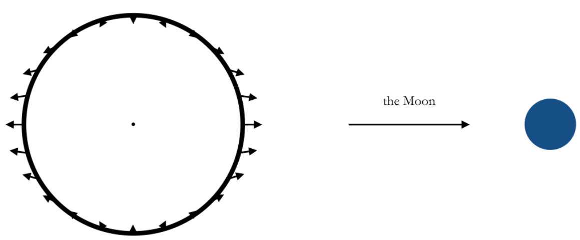
Our initial treatment of gravity assumes that objects are point masses. In other words, we assume that the difference between the higher gravitational attraction on the near side of an object balances out with the lesser gravitational attraction on the far side. But real objects aren't totally rigid and will deform in response to a differential gravitational pull. Consider the Earth and Moon, not to scale!; the arrows represent the gravitational force on the Earth due to the Moon at three different locations:



The gravitational force on the Earth due to the Moon

Figure 1.8

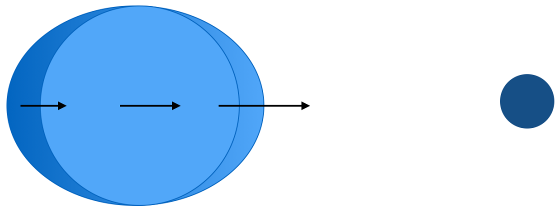
If we look at the difference in the gravitational attraction on a test mass in several locations at the surface of the Earth, it looks like this:



The differential gravitational force on a test mass at various locations on the Earth due to the Moon

Figure 1.9

The Earth's oceans can react fairly readily to the difference in the force and flow toward the location of highest force:



Ocean response to the differential gravitational force.

Figure 1.10

Tidal bulges are raised in the oceans – the two high tides – by the difference in force. You can calculate the difference in the force either using algebra or calculus—either take $\vec{F}_{near} - \vec{F}_{far}$ or take $\frac{d\vec{F}}{dr}$, where $\vec{F}$ is the Newtonian force of gravity. What you'll find is that the tidal force goes down as r^3 and that the difference is larger the larger the diameter across the object (in this case the Earth) is.

Tides get raised in solid objects, as well, as long as they can deform a bit. The Moon, for instance, is a bit elongated in the direction toward the Earth. Tidal flexing due to Jupiter is *the* mechanism for supplying the heat to power the volcanoes on Io.

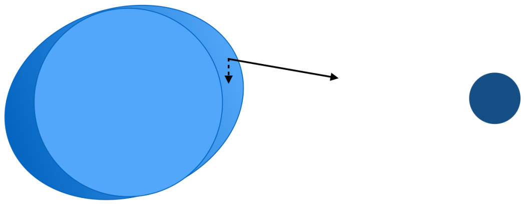
Tides matter for making planetary rings: imagine trying to place a moon-sized readily deformable ("liquid") object in a very tight orbit around Saturn. It will deform. And its outer edge will "want" to orbit around Saturn more slowly than its inner edge (think Kepler's 3rd law). A moon-sized object will shred. Farther out? the difference in the gravity across the satellite will be less the farther away it is. Smaller satellite? the difference in gravity across the satellite will be less the smaller it is. Planetary rings tend to be in close and made of smallish chunks of stuff. How close is close and how small is small depends on what the satellite is made of. The distance within which you can't have big satellites is called the Roche limit; a rough rule of thumb is that it's about twice the radius of the planet.

A bit better: $r_{Roche} = 2.44\left(\frac{\rho_{planet}}{\rho_{satellite}}\right)^{1/3} R_{planet}$, which you get by assuming that the "satellite" is a "liquid" made of two particles held together only by their mutual gravitation and setting that gravitational attraction equal to the differential gravitational force and solving for the distance. Moons with more structural integrity can get closer than this.

Tides will also matter for things like galactic interactions; when two galaxies move past each other, we may see long tidal tails of stars being pulled out from one or both galaxies. Streams of stars have been identified in our galaxy, looking very much as though they used to be small satellite galaxies that have been tidally shredded by the more massive Milky Way.

We will find tides coming into play also when we consider close binary stars. Some star pairs orbit each other so closely that the differential gravitational force deforms them, makes them more egg-shaped than spherical.

Back to the idealized picture of the Moon raising tides on the Earth. The Earth is rotating and there is friction between the water and the ocean floor. The tidal bulge raised by the Moon thus gets ahead of the Earth-Moon line. The gravitational attraction between the tidal bulge and the Moon exerts a torque in the direction opposed to the direction of the Earth's rotation. The Earth and Moon swap some angular momentum.



Torque on Earth due to tidal bulge

Figure 1.11

The Earth's rotation rate is very slowly slowing down. The Moon is very slowly moving farther away from the Earth. Since the forces involved are distance dependent, the rates of the slowing of the day and the lengthening of the month haven't been constant throughout the history of the Earth-Moon system; in other words, we can't just run the movie backwards linearly. Regardless, we can get some help from paleontologists: there are some sea critters whose shells show daily and/or monthly growth bands, allowing us to figure out how long the day or month were using the fossil record. About 45 million years ago the synodic month was 29.1 days long, almost a half day shorter than at present. The evidence suggests that 2.8 billion years ago, the month was only 17 days long. Right now the Moon's moving away at about 3 cm per year; we expect that the day and month will come into synch when each is about 47 times as long as the current day. We'll be like Pluto and its moon Charon, which are tidally locked on each other, both keeping the same face always toward the other. It won't last, though. There's a tide from the Sun, too, which will eventually act to drag the Moon back in toward the Earth.

Real ocean tides on Earth depend on the phase of the Moon, the location on Earth, the underwater topography, etc., etc. Tides will be more extreme if the Moon is New or Full, because the Moon and Sun will be acting along the same line to raise the tide. Probably the most famous example of extreme tides is in the Bay of Fundy, in Newfoundland:

Figure 1.12: Bay of Fundy tides; pictures by Samuel Wantman, in 1972.
By © Samuel Wantman / Wikimedia Commons, CC BY-SA 3.0 https://commons.wikimedia.org/w/index.php?curid=225282 and https://commons.wikimedia.org/w/index.php?curid=225283

More celestial mechanics, specifically pertaining to the solar system, may be found in Chapter 2 (Solar system dynamics).

Relativity

There are times in astronomy when we need to deal with particles that are moving very rapidly (particles other than photons) or objects that are very massive (like black holes) or both (like massive particles in the first few seconds of the universe), so we need to look just a bit at relativity. There are two pieces: Einstein published the special theory of relativity in 1905, which deals with high speed, and the general theory of relativity in 1915, which deals with accelerated motion and is thus the part where we get into dealing with gravity.

The basic principle of special relativity is that the laws of physics should look the same to two observers who are in relative motion. The classic thought experiment is to consider someone in a moving train car performing the same experiments—bouncing light off mirrors, tossing a ball back and forth, swinging a pendulum, whatever—as someone standing on the train station platform. If they do the same things they should get the same results. Everything in the train car is moving together—we assume it moves *very* smoothly—so the train car and its occupants might as well be standing still and the station platform moving. Or both moving at half speed in opposite directions. They are definitely in relative motion—each observer can see the other move past—but the laser pointer, the baseball, the grandfather clock all work the same within the train car as they do on the platform. As you probably know, this has some interesting and counterintuitive consequences. For one thing, there is no absolute frame of reference in the universe, no box whose walls we can use to tell us who's moving and who's standing still on the platform. All motion is relative. Another, critically important, consequence is that all observers measure the same thing for the speed of light. Think about watching *from the platform* as two people in the train car throw a baseball back and forth. Thinking classically, we are accustomed to adding the speed of the train and the speed of the baseball; relative to us on the platform, the baseball is going faster when thrown with the direction the train is moving than when thrown against it. Light isn't like that. Everyone measures the same thing for the speed of light whether the laser pointer is shining in the direction of the train's motion or against it.

Historical note: The first good measurement of the speed of light was made in ~1676 by Danish astronomer Ole Rømer. He timed the eclipses of Jupiter's moon Io, the frequency of which could be predicted from Kepler's laws. Rømer reasoned that when Jupiter was near conjunction with the Sun the light travelling from Jupiter had to go 2 AU farther than when Jupiter was at opposition. His estimate of the travel time across 2 AU was a bit large, and we did not yet know how many kilometers there were in an AU, but nonetheless, it was a very creditable achievement.

If we all agree on the speed of light then we must disagree about measurements of distance and time. Standing on the platform, we think that the observer in the train moving past has clocks that run too slowly and meter sticks that are too short. Specifically, time is dilated and length contracted by a factor of $\gamma = \left(\sqrt{1-\frac{v^2}{c^2}}\right)^{-1}$. Looking back at us, they think our clocks and meter sticks have exactly the same problem.

Here's a classic example to demonstrate the time dilation: Our observers in the train car have a clock consisting of a mirror on the ceiling of the train car and a device that can send a pulse of light from the floor to ceiling and back. When the light pulse returns to the floor the clock records one unit of time. Here's a sketch to show what the observers in the train car record when they measure the amount of time it takes the light to bounce up and back one time:

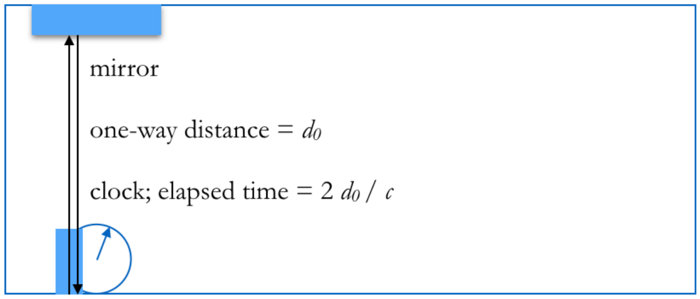


Figure 1.13

Observers, clock, mirror, and light pulse are all in the same frame of reference, i.e., not moving relative to one another; the time measured thusly is called the *proper time*.

Now imagine that the train car is moving toward the right with a speed v with respect to stationary observers on the platform. As the light pulse is going from floor to ceiling, the mirror and the clock device are moving to the right as seen from the platform. Here's a sketch of what these observers see:

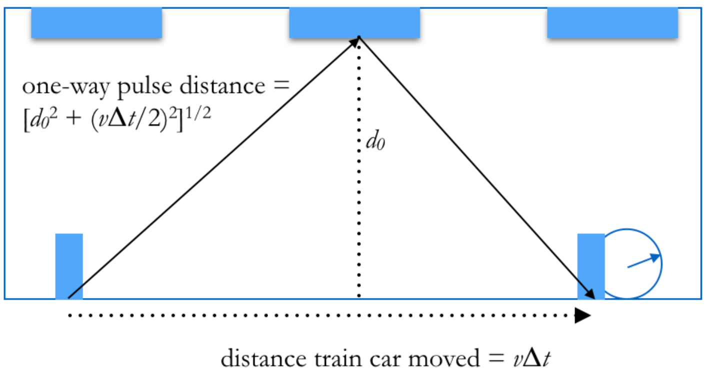


Figure 1.14

The observers on the platform measure the light pulse to have traveled a greater distance than the distance as measured by the observers inside the train car:

$$2\cdot[d_0{}^2 + (v\Delta t/2)^2]^{1/2} > 2d_0.$$

Einstein's realization was that if both sets of observers agree that physics works, i.e., that the speed of light in a vacuum is constant, then the disagreement about the distance the light pulse travels requires that the observers also must disagree about the time it takes for the light pulse to bounce off the mirror and return to the floor. The time measured by the stationary platform observers is longer than the time measured by the observers moving along with the train car:

$$\Delta t_{\text{recorded on the stationary platform}} = \frac{\Delta t_{\text{recorded in the moving train car}}}{\left(1 - v^2/c^2\right)^{1/2}}.$$

Time dilation happens.

So does length contraction. Suppose that this time the light pulse travels lengthwise along the train car, i.e., in the same direction as its motion rather than transversely to its motion. Again, the clock registers one tick when the pulse returns to the source. Observers moving with the train car see this:

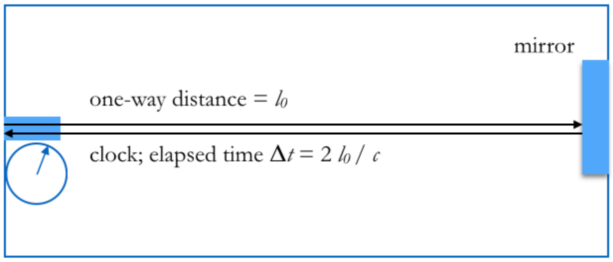
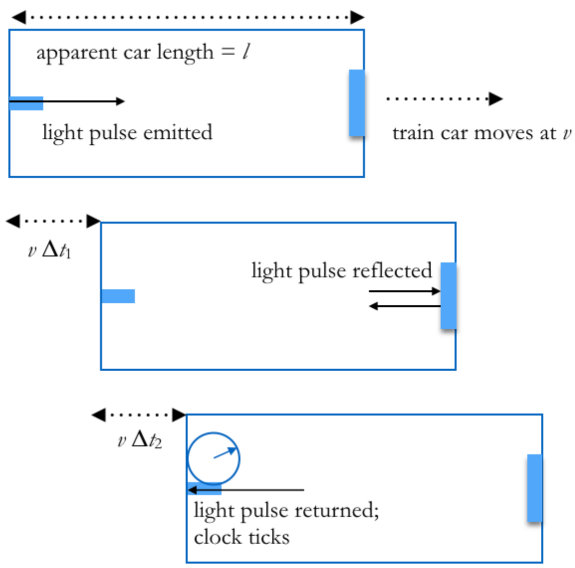


Figure 1.15

Now consider what the stationary observers on the platform record:

Figure 1.16



How far do they record the light pulse as having travelled on its outbound trip?

$$l + v\Delta t_1 = c\Delta t_1$$

And on the return trip, the light travels

$$l - v\Delta t_2 = c\Delta t_2 .$$

The total elapsed time recorded by the platform observers is

$$\Delta t_{\text{recorded by stationary platform observers}} = \Delta t_1 + \Delta t_2 = \frac{l}{c-v} + \frac{l}{c+v} = \frac{2lc}{c^2 - v^2} = \frac{2l/c}{\left(1 - {v^2}/{c^2}\right)} .$$

But we already know that the elapsed time recorded by the platform observers differs from the elapsed time recorded by observers moving along with the train car:

$$\Delta t_{\text{recorded on the stationary platform}} = \frac{\Delta t_{\text{recorded in the moving train car}}}{\left(1 - {v^2}/{c^2}\right)^{1/2}} = \frac{2l_0/c}{\left(1 - {v^2}/{c^2}\right)^{1/2}} .$$

Use these two last equations to eliminate the time and solve for the relationship between the distance as measured by the observers on the platform (l) and the distance as measured by the observers moving along with the train car (l_0):

$$l = l_0 \left(1 - {v^2}/{c^2}\right)^{1/2} .$$

The observers on the platform measure the train car length to be *less* than the length as measured by the observers riding along with the train car.

In a very cool demonstration, folks in New Hampshire observed the numbers of muons at the summit and at the base of Mt. Washington, elevation ~6,300 feet. These muons were formed in the upper atmosphere by collisions between atmospheric particles and high-energy cosmic rays. (Cosmic rays are high-energy particles usually arriving from outside the solar system, sometimes from outside our galaxy, although the term could include particles in the solar wind as well.) Muons decay in about 1.5 10^{-6} s. They travel at nearly the speed of light, we know how many were observed at the top of Mt. Washington and how far it is to the base. Where, yup, we observe way too many muons. Why? Relative to us, their clocks are running slow (or, if you prefer, they don't think they've traveled nearly as far as we think they have). An observer traveling along with the muons observes them decaying at their own normal rate; we observe them decaying more slowly because of the time dilation.

Another consequence of special relativity is the famous expression $E = mc^2$. Particles have a rest mass, often denoted m_0; if they are moving, their kinetic energy is equivalent to saying that their total mass is larger than their rest mass. Unless, again, we are speaking of photons, packets of energy whose rest mass is 0. Consequences of this mass-energy relationship we will talk about when we talk about nuclear fusion or about the creation of particle – anti-particle pairs.

The general theory of relativity deals with, yes, the more general case: rather than simply uniform motion, the general theory tackles particles that are being accelerated. Turn the train car of the previous thought experiment on its tail and morph it into a spacecraft, preferably one with no windows. Our observers in the spacecraft drop a ball and note that it accelerates downward at 9.8 m/s^2. Assuming that this spacecraft has a *very* smooth propulsion mechanism!, there are two possible interpretations: either the spacecraft is at rest on the surface of the Earth or it is being smoothly accelerated at a rate of 1 g. Alternatively, suppose that our observers find that nothing falls, as if they are experiencing no gravity. Two interpretations: they could be in orbit around the Earth, falling around the Earth along with their spacecraft and everything in it, or they could be far off in space, away from the gravitational influence of anything else. Einstein's Principle of Equivalence says that acceleration is acceleration—gravity is not different from the acceleration of the spacecraft. Or, equivalently: inertial mass and gravitational mass are the same thing. We use them the same way—no one stops to note that it's the same m in $\vec{F} = m\vec{a}$ as in $\vec{F} = \frac{GMm}{r^2}$. On the other hand, you could consider them differently—just as electrical charge is the property that particles which respond to the electromagnetic force have, gravitational mass is the "gravity charge" that particles which respond to the force of gravity have. It isn't obvious that that kind of mass is the same thing as the mass that resists changes in motion.

The wildly counterintuitive part here? It doesn't matter whether our observers were dropping a ball or a beam of light. Acceleration applies to light, too. But light travels in "straight lines", doesn't it?. In General Relativity it makes more sense to talk about massive objects curving the space around them than to treat gravity as a force; light travels along a geodesic, which is the shortest path between two points. It makes sense in relativity to speak about time as another dimension, and spacetime as having dimensions r, θ, φ, and ct. The math for describing the distance between two points is, as you might expect, more complicated in General Relativity because mass-energy and the curvature of spacetime are intimately intertwined. Calculus alert: Einstein's Field Equations involve tensors; they are a set of 10 coupled nonlinear partial differential equations. Working with these equations is not trivial.

The first more-or-less acceptable demonstration of the validity of General Relativity came from observations made during a solar eclipse in 1919. A star of known position, the light from which passed near the limb of the Sun during totality, was observed to be slightly farther away from the Sun, indicating that the path of the light passing the Sun was curved. Neither the precision or accuracy of the measurements was really adequate, but it was a step in the right direction. (Note that you could mix together Special Relativity and Newtonian mechanics and get a deflection of the light—ask what the mass equivalent is of the light energy and how much that mass would be deflected by the Newtonian gravity of the Sun; the answer is half the angle that General Relativity predicts.) The modern application of this example is the gravitational lens: light from a distant object passing a massive object on its way to us will get bent. An example is when light a very distant galaxy passes a medium distance galaxy or cluster of galaxies; we may see the distant galaxy multiple times, and distorted in shape. This lets us work out the mass of the intervening galaxy / galaxy cluster. The following Hubble Space Telescope image of a galaxy cluster called Abell 68 demonstrates this lensing effect:

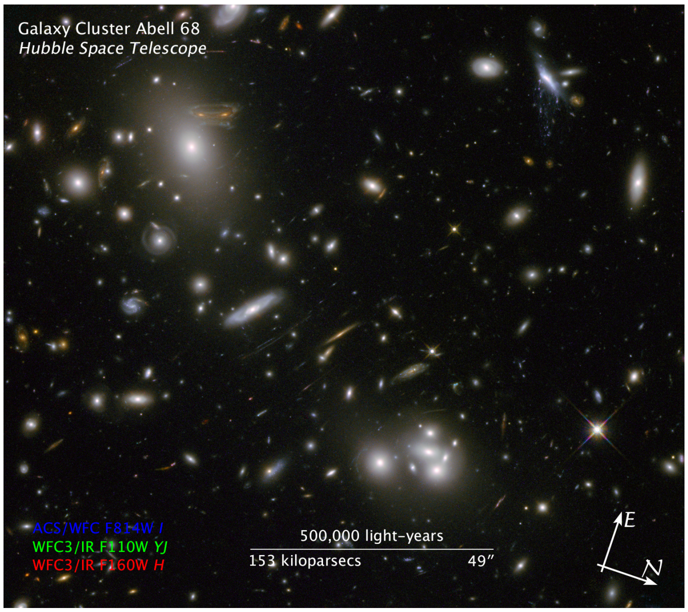


Figure 1.17

Foreground galaxy cluster Abell 68 is ~2.1 billion light years away; the mass in this cluster is causing the light from more distant galaxies to appear curved.

Credit: NASA, ESA, and Z. Levay (STScI)
http://hubblesite.org/image/3157/news_release/2013-09

Other consequences: black holes are the ultimate in gravitational redshifts—light climbing out of a gravitational well will lose some energy which means that we will observe it as redder than it was when it was emitted. It's been measured at the surface of the Earth, although it's a lot easier to observe in light leaving very compact objects such as neutron stars or white dwarfs. Black holes simply have infinite redshift.

In the fall of 2015 LIGO (the Laser Interferometer Gravitational wave Observatory) made the first detection of gravitational radiation. Just as accelerated electric charges radiate electromagnetic radiation, accelerated masses should emit gravitational radiation. In particular, two massive objects in a tight orbit should emit gravitational energy, causing the fabric of spacetime itself to stretch and oscillate as the wave passes. Prior to 2015 we had observed the energy loss: as an example, there is a binary pulsar, PSR1913+16, which is two solar-ish mass objects in a very small elliptical orbit (really small, about an 8-hour revolution period). The pulsar one of the pair emits regular radio pulses, so we can track the orbit. It's decaying, just as predicted. (This 1974 discovery led to the 1993 Nobel Prize in Physics for R.A. Hulse and J.H. Taylor.) The search for actual gravitational waves themselves came to fruition almost exactly 100 years after Einstein's publication of the theory of general relativity. LIGO is a pair of interferometers, located in Livingston, Lousiana, and Hanford, Washington. A source shoots laser light down two perpendicular four kilometer-long evacuated arms of each interferometer. The light bounces off very stable mirrors and back down the tubes to be recombined. When a gravitational wave passes first one arm and then the other will be stretched and compressed by a minute amount, changing the interference pattern as the laser beams recombine. The best estimates are that the event detected on 14 September 2015 was due to two black holes spiralling together at ~0.6 *c* and merging, over the course of ~20 milliseconds, about 1.3 billion years ago (yes, in a galaxy far, far away). The initial masses were ~ 29 and 36 $M_{\odot}$. The merger created a black hole of about 62 solar masses, with three solar masses converted entirely to energy, rippling away across the cosmos. The meager amount of gravitational energy passing Earth expanded and contracted the planet by about the size of the nucleus of a helium atom; the 4-km arms of the LIGO interferometers expanded and contracted by $\sim 10^{-18}$ meters. The detection was announced on 11 February 2016.

Forces and elementary particles

Let's pause for a moment here to discuss forces and particles. There are four fundamental forces: gravity, the electromagnetic force, the strong nuclear force (which holds nucleons together), and the weak nuclear force (which governs radioactivity). Both gravity and the EM force fall off as $1/r^2$; the range of the two nuclear forces is very short, meaning they aren't going to be as obvious in your daily life. Note that gravity is a bit odd: in particle physics it's a force and there's a gravitational field; in general relativity it's a warping of the fabric of spacetime. We do not yet have a successful theory of quantum gravity to unite these two models.

You should be familiar already with the idea that atoms are composed of electrons around nuclei made of protons and neutrons, and that molecules are collections of atoms bound together. But let's lay out where those pieces fit in the current standard model of elementary particles and forces.

There are two types of particles called *fermions*: leptons, of which the electron is the familiar lightweight representative; and the quarks, of which the up and the down are most familiar because they make up protons and neutrons.

All particles have a property called "spin" (because it bears some tiny resemblance to the spin of a top). Fermions have half-integer spins; they obey the Pauli Exclusion Principle, which says that you can't put two of the same particle in the same state at the same time.

Photons, by comparison, are particles of a type called *bosons*; there are no restrictions on how many photons you can shove into one space at the same time. The particles that carry (or "mediate") the fundamental forces of nature are all bosons.

The Contemporary Physics Education Project has excellent charts of particle and nuclear physics and cosmology at http://www.cpepweb.org/. The data in the following tables comes from this site.

Table 1.1 a: Standard model particles

fermions (1/2 integer spin)

Leptons (spin =1/2)			Quarks (spin =1/2)		
flavor	mass (GeV/c^2)	electric charge	flavor	mass (GeV/c^2)	electric charge
e (electron)	0.000511	-1	u (up)	0.002	2/3
ν_L (lightest neutrino)	(0-2) $\cdot 10^{-9}$	0	d (down)	0.005	-1/3
μ (muon)	0.106	-1	c (charm)	1.3	2/3
ν_M (middle neutrino)	(0.009-2) $\cdot 10^{-9}$	0	s (strange)	0.1	-1/3
τ (tau)	1.777	-1	t (top)	173	2/3
ν_H (heaviest neutrino)	(0.05-2) $\cdot 10^{-9}$	0	b (bottom)	4.2	-1/3

force-carrying bosons (spin = 1)

flavor	mass (GeV/c^2)	electric charge	interaction
graviton (not yet observed)			gravity
W^-	80.39	-1	weak
W^+	80.39	1	weak
Z^0	91.19	0	weak
γ (photon)	0	0	electromagnetic
g (gluon)	0	0	strong

Higgs	126	0	scalar boson (spin = 0)

Table 1.1 b: interactions (forces)

interaction	acts on	relative strength at 10^{-18} m	at $3 \cdot 10^{-17}$ m
gravity	mass-energy; i.e., all particles	10^{-41}	10^{-41}
weak	flavor; i.e., quarks & leptons	0.8	10^{-4}
electromagnetic	all particles with electric charge	1	1
strong	color charge; i.e., quarks & gluons	25	60

The bottom table show which force acts on particles having which properties. We will return to these charts several times, most importantly when we get to discussing nucleosynthesis and the first few minutes of the universe. (For instance, you probably have heard about the Higgs boson; you might not be totally solid on why you should care.)

In addition to the particles, there are also anti-particles; for instance, the positron is just like an electron except with a positive electrical charge. Most of the high-mass particles are unstable, created in high-energy collisions in particle accelerators, and have only a fleeting existence before decaying to some lower mass particle. Why they have the masses they do, though, is an interesting hot topic question in high-energy physics and cosmology.

Quarks are held together in twos and threes (and perhaps fours and fives) by the strong force to make things called hadrons. Hadrons made up of a quark – anti-quark pair are called mesons. The three-quark combinations are called baryons and include protons and neutrons, which are combinations of uud and udd quarks, respectively. The property of the quark that makes it susceptible to the strong force is called "color" (having nothing whatsoever to do with what you normally think of as color!). Think of electrostatic attraction or repulsion—it acts on things that have electric charge; the comparable property for the strong force is color charge. We say that quarks have color charges of red or green or blue. The resulting 3-quark nucleon is color neutral, just like an atom, composed of negative electrons and an equal number of positively charged protons, is electrically neutral.

The strong force binds quarks together to make nucleons; its influence doesn't end abruptly at the "edge" of the nucleon, though, so it also serves to hold the nucleons together into an atomic nucleus. Similarly the electromagnetic force that holds electrons in orbit in an atom also extends beyond the "edges" an atom to hold atoms together into molecules. The tables above give the quarks rest masses and you might note that the sum of the rest masses of three up and/or down quarks is only about 1/100 the mass of a proton or neutron. The energy involved in the strong force and the gluons holding the quarks together contributes far and away the majority of the mass of a nucleon.

Why particles have the properties that they do is not a question that we can yet answer. One area of investigation that looks cautiously promising is string theory. String theory suggests that what we interpret as point particles are actually tiny (very tiny) vibrating strings. The frequencies of vibration determine the particles' properties, e.g., electric charge, color charge, mass, etc. In string theory the universe has 10 or 11 dimensions, most of which are curled up tightly so that we don't notice. The example here is of a garden hose or a thick rope: seen from a large distance, it appears to be one dimensional; seen from the point of view of an ant crawling around on the surface there's an extra "around" dimension. Brian Greene's 2000 book *The Elegant Universe* is an excellent and accessible introduction to string theory. On the experimental front, one of goals for the Large Hadron Collider is to address the question of why particles have the properties that they do.

Vocabulary associated with atoms: an atom is whatever element it is because of the number of protons in its nucleus. A neutral atom will have as many electrons in orbit around the nucleus as it has protons in the nucleus. If we add or remove one or more electrons, making the atom electrically charged, we call it an ion. In the nucleus with the protons are some number of neutrons. Neutrons are electrically neutral and just a tad heavier than protons. In light atoms there tend to be about as many neutrons as protons; e.g., the most common type of carbon has 6 protons and 6 neutrons. As we get to the heavier elements we find relatively more neutrons, which makes sense—the protons' electrical charges repel each other, so to hold a large nucleus together you need more neutrons to add more strong force "glue". Most elements have several different possibilities for the number of neutrons; we call these different forms isotopes. Notation: ^{A}X, with A = Z + N, means element X with atomic mass A composed of Z protons plus N neutrons. We call it "atomic" mass when we probably should call it "nuclear" mass, but since the electrons are ~2000 times less massive than the nucleons, we just tend to ignore them. Just to be confusing, you may also see this written X^{A}; it means the same thing as ^{A}X. A subscript after the X means we're talking about a molecule. A ± n superscript after the X or a Roman numeral after the X, e.g., X^{++}, means we've got an ion.

Example: A hydrogen nucleus typically consists of a lone proton with no neutrons. Hydrogen also exists in a stable form with one neutron, called deuterium, and an unstable form with two neutrons, called tritium. In terms of the notation, those are ^{1}H, ^{2}H, and ^{3}H, respectively. A diatomic hydrogen molecule would be written H_2. An ionized hydrogen would be H^{+} or H II. Note that that last bit could be confusing—where the chemist would use one + to refer to a single ionization, spectroscopists use II.

The vast majority of atoms (by number), like 90-ish %, in the universe are hydrogen, followed by helium, and then small and smaller amounts of everything else. It can make for an abbreviated periodic table: hydrogen, helium, "metals". Another notation thing: composition, e.g., of a star, is referred to as *X, Y, Z*, where those are the mass percentages of hydrogen, helium, and metals. By mass, the universe is about 70% H, 28% He, and 2% everybody else. Note that we are starting to recycle letters: don't get the composition *Z* confused with the atomic number Z!

Electrons are not little planets in orbit around a nuclear star; they are more like a fuzzy distribution of charge than like little BBs. Nonetheless, we can do an awful lot by thinking of electrons as little particles in planet-like orbits. One of the big challenges facing early modern physics, say, turn of the 20th century, was the realization that accelerated charges radiate. (That's what makes radio possible, wiggling electrons back and forth.) If you put an electron in an orbit around a nucleus it will be continually accelerated (changing direction), should radiate, should lose energy, should spiral down into the nucleus. That was clearly not what electrons were doing. In 1913 Niels Bohr (1885-1962) made the astonishingly simple proposal that electrons occupy only stable, non-radiating orbits. What makes an orbit stable? its associated angular momentum. Bohr's proposal was that the electron's orbital angular momentum must be an integer multiple of $\hbar$ (= $h/2\pi$), Planck's constant. In other words, permitted electron orbits are quantized. This is the birth of quantum mechanics. We tend to think of electrons as particles, but they have wave properties, too; in terms of wavelengths, the permitted orbits are those where you can fit an integer number of wavelengths around the orbit, i.e., where the electron wouldn't come around and be out of phase and interfere with its own wave. Again, similar to classical orbits (planets, rockets), the size of the orbit corresponds to its energy. Unlike classical orbits, electrons can only occupy discrete energy levels. Electrons are couch potatoes: they "prefer" to occupy the lowest possible energy level. Bump an electron up to a high energy level and it will promptly fall back down to a lower level. The amounts of energy that it takes to bump an electron to a higher energy level or that it emits when it drops back down are quantized; they correspond to the difference in energy between the initial and final energy levels.

Light, spectra

The vast majority of the information that we can obtain about anything and everything beyond the bounds of the Earth comes by way of the light that we receive. There are exceptions—high-energy particles, meteorites, gravitational waves—but what we learn in astronomy is dominated by electromagnetic radiation. So, let's talk about light.

Electromagnetic radiation can be described either as waves or as particles. As a wave, we describe a particular color of light by its wavelength λ (how far between wave crests) and its frequency f (or ν; how many wave crests pass a point each second). As a particle, we can think of light as tiny bundles of energy, called photons. The way to convert from the wave to the particle picture is by way of *Planck's constant, h*: $E = hf$, where $h = 6.63 \cdot 10^{-34}$ J·s $= 4.136 \cdot 10^{-15}$ eV·s. The electron volt is the energy gained or lost by an electron as it crosses a potential of 1 volt; 1 eV = $1.602 \cdot 10^{-19}$ J. If you reduce the Joule to its fundamental units, it's = 1 kg·m^2/s^2.

Wavelengths run from picometer-wavelength gamma rays out to 100+ km long-wave radio. There are no hard and fast rules for dividing the regions of the electromagnetic spectrum, but roughly:

	wavelength	frequency	energy
Gamma (γ)	1 – 10 picometers	30 – 300 Exa Hz	124 keV – 1.24 MeV
X-ray	10 pm – 10 nm	30 Peta Hz – 30 EHz	124 eV – 124 keV
Ultraviolet	10 – 400 nm	750 Tera Hz – 30 PHz	3 – 124 eV
Visible	400 – 700 nm	430 – 750 THz	1.8 – 3 eV
Infrared	700 nm – 1 mm	300 GHz – 430 THz	1.24 meV – 1.8 eV
microwave	1 mm – 1 m	300 MHz – 300 GHz	1.24 μeV – 1.24 meV
radio	1 m – 10^8 m	3 Hz – 300 MHz	12.4 femto eV – 1.24 μeV

Multiplied together, wavelength and frequency equal the speed of light: $\lambda f = c$. This familiar speed limit, 299,792 km/s, is the speed of light *in a vacuum*. A vacuum has an index of refraction, $n = 1$. Most real stuff, whether it's the glass lens in the eyepiece of a telescope or the almost but not quite emptiness of interstellar space, has an index of refraction that's > 1. That means light doesn't usually travel at "the speed of light" but at a speed that slower by a factor of $1/n$. In glass, for instance, where $n \sim 1.5$, light travels at $3.0 \cdot 10^5$ km/sec / 1.5 = $2.0 \cdot 10^5$ km/sec. Indices of refraction are wavelength-dependent, meaning that not all colors get bent the same amount or travel at the same speed.

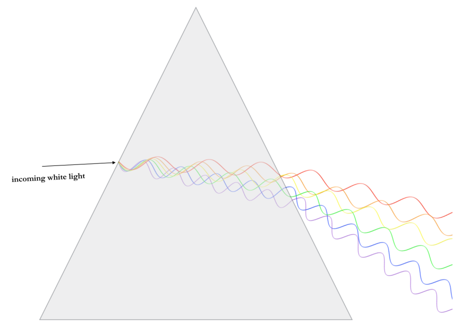


Figure 1.18

Different colors of light being refracted – bent – by different amounts as the light passes through a prism

Waves can interfere with each other. Waves of light coming through slits – or the circular aperture of a telescope – will be diffracted and produce an interference pattern. We will look at this later, when we discuss telescopes, but for now note that the best angular resolution that is theoretically possible with a telescope of diameter D is given by

$\theta''_{\min} = 2.5 \cdot 10^5 \lambda / D$, where the angle is in seconds of arc.

Some examples:

1. In the table above, the energies are given in eV; the usual SI (mks) unit of measurement for energy is the Joule. As noted above, 1 eV = $1.602 \cdot 10^{-19}$ J. A high-energy gamma ray photon might have an energy of 1 MeV; how many J is that?

   10^6 eV · ($1.602 \cdot 10^{-19}$ J / eV) = $1.602 \cdot 10^{-13}$ J

2. If we converted that 1 MeV photon into the rest mass of a particle, what would the particle's mass be? Here we want $E = m c^2$:

   $1.602 \cdot 10^{-13}$ J / ($3 \cdot 10^8$ m/s)^2 = $1.78 \cdot 10^{-30}$ kg.

   For comparison, the mass of an electron is 0.5 MeV or $9 \cdot 10^{-31}$ kg.

3. Check that those units work out; in other words, are kg = J / (m/s)^2?

   J / (m/s)^2 = (kg·m^2/s^2) / (m/s^2) = kg.

4. It takes 13.6 eV to ionize a hydrogen atom from the ground state (more below); what wavelength does that correspond to? $E = hf = hc / \lambda$.

   λ = ($4.135 \cdot 10^{-15}$ eV·s) · ($3 \cdot 10^8$ m/s) / 13.6 eV = $9.12 \cdot 10^{-8}$ m = 91.2 nm.

Some times it will help to consider the electric and magnetic fields themselves, the entities that are doing the waving. Time-varying electric fields $\vec{E}$ produce time-varying magnetic fields $\vec{B}$ (in a perpendicular direction); the electromagnetic wave – the light – propagates in the direction perpendicular to the orientation of the electric and magnetic fields, here shown as the S direction:

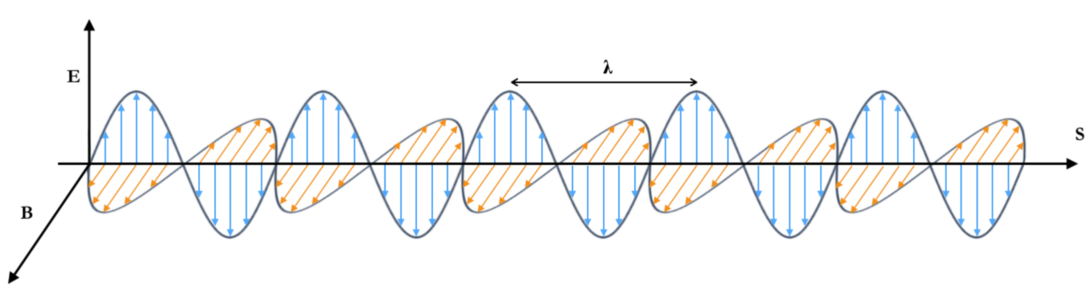


Figure 1.19: Electromagnetic waves

If, as shown in the above figure, the fields continue to oscillate in the same planes, the light is plane-polarized. If the planes vary, the light is elliptically polarized. If there are two distinct planes of oscillation at 90° to each other, the light is said to be circularly polarized. Polarized sunglasses are useful, but why do we care about polarization in astronomy? Polarization has to be induced by something, like, for instance, small elongated grains in the interstellar medium. Measuring the polarization can help tell us something about the space the light passed through on its way to us.

More material about optics, along with telescopes and detectors, may be found in Chapter 20.

Waves tell us about the relative radial motion of the source and us by way of the Doppler shift. A source that is moving towards us, say, emits light waves of a certain wavelength but because it is moving towards us while emitting that light, the waves get scrunched together, so that we see shorter wavelengths than we would have seen if there were no relative radial motion. The faster the motion, the more the shift in wavelength. Sideways motion won't have any effect—only the radial component of the motion matters.

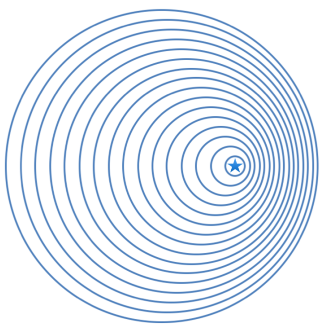


Figure 1.20
Here is a sketch showing a star moving toward the right at a constant velocity while emitting light, shown as circles expanding about the points from which the light was emitted (spherical bubbles, if we considered three dimensions). If the circles represent the crests of waves, we can see that the wavelength looks shorter to an observer on the right, longer to an observer on the left, and unchanged to an observer viewing the system tangentially.

The math looks like this: $\dfrac{\Delta\lambda}{\lambda_0} = \dfrac{\mathrm{v}_r}{c}$.

(Or, if the speed is relativistic, $\dfrac{\Delta\lambda}{\lambda_0} = \left[\dfrac{1+\mathrm{v}_r/c}{1-\mathrm{v}_r/c}\right]^{1/2} - 1$.)

Example: Above, in the examples of how to work problems using the *vis-viva* (orbital velocity) equation, we calculated the central mass for M31 assuming that the measured velocities of matter in orbit around that central mass were 240 km/sec. Suppose that we obtained those velocities by observations of the Ca K line, a spectral line caused by singly ionized calcium atoms. The rest wavelength for the Ca K line is 3933.6614 Å; the Ångstrom is a unit of length = 0.1 nm. Due to the Doppler shift, at what wavelength would we observed the Ca K line as emitted by atoms moving toward us at 240 km/s?

First, 240 km/s is not relativistic, so we can use: $\dfrac{\Delta\lambda}{\lambda_0} = \dfrac{\mathrm{v}_r}{c}$.

Second, the atoms are coming *toward* us, meaning we will observe a *shorter* (blueshifted) wavelength.

$\Delta\lambda = \lambda_0(\mathrm{v}_r/c) = 3933.6614 \cdot (240 \text{ km/s} / 3 \cdot 10^5 \text{ km/s}) = 3.15$ Å.

Subtract this from the rest wavelength:

3933.6614 – 3.15 = 3930.51 Å.

An alternative route by which to tackle this problem is to solve for the observed wavelength before starting to plug in numbers. Expand the Doppler equation as follows:

$$\frac{\Delta\lambda}{\lambda_0} = \frac{\lambda_{\text{observed}} - \lambda_0}{\lambda_0} = \frac{\lambda_{\text{observed}}}{\lambda_0} - 1 = \frac{\mathrm{v}_r}{c}.$$

Solve for the observed wavelength:

$$\lambda_{\text{observed}} = \lambda_0 \left(\frac{\text{v}_r}{c} + 1 \right) = 3933.6614 \cdot \left(\frac{-240}{3 \cdot 10^5} + 1 \right) = 3930.51 \text{ Å}.$$

Here we write the velocity as *minus* 240 km/s because the emitting atoms are coming *toward* us and we know our observed wavelength must be blueshifted.

If we want to work in frequency units rather than wavelength, we need to remember that longer wavelengths correspond to lower frequencies; in other words, the change due to motion will have the opposite sign. Let's convert from the Doppler formula from wavelength to frequency using the relationship between wavelength, frequency, and the speed of light:

$$\frac{\text{v}_r}{c} = \frac{\lambda_{\text{observed}} - \lambda_0}{\lambda_0} = \frac{{c}/{f_{\text{obs}}} - {c}/{f_0}}{{c}/{f_0}} = \frac{\frac{f_0 - f_{\text{obs}}}{f_{\text{obs}} f_0}}{\frac{1}{f_0}} = \frac{f_0 - f_{\text{obs}}}{f_{\text{obs}}} = \frac{f_0}{f_{\text{obs}}} - 1.$$

As noted above, light also behaves as a particle. One chunk of electromagnetic energy is called a photon or a quantum. Different colors of light correspond to different amounts of energy in our photon. Short wavelengths are higher, long wavelengths are lower energy. Again, the following relationship lets us convert between the two pictures: $E = hf = hc/\lambda$, where h is Planck's constant, = 6.623 10^{-34} J·s or = 4.134 10^{-15} eV·s. Units check: remember that Joules are the mks unit of energy. 1 J = 1 Nm = 1 kg m^2 / s^2. The eV is the electronvolt, the energy gained by a single electron as it is accelerated through a potential of one volt. 1 eV = 1.602 10^{-19} J. Masses of particles will often be expressed as their energy equivalents in eV, where we really mean eV/c^2 and often just drop the factor of c^2. The mass of the proton, for instance, could be expressed as 1.67 10^{-27} kg or as 938 MeV.

Because we have waves, particles, and fields, all describing the same phenomenon, this might be a good place to mention the Heisenberg Uncertainty Principle. We could describe a photon as having a position, x, and a momentum, p. It is a fundamental property of quantum systems that pairs of complementary properties ("conjugate variables") such as these cannot be known to arbitrary precision. One way to write this is $\Delta x \cdot \Delta p \geq {\hbar}/{2}$, where $\hbar$ is called the "reduced" Planck's constant, = h / (2π).

A wave is spread out; it might have a well-defined direction, i.e., p, in which case we can say very little about where it is, x. A particle can be thought of as a superposition of many waves, adding together at one point, x, where we say, aha, here is a particle; in that case, we have very little knowledge of where the particle is headed, p.

Look at the units here: momentum is mass times velocity, or kg m /s. Multiply that by position and we have kg m^2/s = J·s. In other words, we could also express the uncertainty principle in terms of the variables energy and time:

$$\Delta E \cdot \Delta t \geq {\hbar}/{2}.$$

We'll see this formulation, below, when we consider the mechanisms responsible for broadening spectral lines.

Detectors are not usually going to count the energy in single photons. More realistic will be a detector with a small surface area receiving a certain amount of energy per second. We call that the *flux*, i.e., J / (s· m^2). If you want to compare with thing such as light bulbs, 1 J/s = 1 Watt. Note that the flux is probably not the same at every color, so you will often see F_λ, to indicate the monochromatic flux, and F to indicate the flux integrated over all wavelengths.

Let's talk about how the light gets emitted. At its most basic, something that had energy is giving that energy up, converting it to electromagnetic radiation. For traditional visible-wavelength astronomy the most obvious thing to talk about are electrons and their atomic energy levels. We will also want, in a bit less detail, energies associated with molecules and energies of charged particles moving in magnetic fields.

Atomic spectra

As noted above, electrons in atoms are permitted to occupy only discrete quantized energy levels. Changes in an electron's energy level correspond to emissions or absorptions of energy. Let's look at a simplified energy level diagram for hydrogen:

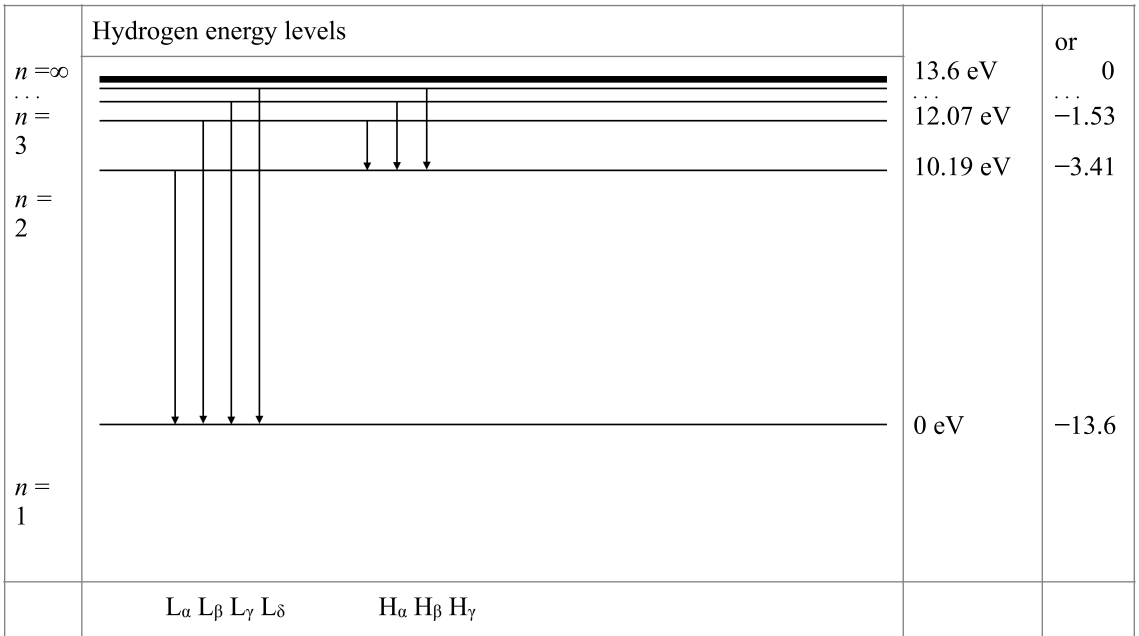


Figure 1.21 (the e⁻ is bound, so properly speaking we should use the column of negative energies)

The electron jumping up is absorption; down, shown in this figure, is emission. Jumps to or from the ground state are called the Lyman series, to or from the $n = 2$ level are the Balmer lines, $n = 3$ the Paschen series. These are in the ultraviolet, the visible, and the infrared, respectively. Give a ground state electron 13.6 eV and it becomes unbound, i.e., ionized. It will probably promptly recombine and drop back down, in one or more steps, to ground. Give a ground state electron 12.07 eV and it will jump to the $n = 3$ level; if it drops back down by way of the $n = 3$ to 2 transition it will emit H_α, a bright red line with $\lambda = 656$ nm. Note that we've shown the energies here as positive; since the electrons in their various energy levels are *bound*, it would make just as much sense to label the ground state as –13.6 eV and ionization as 0 eV.

Note that this continual electron pinball can shuffle the energy distribution of photons around a bit. Imagine a cloud of neutral hydrogen atoms being bathed with a flux of ultraviolet light, say from a hot blue star. The UV photons will ionize the hydrogens, the protons will recombine with the free electrons, and the electrons will cascade back down to the ground state, emitting the appropriate photons on the way down. Some of those will be 13.6 eV photons, i.e., the electron drops all the way back to ground in one step. But many of them will not be. The electron may drop back down by way of some combination of lower energy transitions. The only requirement is that the transition energies all add up to 13.6 eV by the time the electron is back down in the $n = 1$ level. In the process, the 13.6 eV UV photon that did the initial excitation may have gotten converted, e.g., to a couple of IR photons, a visible photon, and a lower energy UV photon.

Each element has its own unique pattern of energy levels, meaning that the emission spectrum from each element is different. Here's what an emission spectrum for the first four hydrogen Balmer transitions looks like:

Figure 1.22

And for another example, here's carbon:

https://imagine.gsfc.nasa.gov/science/toolbox/spectra2.html

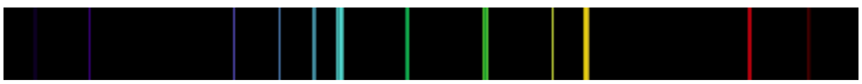


Figure 1.23

Energy level diagrams are more complicated than this example. Electrons have orbital angular momentum and spin angular momentum—it takes four quantum numbers, not just *n*, to adequately describe the state of an electron. Most transitions are permitted but some are not. Electrons can find themselves in states from which they can't legally de-excite. A level like that is called metastable, because an electron in such a state is likely to spend a relatively long time there. Transitions downward from metastable levels are called forbidden transitions, because they do occur, but much less rapidly than permitted transitions. An electron in a normal excited level may spend 10^{-8} s or less in that excited state before de-exciting; an electron in a metastable level may be there for 10^{8} s or more before it drops down. In most cases the easiest way out is up, e.g., a collision occurs, bumps the electron up to a higher energy level, and it drops back down by some other, permitted, pathway. If the electron does stay in the metastable level for a while it may get stimulated to de-excite by the passage of another photon of the same energy as the one the electron would emit. This process—called stimulated emission—is what happens in lasers and their microwave cousins, masers. The two photons have the same wavelength and they are coherent, meaning their oscillations are in phase.

Electrons affect each other, so the spectra of neutral and ionized atoms are different, since they have different numbers of electrons. E.g., singly ionized carbon will have a different pattern of lines than neutral carbon. The mass of the nucleus also matters, a bit, so isotopes will have unique spectra; for instance, the lines of hydrogen and deuterium will be slightly different.

Molecules and molecular spectroscopy

The individual atoms in a molecule can make electronic energy transitions such as those described above. Molecules, though, have more ways to store and release energy than individual atoms. Molecules can vibrate, as if the atoms were connected by springs, and they can rotate. The angular momenta of rotation and vibration are also quantized in units of $\hbar$. In other words, molecules have rotational and vibrational energy transitions. Most of these transitions are in the infrared. Vibrational energies tend to be larger by about an order of magnitude than rotational energies. A molecule is probably doing both at once, so for a particular vibrational energy state there are going to be several rotational levels. Energy transitions may involve a combination of electronic, vibrational, and rotational levels. This leads to a banded structure for molecular spectra.

Let's look at molecules in more detail. A molecule usually refers to a neutral group of two or more atoms held together by a chemical bond, often involving sharing one or more electrons. Sometimes noble gas atoms, not bonded to other atoms, are considered molecules. Sometimes electrically charged groups would be called ions just as charged atoms are. In astronomy the molecules we encounter usually only contain a only few atoms – H_2, CO, CH_4, etc. – although there a few interstellar hydrocarbon molecules containing 12-13 atoms, as well as detections of the fullerenes C_{60} and C_{70}.

Molecules have more available paths for absorbing and emitting energy than individual atoms. Atoms have electrons with quantized energy levels. Molecules, with more than one atom, also have quantized vibrational and rotational energies. Let's look at diatomic molecules.

For the vibrational energy levels, we can approximate a diatomic molecule as a simple harmonic oscillator, which isn't bad as long as the nuclei aren't too far apart. The potential energy for simple harmonic oscillation is given by

$V(x) = ½\, C\, x^2$,

where x is the distance from the center of mass and C is the force constant.
We also have a kinetic energy associated with each of the nuclei. Express the kinetic energies of each of the atoms in terms of their momenta:

$½\, mv^2 = ½\, p^2/m$.

Note that the momenta for the two nuclei are equal in magnitude and the fact that they are opposite in sign won't matter when we square them. In other words, the total kinetic energy is

$$K.E. = \frac{1}{2}\frac{p_1^2}{m_1} + \frac{1}{2}\frac{p_2^2}{m_2} = \frac{1}{2}p^2\left[\frac{1}{m_1}+\frac{1}{m_2}\right] = \frac{1}{2}p^2\left[\frac{m_1 m_2}{m_1+m_2}\right]^{-1} = \frac{1}{2}\frac{p^2}{\mu}.$$

The term μ is the reduced mass — if you've read the section above on celestial mechanics, you saw the reduced mass in terms of two gravitationally bound objects; same basic idea here. Thus we can write the total energy for the molecule as

$$E_{vibration} = \frac{p^2}{2\mu} + V(x) = \frac{p^2}{2\mu} + \frac{1}{2}Cx^2.$$

When we move from the classical mechanical treatment of a macroscopic oscillator to the quantum mechanical treatment of a vibrating molecule, we find that the allowed energies for vibration take the form

$$E_{vib} = (\mathrm{v} + 1/2)\cdot \hbar\sqrt{C/\mu},$$

where v is the quantum number associate with the vibrational energy levels. Note that the lowest possible energy is not zero. Vibrational energy transitions tend to be in the infrared.

Rotational energies are smaller than vibrational. We will approximate our molecule as a rigid rotator. Classically, we have

$$I = \sum_i m_i r_i^2 \text{ and } L = \sum_i m_i \omega r_i,$$

where I and L are the moment of inertia and angular momentum, respectively. For a diatomic molecule $I = \mu\, r^2$, where we are again using the reduced mass. The kinetic energy is given by

$$E = I\omega^2 = \frac{L^2}{2I}.$$

Quantum mechanics tells us that the angular momenta, and hence the energies, are quantized. Using J as the rotational quantum number (R has too many other meanings!),

$$E_J = J(J+1)\cdot\frac{\hbar^2}{2I}.$$

Rotational energy transitions are typically in the microwave.

Molecular spectra are made up of bands of closely spaced lines that arise from combinations of electronic, vibrational, and rotational energy transitions. These energy transitions involve coupling of an electromagnetic field (i.e., passing light) to the electric dipole moment of the molecule. Small linear highly symmetric molecules — e.g., N_2 — don't have permanent dipole moments. They may have temporarily induced electric dipoles, as well as electric quadrupole and magnetic dipole moments, but these are weaker and result in weaker transitions and harder-to-detect molecules. In other words, it's harder for small symmetric molecules to absorb energy from an electromagnetic field. Imagine an exoplanet moving in front of its parent star. Light from the star will pass through the planet's atmosphere and some wavelengths will be absorbed. It could be the case, though, that some of the dominant molecules in that planet's atmosphere will be very hard to detect.

Excitation energies

Where does the energy come from for the jump up? Either by hitting the atom or molecule with a photon of just the right energy (radiative excitation) or by hitting it with another particle in a collision that involves enough kinetic energy to excite the energy transition (collisional excitation). In other words, we need to talk about how energy is distributed among large collections of particles and photons.

Particles first. Whenever possible we hope to be able to treat a collection of particles as an ideal gas. This could be the atmosphere of a planet, the outer layers of a star, stars in a galaxy, even galaxies within a cluster of galaxies, although, granted, the atmosphere of a planet is more likely to come to mind when you think of an ideal gas than is the idea that a galaxy could be a "particle". In an ideal gas collisions are elastic. This will result, over time, in a statistically predictable distribution of kinetic energies among the particles, called the Maxwell-Boltzmann distribution. In terms of particle speeds, it looks like this:

$$g(\mathrm{v}) = 4\pi \mathrm{v}^2 \left(\frac{m}{2\pi kT} \right)^{3/2} e^{-m\mathrm{v}^2 / 2kT}.$$

This equation describes the number of particles of mass m at temperature T having speed v. The constant k = 1.38 10^{-23}J/K = 8.61 10^{-5}eV/K; it's called the Boltzmann constant, and as you can see from the units it allows you to convert temperature into energy units. Graphically, the function g(v) looks like this:

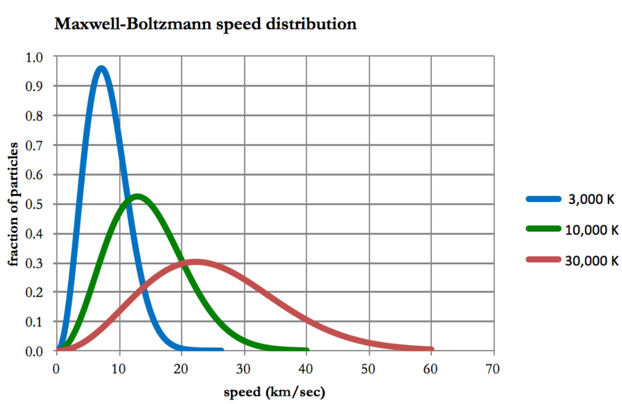


Figure 1.24

Particle speed distribution.

Plotted for m = hydrogen atom, and three temperatures, T = 3,000, 10,000, and 30,000 K.

The vertical axis scaling is arbitrary.

Let's look at some of the features of this plot.

- Where does the $4\pi \mathrm{v}^2$ come from? Because we are plotting *speed*, not velocity. Remember that velocity is a vector quantity; we assume that the velocities are randomly distributed around all angles, so when we integrate over all angles we get a factor of $4\pi \mathrm{v}^2$.
- Related question: what is the average *velocity* of the particles? 0 (because equal numbers of particles have positive and negative velocities).

- What does the area under the curve represent? g(v) gives us the number of particles having each speed, so if we add them all up, i.e., if we integrate g(v) over all v, we get the total number of particles.
- What is the most probable speed for a particle to have? That's a calculus question. To get this you take the derivative of g(v), set it equal to 0, and solve for the speed (i.e., solve for the speed at which the curve reaches a maximum, flattens out, and turns over). The result is $v_{\text{most probable}} = \sqrt{2kT/m}$.
- Note that the most probable speed is not the same as the average speed. Traditionally people are more interested in the average kinetic energy, which means we want the average v^2. This gives us what's called the root mean square speed: $v_{rms} = \sqrt{3kT/m}$.
- If you want the average speed, that's yet another value: $v_{ave} = \sqrt{\dfrac{8kT}{\pi m}}$.
- What happens if we increase the mass of the particles? There are two terms with mass in them. Notice that term in the exponential is the ratio of kinetic energy of the particle to the average thermal energy: $\frac{1}{2}mv^2 / kT$. And that it's negative. Should make sense what happens: higher mass particles are harder to get moving fast, so there will be relatively more particles with lower speeds. The v_{mp} will decrease.
- What happens if we increase the temperature? This corresponds to dumping in more energy to be shared among the kinetic energies of all the particles, so the peak of the curve shifts to higher speeds. Note that there will still be some particles moving slowly. Note also that the area under the curve didn't change. We didn't change the number of particles, just their speed distribution.

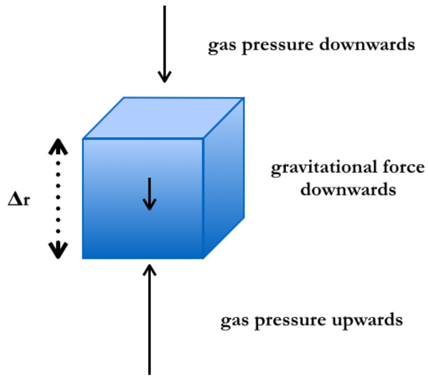
In addition to knowing the particles' kinetic energy, we also want to know how densely they are packed, how much pressure they exert, that sort of thing. You've probably seen the Ideal Gas Law before, although exactly the notation that's used may depend on whether you used it in a physics class or a chemistry class or somewhere else. Generically this is known as an equation of state—there are non-ideal gases and they have other expressions relating their pressures, temperatures, and densities. Here are several formulations that all say the same thing for an Ideal Gas, one where the colliding particles just bounce off each other elastically (like billiard balls):

$PV = NR_gT$	N is the number of moles, V is volume
$P = \rho TR_g/\mu$	μ is the average particle "weight", ρ is mass density
$P = nkT$	n is the particle number density
$P = \rho kT/\mu m_u$	m_u is the average mass of a nucleon

R_g is the universal gas constant, = 0.0821 liter – atm / mol – K = 8.314 J / mol – K , where 1 mole = 6.023 10^{23} particles.

While we're talking about particles and pressures we've been ignoring gravity. Any real system of any appreciable size, i.e., bigger than a box on a lab bench, is going to have a pressure gradient because there's a distinct direction ("down") because of gravity. The atmospheric pressure at the surface of the Earth is greater than it is 5 miles up. But the same physics tells us that stars will be concentrated toward the plane of our galaxy and decrease in number density as we move above or below the plane. And tells us that for a star to be in equilibrium the outward pressure must equal the inward force of gravity—otherwise our star will either collapse or explode.

Let's take the latter principle first. It's called hydrostatic equilibrium. Imagine a small cube inside a bigger box (star, planet, whatever) for which there's a distinct gravitational up and down:



A cube of mass *m* at some distance *r* from the center of mass of the planet (or star or whatever), subject to a gravitational acceleration *g*.

Figure 1.25

Consider the forces acting on the cube; first there is gravity:

$$F_{grav} = mg = -\rho(r)A\Delta r\left(GM/_{r^2}\right),$$

where we are writing the volume of our box as area of one face x height (i.e., $A\Delta r$).
Recall that pressure = force / unit area, so the force due to the gas pressure can be expressed as $P(r)A$. Specifically, we have

$$F_{\text{gas, upward}} = P(r)\cdot A \text{ and } F_{\text{gas, downward}} = -P(r+\Delta r)\cdot A.$$

Note that if we are in equilibrium, i.e., if our cube is not rising or falling, $P(r+\Delta r) < P(r)$, the total force equals zero and there's a net upward pressure that balances the gravitational force downward. In other words, expressing the pressure difference across the height of the cube as ΔP,

$$\Delta P\cdot A = -\rho(r)A\Delta r\left(GM/_{r^2}\right).$$

Divide both sides by A and Δr:

$$\Delta P/_{\Delta r} = -\rho\left(GM/_{r^2}\right) = -\rho g.$$

Calculus alert: in the limit as $\Delta r \to 0$, this goes to dP/dr.

The pressure clearly varies as a function of r. Go back and retrieve the ideal gas law and substitute in for the density in the hydrostatic equilibrium equation. This gives

$$dP/_{dr} = -\rho g = -P\left(m/_{kT}\right)g;$$

$$dP/_{P} = -\left(m/_{kT}\right)g\cdot dr.$$

A ratio of the form $dX / X = d \ln (X)$. When we integrate, we will get

$$P(r) = P(r_0)\cdot e^{-(gm/kT)h} = P(r_0)\cdot e^{-h/H},$$

where we've defined $H = kT/gm$ and assumed that a whole lot of things—m, T, g—don't vary with r. H is called the *scale height* and it is the distance over which the pressure (or density) falls by a factor of e. Whether you are talking about how the amount of oxygen drops as you climb a mountain or how the number of stars drops off as you move up from the plane of our galaxy, the same physics tells you that the decrease in density will be exponential.

The distribution for photons looks quite similar to the Maxwell-Boltzmann distribution, with a couple of subtle differences: first, photons are bosons and don't collide elastically; second, when you heat an object it is likely to emit *more* photons, so that the total number of particles in our "box" isn't a constant.

Here we won't plot number vs. speed but intensity vs. wavelength (or frequency). Intensity, which you haven't met before, comes in units of J / s·m²·ster; in other words, it's the flux per unit solid angle (ster stands for steradian; there are 4π ster in a sphere). Again we've got three different temperatures plotted, 3,000, 10,000, and 30,000 K:

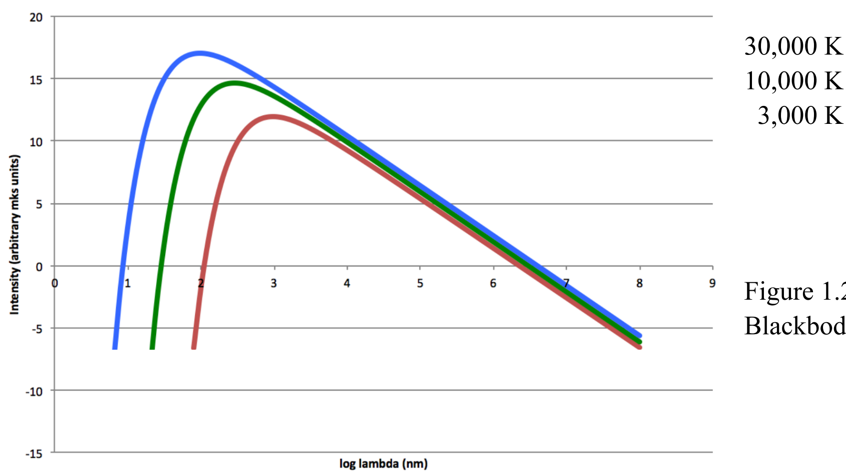


Figure 1.26: Blackbody spectra

This is called the Planck function; in wavelength units it is

$$I(\lambda)\Delta\lambda = \frac{2hc^2}{\lambda^5}\left[\frac{1}{e^{hc/\lambda kT}-1}\right]\Delta\lambda,$$

where the $\Delta\lambda$ indicates that we are never going to measure a monochromatic intensity but will always measure the energy in some wavelength interval $\Delta\lambda$. If we wanted the Planck function in frequency units, it looks like this:

$$I(f)\Delta f = \frac{2hf^3}{c^2}\left[\frac{1}{e^{hf/kT}-1}\right]\Delta f.$$

Calculus-physics note: $\Delta\lambda \neq \Delta f$.

The Planck function describes the emission of a blackbody, i.e., a perfect emitter / perfect absorber. Think asphalt pavement, which is close. It absorbs the sunlight falling on it, comes to an equilibrium temperature, and radiates energy according (well, nearly according) to the Planck function description for how much energy should be radiated at which wavelengths given a particular temperature. Objects which come closest to being blackbodies are those which are *opaque*. Opacity is a term that refers to how hard it is for light to get through a material. High opacity means that photons are going to bounce around a lot, get absorbed and reemitted, get scattered, generally get their distribution of energies shifted around a lot before being emitted. The surface of a planet, the insides of a star, the background radiation from the Big Bang—these are all pretty good blackbodies and can be described pretty well by the Planck function. Let's take a look at some features of this plot:

- What does the area under the curve represent? If we integrate the Planck function over all wavelengths we will get the total amount of energy being emitted by the blackbody per unit area. The result is called the Stefan-Boltzmann law and is written $F = \sigma T^4$, where we've taken all the various constants and piled them all into one: $\sigma = 5.67 \cdot 10^{-8} \frac{\mathrm{J}}{\mathrm{s} \cdot \mathrm{m}^2 \cdot \mathrm{K}^4}$. Note that this is still energy per second *per unit area*; if you want the *total* amount of energy that an object is emitting per second you have to multiply by the surface area to get the luminosity. If the object is spherical, that's easy: $L = A \cdot F = 4\pi R^2 \sigma T^4$ J/s.

- What is the most probably wavelength, i.e., what is the dominant color of our blackbody? Calculus alert again: take the derivative of the Planck function, set it equal to zero, and solve (messily) for λ_{max}. The result is Wien's law:

$$\lambda_{\text{max}} = \frac{2.898 \cdot 10^{-3} \text{ m} \cdot \text{K}}{T \text{ K}}$$

which gives the peak wavelength (or the wavelength at which the intensity is a peak) in meters.

- What happens when we increase the temperature of our blackbody? Two things: our object becomes brighter *at all wavelengths* and the peak wavelength shifts blueward (i.e., to higher energy). The area under the curve changes!!

Examples

1. What is the ratio of the flux emitted by a 20,000 K surface temperature star to that emitted by a 5,000 K surface temperature star?

$$\frac{F_{20,000}}{F_{5,000}} = \frac{(\sigma T_1^4)}{(\sigma T_2^4)} = \left(\frac{20,000}{5,000}\right)^4 = 4^4 = 256.$$

2. What is the wavelength of the peak of the Planck spectrum for a 10,000 K star?

$$\lambda_{\text{max}} = \frac{2.898 \cdot 10^{-3} \text{ m} \cdot \text{K}}{T \text{ K}} = \frac{2.898 \cdot 10^{-3} \text{m} \cdot \text{K}}{10,000 \text{ K}} = 2.898 \cdot 10^{-7} \text{ meters or } 289.8 \text{ nm.}$$

You might be starting to wonder by now what we mean by "temperature". Temperature can be a slippery beast. Based on the spectrum of the sunlight we can say that the temperature of the solar photosphere (the layer from which most of the light is emitted) is ~5800 K. Stand outside on a sunny day and 5800 K sunlight is falling on you. The temperature of the air around you is, depending on the time of day, year, etc., etc., maybe 295 K. The color temperature of the sunlight is clearly not the same type of temperature as the kinetic temperature of the air molecules.

We slid from talking about individual atoms absorbing or emitting to talking about opaque ensembles of lots of atoms emitting blackbody radiation. Often we're going to have some combination; e.g., more-or-less blackbody radiation emitted from the surface of a star or planet is going to interact with a non-opaque bunch of atoms on its way to us. What's the resulting spectrum going to look like? A collection of statements called Kirchhoff's rules say: 1) opaque objects emit blackbody spectra; 2) cool gases in front of opaque objects produce absorption spectra; and 3) hot gases produce emission spectra. A diagram:

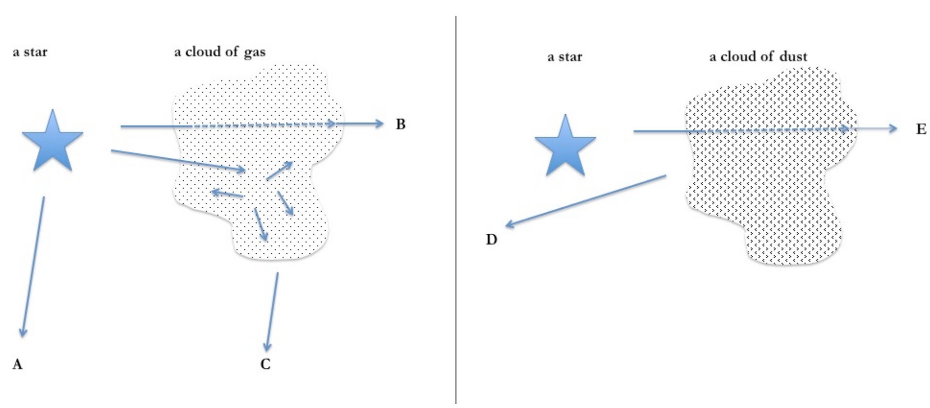


Figure 1.27 Geometry for understanding Kirchhoff's rules

What do you see

- from position A? The more-or-less blackbody spectrum of the star (the opaque object) alone.
- from position B? Let's suppose that our opaque object is "hot", by which we mean that the light it emits has enough energy to excite the atoms in the cloud. We will see dark absorption lines where specific wavelengths have been removed from the underlying spectrum by the process of bumping the electrons in the atoms in the cloud up to higher energy levels. Which wavelengths are missing depends on the composition of the cloud (and on its excitation level; remember that if we have lots of energy we could have ions, which have different energy level diagrams than neutrals).
- from position C? All those electrons that get bumped up to high energy levels are going to de-excite, drop back down, and emit the specific wavelengths that correspond to the energies of the electronic transitions. The emissions occur in random directions. From C we will see an emission spectrum, since some of that light is coming toward us. Against the cold background of empty space, our cloud is relatively "hot". What bright lines we see depends on the composition and temperature of the cloud.

There's another possibility, as long as we're on clouds of stuff, shown in the right-hand panel. This time we have a cloud of dust – not a rock, so it's not totally opaque but still having distinct solid grains in it. Dust preferentially transmits longer wavelengths and blocks or scatters visible and UV. (Very high energies, e.g., X-rays and Gamma rays, are also likely to get through.) This is called Rayleigh scattering and its efficiency is proportional to $1/\lambda^4$.

- What you see from D will be reflection of the bluer wavelengths by the dust.
- What you see from E will be a dimmer, reddened version of the spectrum from our opaque object.

Think about the sky: particles in the air scatter blue light out of the path from the Sun, making the sky blue; when there are *lots* of particles in the air, the setting Sun will look extra red, as even the green and yellow light will have a tough time making it through the atmosphere.

Blackbody radiation is often called thermal radiation because of the temperature dependence in the Planck function. That can be a bit confusing, though, when you recall that temperature takes many guises. We can talk, for instance, about the excitation temperature of the hot gas that's producing an emission spectrum. And kinetic energy, when paired with magnetic fields, can produce prodigious amounts of energy. The latter, cyclotron or synchrotron radiation, is usually called non-thermal to distinguish it from blackbody.

Suppose you have a source of free electrons (any charged particles will do, but electrons, being light, are handy) and a moderate to strong magnetic field. Further suppose that some component of the electrons' motion is perpendicular to the direction of the magnetic field. Recall from physics that the magnetic field will exert a force on the charged particle – $\vec{F} = q \cdot \vec{\mathrm{v}} \times \vec{\mathrm{B}}$ – that will cause the direction of its motion to curve. Assuming that our electrons also have some component of their velocity that is parallel to the direction of the magnetic field, the resulting motion for the electrons will be helical. In other words, the electrons are being accelerated, and accelerated charges radiate. If the electrons are non-relativistic, they radiate at a frequency that's equal to the frequency with which they spiral around the field lines. That's cyclotron radiation. If they are relativistic there are, as you might expect, consequences: the frequencies are messier and the radiation is tightly beamed in the direction of motion. This is called synchrotron radiation if the motion is dominated by circling around the field lines and curvature radiation if the motion is predominately along the field lines. The radiation may be called gyro-synchrotron if the electrons are mildly relativistic. All these types of radiation are polarized. We're often sloppy and simply call it all synchrotron radiation. Synchrotron emission shows up often in astronomy—Jupiter, neutron stars, active galaxies; lots of objects have magnetic fields and charged particles. In general the spectrum of synchrotron emission follows a power law:

$$F_f = F_0 \cdot f^{-\alpha},$$

meaning that if we plot log flux vs. log frequency we'll get a straight line with slope $-\alpha$; α is called the spectral index and tends to be about 1. A wrinkle: Electrons can interact with electromagnetic radiation. If we have a lot of electrons producing radiation at a lot of frequencies, the electrons will interact with each other's emissions, particularly at low frequencies. This means that our power-law spectrum tends to drop off at low frequencies (steeper slope); this is called synchrotron self-absorption.

Just to be thorough, note that the energy transfer in electron-radiation interactions can go the other way: in what's called inverse Compton scattering, relativistic electrons give up energy to passing photons, boosting the photon energies. And, one more: a free electron whizzing past an ion will have its path bent by the electric field of the ion. And yes, that counts as an accelerated charge, so we get radiation; this process is called bremsstrahlung, or "braking radiation", because it's likely that the electron gives up some of its kinetic energy in the process.

Sometimes the particles themselves arrive at Earth: cosmic rays are high-energy particles (on the order of 10^9 eV, although some are many orders of magnitude more energetic), usually charged, including some heavy nuclei (e.g., iron) as well as electrons and protons. High-energy gamma rays are often classified as cosmic rays, as well. A steady stream of protons, electrons, and a few helium nuclei from the Sun—the solar wind—continually arrives at Earth; solar wind particle are sometimes consider cosmic rays. Other particles arrive from galactic, and even extragalactic, sources. The low-level galactic magnetic field —a few microgauss—often bends cosmic ray paths with such a huge radius of curvature that it's nearly impossible to figure out what the original source of the charged particles was.

Neutrinos and gamma rays, on the other hand, won't be deflected and thus if we detect them in the cosmic ray flux and if we are able to determine their direction, they will point back to their source. In September 2017 just such an event was detected: a high-energy neutrino (~ 290 TeV) was recorded by the IceCube Neutrino Observatory at the South Pole, accompanied by near-simultaneous gamma rays detected by several telescopes, including the Fermi Gamma-ray Space Telescope and by an atmospheric Čerenkov radiation detector (MAGIC) in the Canary Islands, all originating from the same point in space. Čerenkov radiation usually occurs when an electrically charged particle travels faster than the local speed of light; it's like the optical equivalent of a sonic boom. Gamma ray photons are energetic enough that when they strike the atmosphere they can produce pairs of charged particles which are themselves moving fast enough to produce a flash of Čerenkov radiation. Neutrinos interact weakly with atoms in water or ice, albeit with low probability!, producing charged leptons that can also sometimes be energetic enough to produce Čerenkov radiation. The source of these particular cosmic rays appears to be an active galaxy denoted TXS 0506+056. This galaxy is an example of a type called a blazar or a BL Lac object, after the first such object identified. The nuclei of active galaxies often have jets, which act as natural particle accelerators; in the case of blazars, one jet is pointed straight at us. (For more on active galactic nuclei see chapter 18.) The light travel time from this galaxy was approximately 3.8 billion years.

We know that some of the cosmic rays, i.e. radioactive isotopes with modest half lives, must originate relatively nearby. For instance, ^{60}Fe is produced in core-collapse supernova explosions and has a half life of 2.6 million years; detections of a handful of ^{60}Fe cosmic ray nuclei in recent years by NASA's Advance Composition Explorer (ACE) satellite suggest they were produced by a supernova explosion within a kiloparsec or so and within that last few million years.

Physics note: there's an effective upper limit, called the GZK limit after the three folks who first calculated it, to the energy an intergalactic cosmic ray could have. The reason is that when cosmic rays energies get up near $5 \cdot 10^{19}$ eV they will interact relativistically with photons in the cosmic microwave background (CMB) and lose their energies into the production of particles such as electron-positron pairs and low-mass mesons called pions. This limits the energies for cosmic rays arriving from distances larger than ~50 Mpc (~160 Mly).

Nucleosynthesis

Nuclear processes show up at several points in astronomy. Hydrogen and helium, and small amounts of a few other light elements, are produced in Big Bang nucleosynthesis, during the first few minutes of the universe. Other atoms – all the oxygen, carbon, iron, everything else – get produced by stars. Some of the elements produced are radioactive and decay with known half lives. Making use of that radioactive decay, in what's called radioisotopic dating, can give us ages for meteorites, lunar rocks, things that tell us how old the solar system is. Let's look at some of these nuclear processes.

Building light elements up into heavier elements – fusion – produces energy. Breaking heavier elements apart – fission – produces energy. The break point for energy production is at iron and nickel. Why? ^{56}Fe is the peak of the nuclear binding energy curve, meaning has the lowest mass/nucleon; nearby ^{62}Ni has the highest binding energy/nucleon. The mass of one helium nucleus is less than the combined mass of 4 protons. That mass difference gets converted into energy ($E = mc^2$) during hydrogen fusion. If you want to fuse past iron, you have to put energy in. Here's what the plot looks like:

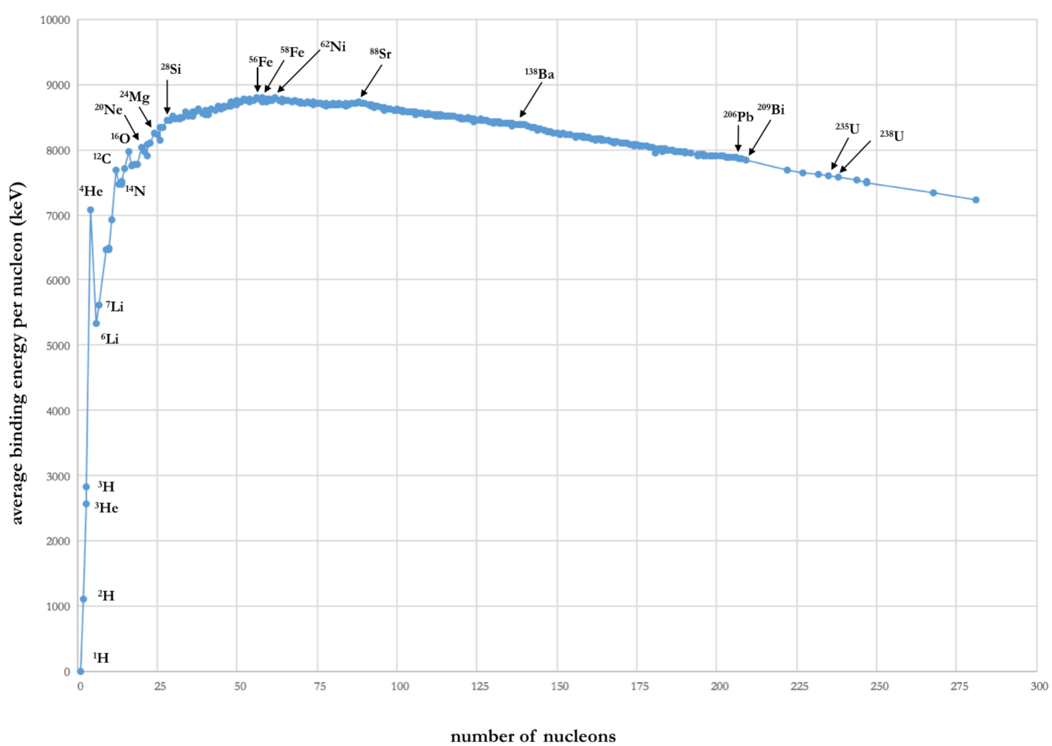


Figure 1.28: Nuclear binding energies for common and / or interesting isotopes.
Data from National Nuclear Data Center, Brookhaven National Laboratory http://www.nndc.bnl.gov/nudat2/index.jsp

Fusion requires getting two nuclei close enough that the short-range strong force can bind them together before their mutual electromagnetic Coulomb repulsion forces them apart. The pressures in the centers of stars are high and the particle kinetic energies are large; still, protons rarely slam together hard enough to fuse. Fusion requires some assistance from the quantum mechanical uncertainty principle and a process called tunneling. If you throw a baseball at a solid barrier you expect it to bounce back; you don't expect to find that the baseball somehow "magically" appears on the other side of the barrier. If you "throw" a proton at the barrier representing the Coulomb repulsion between your proton and another proton, tunneling says that there is a non-zero probability that the protons will both wind up on the same side of the barrier. Protons can be thought of as waves; when it hits the Coulomb barrier the incoming wave decays, but not instantly. Without tunneling the Sun wouldn't shine. Here is a plot of what's going on, in particular for the first step in hydrogen fusion. We have one proton at the origin; we bring in another from the right, with some energy that places it above 0 MeV but not high enough to get over the top of the Coulomb barrier. Classically, the incoming proton experiences the Coulomb repulsion and is deflected; quantum mechanically, some of the time, the incoming proton finds itself within the potential well, within a femtometer or so of the target proton, close enough for the nuclear forces to take over.

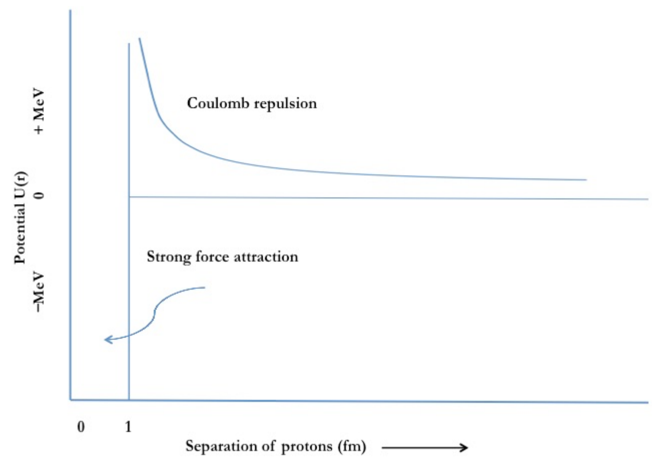


Figure 1.29:
Potential for nucleosynthesis

Let's talk about hydrogen fusion in stars. Assuming we can get our protons close enough for the nuclear forces to take over, here are the basic steps in what's called the proton-proton chain:

${}^{1}H + {}^{1}H \rightarrow {}^{2}H + e^{+} + \nu_e$	In a weak interaction, one proton becomes a neutron, with the release of a positron to balance the electric charge and a neutrino to balance the numbers of leptons. The e^{+} annihilates with one of the original hydrogens' e^{-}. The strong force glues to p^{+} and n together into a deuterium nucleus. Because this step involves the weak force, which is, well, weak, this is the rate-determining step.
${}^{1}H + {}^{2}H \rightarrow {}^{3}He + \gamma$	Mass is converted into energy.
${}^{3}He + {}^{3}He \rightarrow {}^{4}He + 2{}^{1}H$	Net result: 4 ${}^{1}H$ turned into 1 He + energy.

There are a couple more possibilities here, involving adding on to a ${}^{4}He$ nucleus, which we'll talk about in the section on stars and their evolution.

Helium fusion is another process, called the triple-α process (because helium nuclei were traditionally called α particles).

$^{4}He + {}^{4}He \leftrightarrow {}^{8}Be$	This process is almost reversible, because the ^{8}Be nucleus is unstable and decays with a lifetime of about 10^{-16} s. Look at the binding energy curve—the mass of ^{8}Be is *more* than the mass of two helium nuclei, not less.
$^{8}Be + {}^{4}He \rightarrow {}^{12}C + \gamma$	If another α particle hits the beryllium nucleus fast enough it's possible that the two will fuse. This time, mass is converted into energy.

The triple-α process takes higher temperatures than the proton-proton chain because there are more positive charges involved and a stronger repulsion to overcome. We'll see that this process takes place in the cores of Red Giant stars.

Just to confuse matters, there is another means of doing hydrogen fusion, one which involves carbon. That's assuming you *have* some carbon available—this process didn't take place in the very first generation of stars. It's called the CNO cycle:

$^{1}H + {}^{12}C \rightarrow {}^{13}N + \gamma$

$^{13}N \rightarrow {}^{13}C + e^{+} + \nu_e$

$^{1}H + {}^{13}C \rightarrow {}^{14}N + \gamma$

$^{1}H + {}^{14}N \rightarrow {}^{15}O + \gamma$

$^{15}O \rightarrow {}^{15}N + e^{+} + \nu_e$

$^{1}H + {}^{15}N \rightarrow {}^{12}C + {}^{4}He$ Note that the carbon gets spit back out at the end.

This process is more temperature sensitive than the proton-proton chain and tends to occur in the cores of hotter (more massive) stars. And as with the proton-proton chain, some of the reactions will go farther, i.e., the ^{15}N won't split, but will build up to ^{16}O, and so on. To reiterate: the CNO cycle is *hydrogen* fusion. The triple-α process, which produces carbon, is *helium* fusion.

Obviously there have to be some other nucleosynthesis processes because there are other nuclei—iron, uranium, whatever—that have to come from somewhere. Let's mention a few that we'll meet again when we discuss the advanced stages of stellar evolution. One option is to take the nuclei produced by the processes above and add them together. I.e., it's possible to fuse $^{12}C + {}^{12}C$ or $^{16}O + {}^{16}O$ or to add on more α particles. This gets us nuclei such as ^{20}Ne, ^{24}Mg, ^{28}Si. Some processes are going to spit out free neutrons; e.g., $^{12}C + {}^{12}C$ could produce $^{23}Mg + n$. Once we get some heavy nuclei, some of their radioactive decays will also produce neutrons.

The principal means of building up heavy nuclei is through neutron capture. Which isotopes get produced depends on how many neutrons we've got and whether the elements that are produced are stable and hang around long enough to capture more neutrons before they beta-decay into another element. β^{-} decay is the weak interaction where a neutron turns into a proton: $n \rightarrow p^{+} + e^{-} + \bar{\nu}_e$. Electrons are traditionally called beta particles. Later we'll meet the *r*- and *s*-processes, which refer to rapid and slow neutron capture with respect to the beta decay time.

Stars don't normally have free neutrons running around. Free neutrons are slightly heavier than protons and, if left to their own devices, will beta decay into protons in about 15 minutes. Protons and neutrons are both produced in copious amounts during the Big Bang, though, and when the universe was only a few minutes old there were still plenty of neutrons around to participate in nucleosynthesis. And it was hot enough. Recall that the first step in the proton-proton chain involves both weak and strong interactions. It'd be lots easier if it only had to involve the strong force. I.e., getting a proton and a neutron to fuse directly into deuterium is easier than getting two protons to fuse. It *was* hot enough for nucleosynthesis, for the first few minutes, but the early universe was expanding, the pressure and temperature were dropping; some protons and neutrons fused into helium, but remember that the triple α process takes *higher* temperatures. And without that we're stuck, because it turns out that there are no stable nuclei with masses of 5 or 8. The result is that Big Bang nucleosynthesis produces a universe that's about ¼ ^{4}He, by mass, a little bit of ^{7}Li, ^{3}He, ^{2}H, and the rest is ^{1}H. Everything else has to wait for stellar nucleosynthesis.

Stars form, eventually; high-mass elements get produced, and, in a profound act of cosmic recycling, aging stars eject those high-mass elements out into the interstellar medium where they seed the material from which the next generation of stars forms. We are star stuff.

Sample problems

1. The Trans-Neptunian object Eris has these orbital properties: a = 67.78 AU, e = 0.44.
   a) Calculate Eris' orbit period
   b) Calculate Eris' perihelion distance
   c) Calculate how fast Eris will be moving in AU/yr when it is at perihelion

2. An electron has a mass = $9.11 \cdot 10^{-31}$ kg; the positron is the same. How much energy would it take to produce an electron - positron pair? Express your answer both in Joules and in MeV.

3. Hα, a prominent red spectral line due to hydrogen, has a wavelength = 656.3 nm.
   a) What frequency is Hα?
   b) What is the corresponding energy in eV?

Advice: you could practice by working the problems backwards as well; i.e., once you have the answer, could you work back to the wavelength?

4. The pair of stars called 61 Cygni has a radial velocity relative to us of ~ −65 km/sec. There's a spectral line of Na (called "D_1) with a rest wavelength of 589.592 nm. At what wavelength would we observe this line in the spectrum of 61 Cygni?

5. The New Horizons spacecraft fly-by of the Kuiper Belt object 2014 MU_{69} is at a distance of 43.4 AU from the Sun. How much less sunlight, per square meter, is there at that distance than at Earth's average distance from the Sun?

6. The star Sirius is approximately 25.4 times more luminous than the Sun. Its distance is 8.60 light years. How much energy do we receive from Sirius in $J / (s \cdot m^2)$?

7. The temperature of the visible surface of the Sun (the photosphere) is ~5,772 K. Calculate the peak wavelength for the solar spectrum; express your answer in nm.

8. Calculate the most probable speed for a hydrogen atom in the Sun's photosphere.

9. Estimate the scale height for the solar photosphere. Hint: two-step problem; you could look up the surface gravity of the Sun but it would be better to calculate it yourself first and then check your answer.

10. Calculus problem: show that $\Delta\lambda \neq \Delta f$.

11. The mass of ^{1}H = 1.007825 u; ^{4}He = 4.002602 u; u, the unified atomic mass unit, = $1.66 \cdot 10^{-27}$ kg.

a) How much energy will be released when four H fuse to form one He atom?

b) With a luminosity of $3.282 \cdot 10^{26}$ J/s, how many kg of H is the Sun converting to He every second? Hint: this is *not* the mass equivalent to the solar luminosity but the mass of hydrogen used to produce that luminosity.

12. Mars is more massive than the Moon but it's also larger; would you weigh more on Mars or on the Moon? Don't just look up their gravitational accelerations; calculate them from masses and radii.

13. Proteus orbits Neptune in 1.122 days at an average distance of 117,647 km. Use this information to calculate the mass of Neptune.

14. Mars' moon Phobos isn't liquid, but does have a fairly low density, 1.876 g/cm^3, and likely not a lot of structural integrity. The density and radius of Mars are 3.934 g/cm^3 and 3,389.5 km, respectively. Phobos' average distance from Mars is 9376 km. Is that inside the Roche limit?

15. Eratosthenes, in his famous measurement of the Sun on the day of the summer solstice, found the Sun to be 7° south of the zenith at noon, or 83° above the southern horizon. Hopefully he wasn't looking at the Sun to make this measurement but rather at the shadow cast on the ground by a pole. If the pole he was using was 3 m high (and perpendicular to the ground), how long was the shadow it cast at noon?

16. Reading carefully?

a) describe the virial theorem

b) explain why Earth has two high tides each day

c) explain the difference between ^{3}H and ^{3}He

d) circular speed and escape speed differ by a factor of $\sqrt{2}$; explain why, with reference to the *vis viva* equation

e) what's the fusion product that results from the CNO cycle?

f) explain Kirchhoff's rules for spectra; i.e., explain under what conditions we expect to see absorption, emission, or blackbody spectra

g) describe synchrotron radiation

Answers to selected problems are on the next page:

1. 558 years; 38 AU; 1.2 AU/yr
2. $1.6 \cdot 10^{-13}$ J = 1 MeV
4. 589.46 nm
6. ~$1.2 \cdot 10^{-7}$ J / (s · m^2)

7. If you want to practice "guesstimating" this:

$$\lambda_{\text{peak of spectrum}} = \frac{2.898 \cdot 10^{-3}}{T}\text{, which for the Sun is } \approx \frac{3 \cdot 10^{-3}}{6000}$$

which is $0.5 \cdot 10^{-6}$ m or 500 nm.

8 ~9.8 km/s
9. 99-ish km
11a. $4.3 \cdot 10^{-12}$ J / reaction
11b. ~$5 \cdot 10^{11}$ kg/sec
14. If Phobos were liquid, yes; for rigid objects the Roche limit is a bit closer to the planet (instead of 2.44, use 1.26 in the Roche limit equation) and so for the moment Phobos is safe.
15. 37 cm