Chapter 2: **Solar system dynamics: phases and eclipses; celestial mechanics**

- introductory terminology on object orientations
- phases of the Moon
- lunar and solar eclipses
- a few more thoughts about related events
- review gravity and orbits
- the three-body problem, tides, rings, orbital resonances
- dynamical effects of solar wind and radiation on molecules and small particles
- moment of inertia and real gravitational fields.
- sample problems

This chapter contains additional material more specifically relevant to the study of the solar system, beyond that covered in Chapter 1 (Introduction) on gravity and orbits.

Introductory terminology on object orientations

Eclipses occur when one celestial object blocks the light from another. They are related to *transits*, in which a smaller object moves across the disk of a larger object; examples of transits include moons of Jupiter moving in front of the disk of the planet, or the inner planets moving across the disk of the Sun, or similar, if fainter, transits of exoplanets crossing the disks of their respective home stars. During *occultations*, an object of larger angular size crosses in front of a smaller one; examples here include planets or asteroids temporarily blocking the light from more distant stars.

Earth is unique among planets in our solar system in that the angular size of the Sun and Moon on the sky are very similar, leading to the regular occurrence of *solar eclipses*, in which the disk of the New Moon blocks all or a large part of the disk of the Sun, at least as seen from some location on Earth. During a *lunar eclipse*, the Full Moon moves through the Earth's shadow.

Phases of the Moon

It takes the Moon 27.32 days to complete one revolution around the Earth with respect to the background stars. That's the same as the Moon's sidereal rotation period, which is why the Moon pretty much keeps the same face towards us all the time. It's not exactly the same face: Because the Moon's orbit is an ellipse, it doesn't move at a constant orbital speed, meaning that sometimes the Moon gets a little ahead or a little behind the location you'd expect based on its average orbital speed. But the rotation rate doesn't vary, which leads to an effect, called libration, in which the Moon appears to be rocking side to side. Over the course of a month we wind up able to see about 59% of the Moon's surface from the Earth. As it moves closer or farther from us, the Moon will also vary slightly in apparent size on our sky; on average it's about half a degree across.

Celestial mechanics note: The Moon doesn't orbit the center of the Earth, it orbits the center of mass (or barycenter) of the Earth-Moon system, a point about 4,670 km from the center of the Earth. Recall that the center of mass is defined by

$$m_1 r_1 = m_2 r_2 .$$

For Earth and Moon, this becomes

$$7.342 \cdot 10^{22}\,\text{kg} \cdot r_{Moon} = 5.972 \cdot 10^{24}\,\text{kg} \cdot r_{Earth}$$
$$r_{Moon} + r_{Earth} = 384{,}400\ \text{km}.$$

The second equation expresses the fact that the Moon's orbital semi-major axis is ~384,400 km, meaning that on average it's ~384,400 km from the barycenter. Solving this pair of equations for r_{Earth} gives us 4,670 km, or about 73% of the way from the center to the surface of the Earth. The Moon's orbit is ~5%

eccentric, so the apogee and perigee distances should be about 20,000 km more and less than that average. But the Moon and Earth are also tugged on enough by the Sun that the Moon's perigee and apogee distances vary quite a lot: the apogee distance varies between 404,000 – 406,700 km and the perigee varies even more, from 356,400 – 370,400 km. In other words, over many years the Moon's distance from us varies by ~50,000 km, not 20,000 km. A curious side effect of the Sun's influence is that the Moon's path is always concave toward the Sun.

Roughly speaking, though, the Moon goes around us. Its orbit is prograde, which means that from day to day the Moon moves toward the east among the background stars. It also means that the Moon rises later from day to day, by about 50 minutes. The Moon goes through phases as observed from the Earth, as the portion of the illuminated half that we can see waxes, reaches Full, and wanes. The synodic month, the time it takes for the Moon to return to a given phases (e.g., Full to Full) is about 29.5 days.

Math note: Why is the synodic month 29.5 days? We are beating together two sidereal frequencies to obtain the synodic period. The relevant equation is

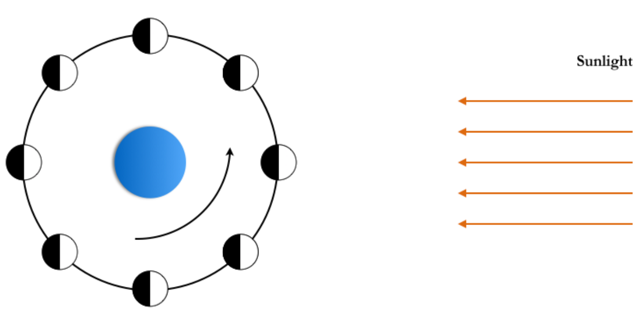
$$\left[\frac{1}{P_1}-\frac{1}{P_2}\right]=\frac{1}{S},$$

where S is the synodic period and the sidereal periods, in this case, are the year and the month. In other words,

$$\left[\frac{1}{27.3^d}-\frac{1}{365.25^d}\right]=\frac{1}{29.5^d}.$$

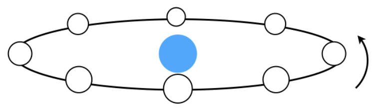
The New Moon is roughly in the same direction in the sky as the Sun; it rises and sets with the Sun and on that day the illuminated portion of the Moon is entirely facing away from us. Over the next few days, as the Moon gets farther to the east of the Sun in the sky, we see a waxing crescent. About a week after New Moon the Moon will be about ¼ of the way around the sky from the Sun and we'll see a First Quarter Moon. At First Quarter the Moon rises about noon and reaches its highest point in the sky in the south (transits the celestial meridian) at about 6 p.m. At First Quarter the side of the Moon facing the west is the side we see illuminated and the terminator, the line dividing the dark and lit sides, is straight. Following First Quarter the Moon waxes gibbous for about a week until it reaches Full Moon at the point where it is pretty much opposite the Sun in the sky. The Full Moon rises about sunset and is up all night. After Full, the Moon is waning gibbous, Third Quarter, waning crescent, and back to New again.

Let's look at this graphically. Here is a sketch of the Moon in several positions in its orbit around the Earth, looking down from the north:



Sunlit side of the Moon
Figure 2.1

So far, this isn't complicated – the sides of Earth and Moon facing the Sun are in daylight and the sides facing away are in darkness. There's a problem, though: it looks as though the Moon should be crossing in front of the Sun at New Moon and moving through the Earth's shadow at Full Moon. The reason we *don't* have eclipses every month is that Moon's orbit around the Earth is tilted (inclined) approximately 5 degrees to the ecliptic (the plane of our orbit around the Sun). The Moon's orbit crosses the ecliptic in two points called *nodes*. The word ecliptic comes from the fact that when the Moon is crossing the ecliptic, and is in the right phase, we do have eclipses; more on that below. Every six months there are eclipses, but if you were looking at the Earth and Moon from the Sun in one of the non-eclipse months it might look more as in the following sketch, where you could see both Earth and Moon all month long:



Moon and Earth as seen from the Sun Figure 2.2

The question, though, is what someone on Earth is going to see when looking at the Moon while the Moon is in one of these positions. Return to the view of the Moon's orbit from the north and add a person *A* on Earth looking at the Moon as indicated by the arrow in the following sketch, i.e., with the Moon at just one location:

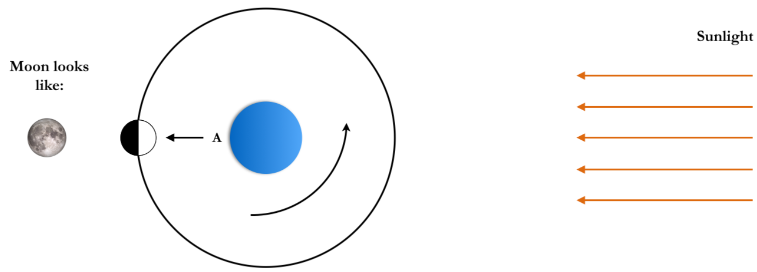


Figure 2.3: Full Moon

For *A* it must be midnight, since this point is on the opposite side of the Earth from the location the sunlight is hitting. The Moon is relatively high in the sky for *A*, and the Earth-facing side of the Moon is fully illuminated. In other words, this is a Full Moon. Looking down from above we can see the dark half of the Moon as well, but *A* simply sees a fully lit Moon.

A week later the Moon has moved a quarter of the way around its orbit. In the following sketch, let's add points to indicate *A*'s position at different times of day:

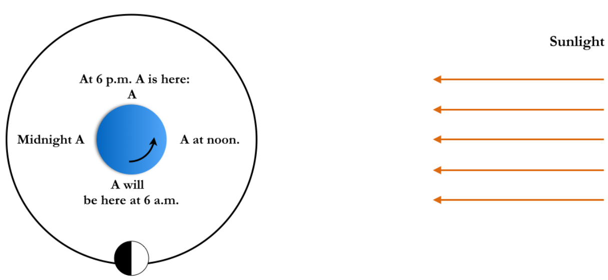


Figure 2.4: 3rd Quarter Moon

With the Moon at this location in its orbit, *A* couldn't see the Moon at 6 p.m.; at midnight, the Earth's rotation will have carried *A* around so that the Moon will be rising. This Moon will be highest in the sky for *A* at 6 a.m. and it will set about noon. This is a 3rd Quarter Moon. What does it look like from the ground?

Imagine looking at the Moon from *A*'s position at 6 a.m.; from the surface of the Earth, the lit side of the Moon is on the *left*. Here's a photo montage to help visualize what the 3rd Quarter Moon would look like. The horizon is real, although you have to pretend that it's 6 a.m.! Images (not to scale) of the 3rd

Quarter Moon, relatively high in the southern sky, and Sun, just rising, have been added to indicate their positions at 6 a.m. on the day of a 3rd Quarter Moon.

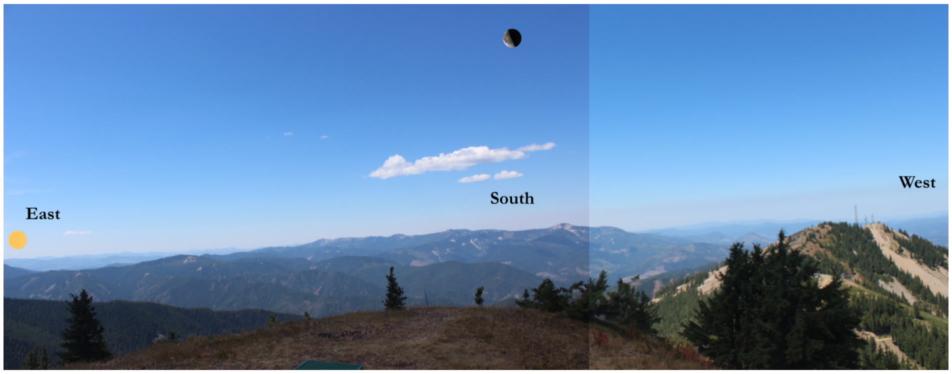


Figure 2.5: The Moon at 3rd Quarter. Moon image: NASA Scientific Visualization Studio; horizon photo: AKD.
http://svs.gsfc.nasa.gov/cgi-bin/details.cgi?aid=4404

Pause for a question: Can you visualize what the Moon will do in the next few days following 3rd Quarter? In the above photo, where will the Moon be in the sky at 6 a.m. the next 3-4 days? How will its phase be changing? Second, where in its orbit will the Moon be? Hints: the Moon orbits prograde, meaning from west to east; the Moon rises later each day by about 50 minutes; the 3rd Quarter Moon is waning, on its way toward New. In the view down onto the orbit from above, prograde means counterclockwise. The Moon takes about a week to get from 3rd Quarter to New, so 3-ish days should get us half way from 3rd Quarter to New.

In other words, relative to its position in the 3rd Quarter photo montage, on subsequent days the Moon will be farther to the east / closer to the Sun at sunrise. In the following photo montage we've enlarged the Moon, so you can see the phase more clearly and added its position at approximately 6 a.m. in the three days following 3rd Quarter.

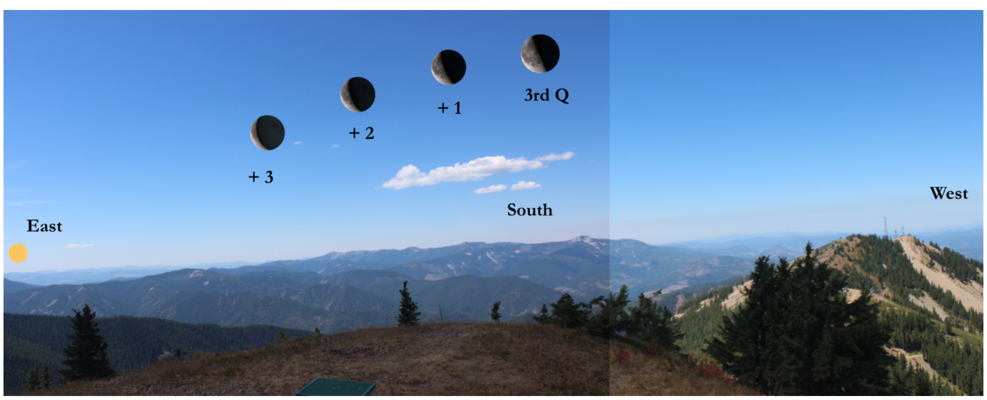


Figure 2.6: The Moon in the days following 3rd Quarter. Moon images: NASA Scientific Visualization Studio

The following sketch of the orbit looking down from above shows the Moon in the few days following 3rd Quarter and also at New. At 3rd Quarter the Moon is highest in the sky at ~6 a.m. It rises later from day to day, and by New Moon the Moon is highest in the sky at noon, with the Sun. The images outside the orbit show what the Moon looks like from the ground on those days.

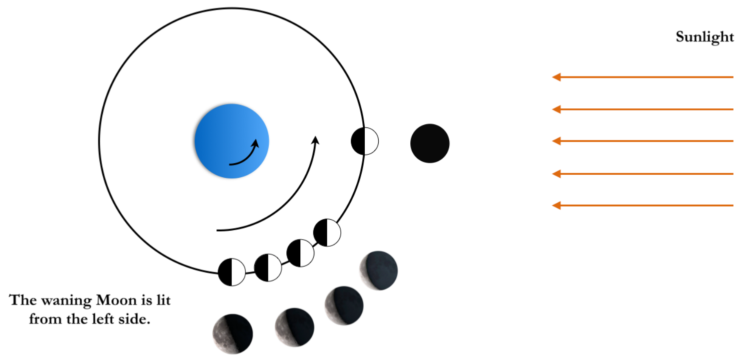


Figure 2.7: The 3rd Quarter Moon and subsequent few days. Moon images: NASA Scientific Visualization Studio

In the following sketch we will complete the month. New Moon is highest in the sky at noon with the Sun; the 1st Quarter Moon rises at about noon, is highest in the sky about sunset, and sets about midnight. Full Moon, opposite the Sun, is up all night; 3rd Quarter, as you've seen above, rises about midnight, is highest in the sky about sunrise, and sets about noon.

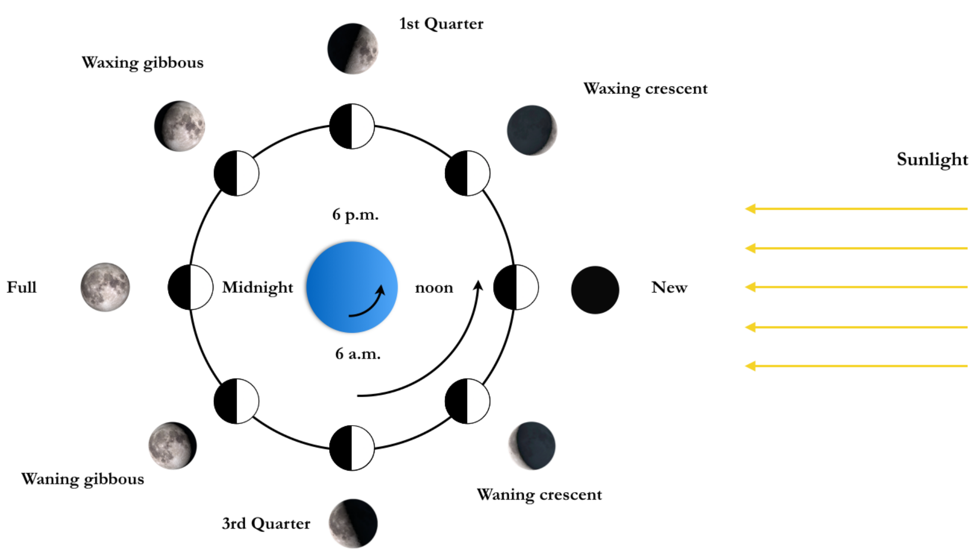


Figure 2.8: Moon phases from above and, outer images, from the ground. Moon images: NASA Scientific Visualization Studio

Imagine being on the Moon looking back at the Earth. You would see the Earth in exactly the opposite phase from the Moon phase being observed from the Earth. That's easiest to visualize at New or Full Moon. If you were on the Moon at Full, for instance, and looked back at the Earth, you'd see the nighttime side of the Earth, i.e., you'd see a New Earth. At New, or perhaps just a day or two past New, you can see this without having to travel to the Moon. An almost New Moon is seeing an almost Full Earth. Light reflects off Earth and then again off the Moon and we see a bit of the "dark" side of the Moon illuminated by Earthglow. If you haven't observed this effect, try catching the Moon shortly after New, shortly after sunset, and look at the dark portion of the Moon.

Lunar and solar eclipses

As noted above, about every six months we have lunar and solar eclipses. The ecliptic is the plane of the Earth's orbit around the Sun. For the Moon either to move through the Earth's shadow or to cross in front of the Sun the Moon must be very close to the ecliptic as well. That means the Moon must be near one of the nodes of its orbit, the points where the orbit comes up through and drops back down below the ecliptic.

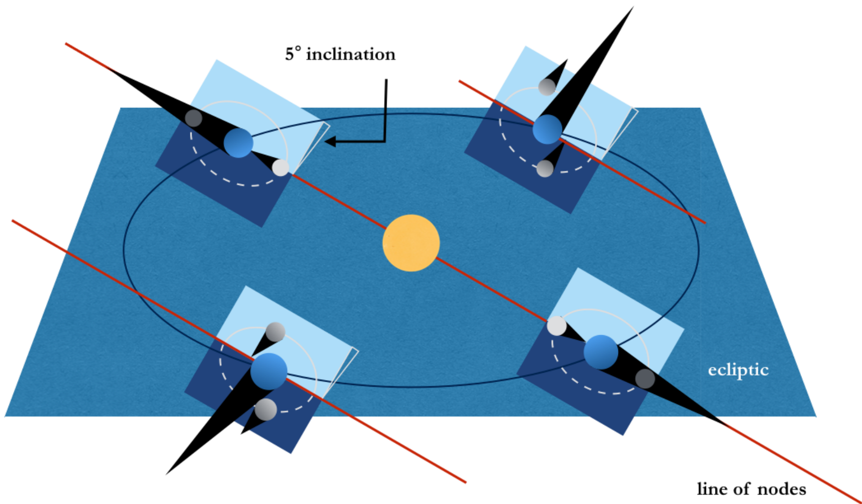


Figure 2.9: Line of nodes at different times of year. Solar image courtesy of NASA/SDO and the AIA, EVE, and HMI science teams.

The sketch above illustrates the inclination of the Moon's orbit and the fact that eclipses happen when the line of nodes is aligned toward the Sun. At those times of year the New and Full Moons will coincide with the Moon being at either the ascending or descending node of its orbit, meaning that at the times of those New and Full Moons the Moon must be in the ecliptic. At those times, called *eclipse seasons*, eclipses can occur. During the other months the Moon will be sufficiently above or below the ecliptic at New and Full Moon that there will not be eclipses.

A central lunar eclipse, where the Moon is right at a node, can last for several hours and it's visible anywhere on Earth where the Moon's up (weather permitting, of course). As the Moon enters the *umbra*, or darkest part of the Earth's shadow, you can see the curvature of the Earth on the disk of the Moon. The edge of the Earth's shadow is always round and always more gently curved than the limb of the Moon, meaning that the Earth must be round and must be larger than the Moon. For a roughly central eclipse there are four distinct points of contact, shown in the following sketch (for a lunar eclipse, but the same terminology applies for any eclipse).

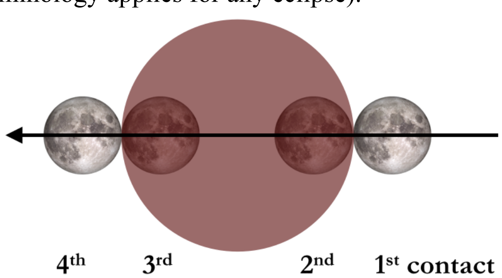


Figure 2.10: Contact points for an eclipse; Moon images: NASA Scientific Visualization Studio

The following image shows the Moon during a lunar eclipse in October 2014.

Figure 2.11: Lunar eclipse
Photo by Nathaniel Paust, 8 October 2014.

The Moon could go straight through the umbra or be offset slightly. The following figure is a sketch of possible configurations for lunar eclipses. The Moon is moving into the page. If the Full Moon is very close to the ecliptic, it will pass through the umbra during a total lunar eclipse. If the Full Moon is a bit off the ecliptic, it might fall partially in the umbra and partially in the *penumbra*. Off a bit further and the eclipse might only be penumbral, which is generally not as exciting as a total eclipse!

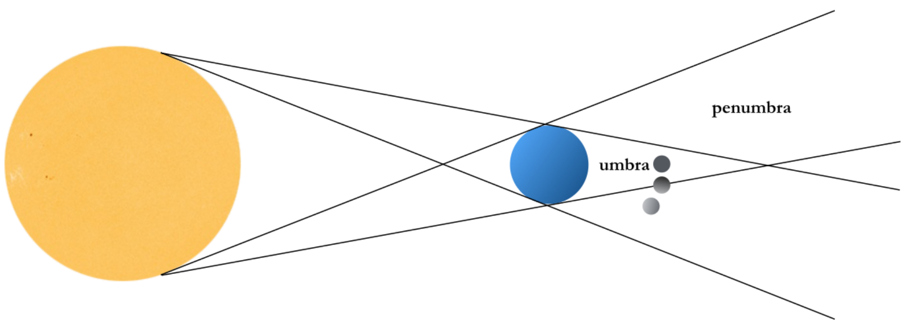


Figure 2.12: Total and partial lunar eclipses. Solar image courtesy of NASA/SDO and the AIA, EVE, and HMI science teams.

During a total lunar eclipse the Moon often looks a bit reddish. The cause is the Earth's atmosphere. Our sky is blue because the atmosphere scatters the blue and violet light in the solar spectrum. Sunlight grazing the Earth goes through the largest amount of atmosphere possible, so much that the only color that doesn't get scattered is the red. The red gets bent (refracted) into the Earth's shadow. Putting something reflective in that shadow, such as the Moon, let's us see that the Earth's shadow is reddish rather than totally dark.

The shadow cast by the Moon during a solar eclipse just barely reaches the surface of the Earth under the best conditions, meaning when the Moon is at perigee and thus relatively larger in our sky. Solar eclipses occur every six months, just like lunar eclipses, but people tend to think of them as rare, largely because the shadow path is not very wide and the likelihood of randomly being in the right place to see a solar eclipse is much lower than the chances of seeing a lunar eclipse.

If the Moon is at or near perigee when it is New and at a node, the solar eclipse will be central and the disk of the Moon will totally cover the disk of the Sun. If the Moon is at its average distance or farther away, it will be too small on the sky to cover the entire Sun and a central eclipse will still leave a visible ring of sunlight around the Moon. This type of eclipse is called an annular eclipse, from the word annulus, which means ring. If the Moon is not quite at a node the eclipse will be partial. Partial also describes the view of an eclipse for those who are just outside the path of totality. Here is an illustration of these possible orientations for the Sun and Moon during solar eclipses.

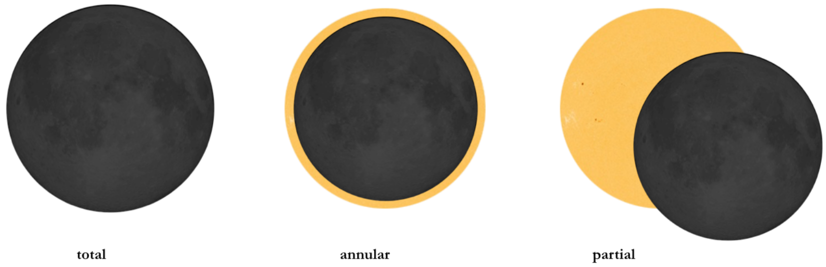


Figure 2.13: Solar eclipse geometries. Moon images: NASA Scientific Visualization Studio. Solar image courtesy of NASA/SDO and the AIA, EVE, and HMI science teams.

During a total solar eclipse it is usually possible to see the Sun's *corona*, the outer "atmosphere" of the Sun that is very hot but so tenuous that it isn't normally visible, and sometimes red arches of plasma in the *chromosphere*, closer to the Sun's visible surface.

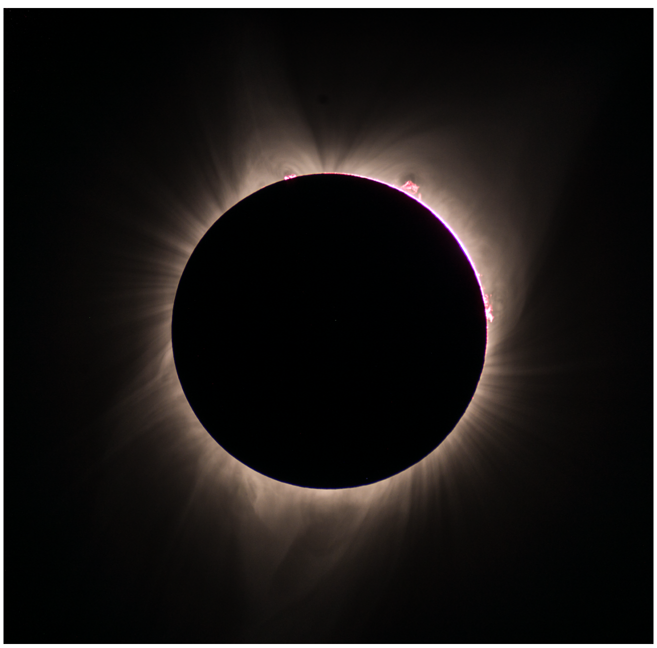


Figure 2.14: The chromosphere and inner corona during the solar eclipse of 21 August 2017. Photo by Nathaniel Paust.

When the line of nodes is pointing toward the Sun there *will* be an eclipse. This is because the node passes the Sun (or the Sun passes the node) sufficiently slowly that the disks of the Moon and Sun can't help but intersect at least partially.

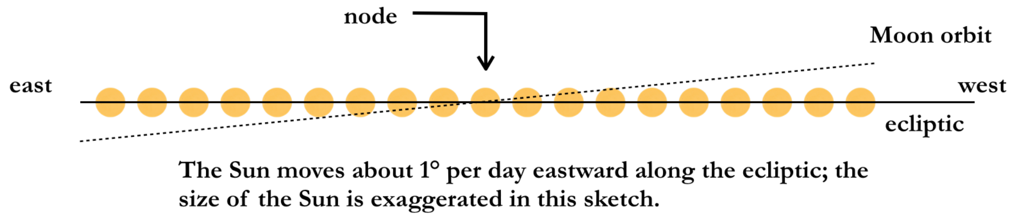


Figure 2.15: Motion of the Sun past the descending node of the Moon's orbit.

The above sketch shows a portion of the ecliptic where it intersects the descending node of the Moon's orbit. From day to day the Sun's apparent position on the sky moves from west to east along the ecliptic by about 1 degree per day. The sketch shows the Sun about every other day. If the New Moon occurs when the Sun is at the outermost positions shown in this sketch, the disks of the Moon and Sun would not intersect. But if the Sun is closer to the node when New Moon occurs, the disks will intersect. In the following sketch we've added possible positions for the New Moon to illustrate this.

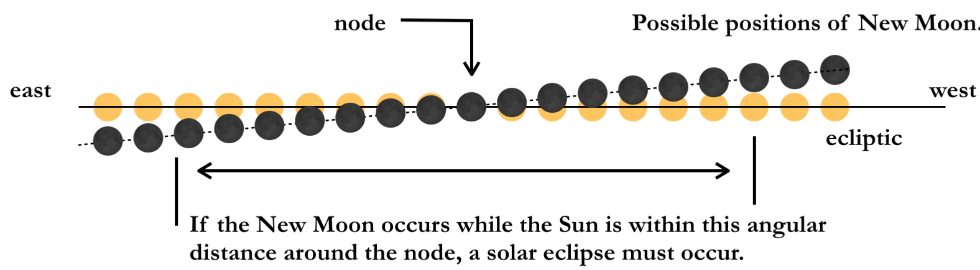


Figure 2.16: Possible positions of New Moon given possible positions of the Sun relative to the node.

It is not possible for the Sun to pass the region around the node within which an eclipse must occur in less than one month. In other words, at some point as the Sun is crawling past the node the Moon will come whipping along, passing through New. In fact, it might come around twice: if New Moon, and a partial eclipse, occurs just inside the western edge of the eclipse zone then there is likely to be a partial eclipse the next month, when the Sun is at the eastern edge of the eclipse region. If that's the case, there will have been a total lunar eclipse at the Full Moon between these two — half way in between these two New Moons the Sun must have been at the descending node and the Moon at that time be Full and at the *opposite* node of its orbit.

The line of nodes is not fixed in space; it regresses around the Moon's orbit with a period of 18.6 years. The nodical (or draconic) month is 27.2 days, a bit shorter than the sidereal month (27.3 days). This regression of the nodes means that eclipse seasons are slightly less than six months apart. In a calendar year, then, we could have as many as seven eclipses — two solar and one lunar, two solar and one lunar again, and the first solar of the next eclipse season.

The longest possible duration for the totality of a solar eclipse is about $7^1/_2$ minutes. A solar eclipse will last longer if the Moon is relatively larger on our sky, i.e., near perigee. The line connecting the apogee and perigee points of the Moon's orbit (the major axis, also called the line of apsides) precesses round the orbit with a period of 8.85 years. This means that the period between perigees, called the anomalistic month, is slightly longer than the sidereal month — 27.55 days compared to 27.32 days.

The *saros*, a period of $6585^1/_3$ days, is almost exactly equal to integer numbers of sidereal, synodic, draconic, and anomalistic months (241, 223, 242, 239, respectively). That best possible solar eclipse, when the Moon is New, at a specific node, and at perigee, will repeat about 18 years later, one-third of the way around the Earth. The saros period isn't exactly integer numbers of months. A saros series, a string of eclipses one saros period apart, starts with a partial eclipse, with the Sun just at the end of the range around the node where eclipses are possible; the eclipses in the series slowly become more central as the Sun gets slightly closer to the node for successive eclipses in that particular saros series, and then the Sun slowly migrates off to the other end of the range around the node and that saros series ends. A given saros series, i.e., a given string of these eclipses with almost exactly the same geometries occurring at ~18-year intervals, lasts for over a thousand years. Approximately 40 saros series are running at any give time. They are numbered, with even numbers corresponding to the Sun at the descending node of the Moon's orbit and odd numbers to the ascending node. For example, the total solar eclipse on August 21st, 2017, is part of saros series #145. It will be total across much of the United States. The next solar eclipse of saros 145 is on September 2nd, 2035; it will be total in northeastern Asia. Another 18 years later, September 12th, 2053, the saros 145 solar eclipse will be total around the Mediterranean. The eclipses in a saros series drift in latitude; September 23rd, 2071, people in Central America will see the solar eclipse of saros cycle 145.

A few more thoughts.

Transits, occultations, and eclipses continue to provide useful information about astronomical objects. As noted above, lunar eclipses told early sky watchers that the Earth is round and larger than the Moon. Solar eclipses gave us our first views of the Sun's corona. Here are a few more examples.

In the late 1980s the orbit of Pluto and Charon was aligned such that we, from Earth, observed a long series of mutual eclipses. The timing of the events improved our knowledge of the diameters of Pluto and Charon. The following figure is an illustration of the dip in the observed amount of light we would receive from Pluto and Charon as Charon moves in front of Pluto (considering Pluto as if it were stationary). Outside of the eclipse we receive light reflected from both of them; during the event, part of the light is blocked. Knowing the orbit size and the speeds of the two bodies we can calculate distances. Charon moves its own diameter between 1st and 2nd contact. It moves Pluto's diameter between 1st and 3rd contact.

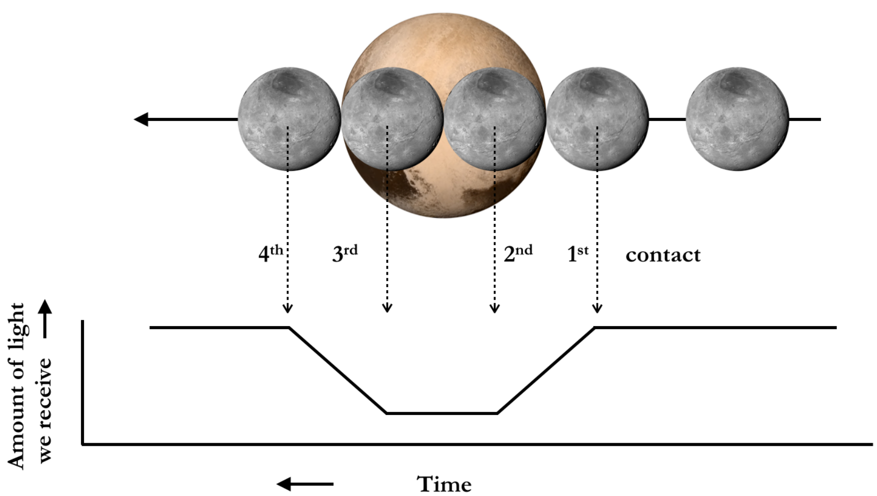


Figure 2.17: Pluto - Charon central occultation. Image credit: NASA / Johns Hopkins University Applied Physics Laboratory / Southwest Research Institute.
http://pluto.jhuapl.edu

In the chapter on properties of stars we will look at eclipsing binary stars. Binary star light curves exhibit the same behavior as the Pluto-Charon eclipse light curve, namely that the light we receive drops when one star is behind the other. From the duration of the drop, we can determine the sizes of the stars.

By early 2015 the Kepler space telescope, launched in 2009, had confirmed detections of over 1,000 exoplanets around relatively nearby stars. These planets have been discovered through observations of the dips in the light from their parent stars as the planets transit across the disk of the star. It takes multiple observations to confirm that the drop in light is due to a planet, rather than, e.g., a sunspot or a nearby variable star. (The target star isn't necessarily well separated on the sky from other stars and follow-up observations are required to determine which object is really varying and why.) Transit detections are somewhat biased toward relatively large planets close to their parent stars. Larger planets block more light from their parent star and planets with small orbits will have shorter periods which are amenable to follow-up observations. That said, Kepler detections include several potentially Earth-like planets.

Closer to home, observations of asteroid occultations provide useful information about the sizes and shapes of asteroids, including potentially hazardous objects whose orbits intersect Earth's. When an asteroid's orbit takes it in front of a star, the asteroid casts a shadow along a path on the ground similar to the shadow path of the Moon during a solar eclipse. Groups of astronomers, many amateurs with portable telescopes, position themselves across the predicted shadow path to record the timing of the drop in the star's light as the asteroid occults it. The duration of the occultation at a point in the path translates into a chord across the asteroid. Many chords analyzed together translates into a size and shape for the asteroid. By this means astronomers have determined, for example, that some asteroids are binaries. The first known rings around a minor planet, the centaur Chariklo (whose orbit is between Saturn and Uranus), were detected during an occultation in 2013.

Transits of Venus across the disk of the Sun played a role in determining the scale of the solar system. By the early 1600s we knew the sizes of the planets' orbits in AU (astronomical units; 1 AU = the average Earth-Sun distance) but it was not so easy to determine the number of kilometers in an AU. That's where the transits came in. Venus' orbit doesn't lie quite in the ecliptic; if it did, Venus would cross the disk of the Sun every time it passed us in its orbit. As it is, transits come in pairs, one corresponding to Venus at the ascending node of its orbit and one the descending node. The two transits in the pair occur 8 years apart and the pairs are separated by 105.5 and 121.5 years. The most recent transits were in June 2004 and June 2008 and before that the pair in December 1874 and December 1882. The next pair are in December 2117 and December 2125. Venus transits last for several hours. During a transit Venus is closer to us than it is to the Sun, which means that its observed path across the Sun varies depending on the observer's latitude. The following sketch shows how we can use observations of a Venus transit from known locations on Earth (with different latitudes) to determine the distance to Venus in kilometers.

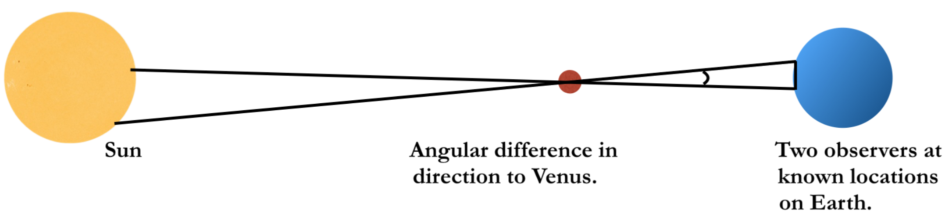


Figure 2.18: Using the transit of Venus to determine the number of kilometers per AU. Solar image courtesy of NASA/SDO and the AIA, EVE, and HMI science teams.

Having the transit timings helps make it possible to determine more precisely the angular separation of the apparent positions of Venus on the Sun for the two observers. Astronomers fanned out across the world to observe the transits of 1761 and 1769. London's Royal Society, for instance, funded the expedition of James Cook to the South Pacific in large part so that he would be in Tahiti for the 1769 transit. The analysis of the data from the combined observations from this pair of transits produced a value for the astronomical unit that was only about 2% too high. Today we combine a knowledge of the speed of light with travel times for radar reflected off of objects such as Mercury, but for the eighteenth century the transits of Venus were high science.

The following false-color image of the Sun at 170 nm was taken by the Solar Dynamics Observatory space telescope during the 2012 transit (5 June 2012, UT = 23:48:56). You can see Venus, mid-transit, and several sunspots that were present on the disk that day.

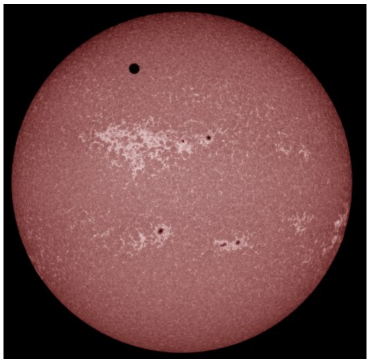


Figure 2.19:
Solar Dynamics Observatory 2012 Venus transit.
Courtesy of NASA/SDO and the AIA, EVE, and HMI science teams.
http://sdo.gsfc.nasa.gov/data/aiahmi/browse/

Gravity and orbits (with some reference to / duplication of material in the Introduction chapter)

It is not easy to tell, using just your eyes, that we are not the center of the cosmos. The Sun and Moon rise and set, stars circle around the pole, and the visible planets (the word is from the Greek for "wandering stars") move among the background stars. It doesn't feel as though we are on a chunk of rock hurtling around the Sun at ~30 km/sec or whizzing around an axis that happens, these days, to be pointed roughly toward the star α Ursae Minoris. You can tell, though, that the brightnesses of the planets vary over time, that Mercury and Venus never get too far from the Sun in the sky, and that planets periodically go "backwards", or *retrograde*, with respect to their normal eastward motion among the stars.

The development of a geocentric model of the cosmos, based on what can be seen by eye, is not unreasonable. The Ptolemaic model, named for Claudius Ptolemy (c. 100 C.E.), one of those involved in its development, is based on the assumption that the Earth is central and fixed and that the Sun, stars, and planets move around us in circular orbits. To get the variation in the planets' brightnesses, planets are assumed to trace out small circles, called epicycles, which ride along the larger circular orbit, the deferent, around Earth. To improve on that model and better predict the observed phenomena, those orbits are eventually presumed to move at a uniform rate not around the Earth but around a point called the equant, offset from the Earth. The net result of the circles upon offset circles is that an orbit tends to approach an ellipse. The basic Ptolemaic model dominated European astronomy for well over a millennium and didn't really lose sway until after the development of the telescope.

Here is a sketch of a planet on its epicycle, which is itself moving on the deferent around the Earth at a uniform rate around the equant point.

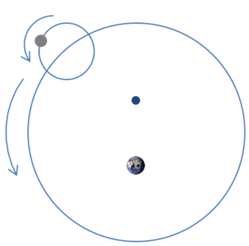


Figure 2.20: A planet on an epicycle, which is itself moving on the deferent around the Earth at a uniform rate around the equant point. Earth image: NASA / Visible Earth
https://visibleearth.nasa.gov/view.php?id=57723

There were a few ancient astronomers who thought that it would be more reasonable if the Sun were the center of the cosmos but it was not until ~1600 C.E. and the first telescopes that astronomy in Europe was able to wrap its collective head around that notion. There are some terminology notes and a few names worth noting in this regard.

First the terminology. *Elongation* describes the angle between the planet and Sun as seen from Earth at any point in time. An object that is, roughly, aligned with the Sun in the sky is said to be at *conjunction*. Opposite to the Sun, and hence up in the middle of the night, are objects that are at *opposition*. For outer planets, a 90° elongation is called *quadrature*. For Mercury and Venus, which are closer to the Sun than Earth, *greatest elongation* describes the largest angular separation between the planet and the Sun. Mercury and Venus are called *inferior* planets because their orbits are interior to Earth's; Mars and the giant planets are called *superior* planets. Just to confuse that a bit, note that Mercury and Venus will have two points of conjunction, one on the far side of the Sun and one when they move between us and the Sun. The near point is called inferior conjunction and the far point is called superior conjunction (yes, even though it is a point in the orbit of an inferior planet. . .). The following sketch illustrates these terms:

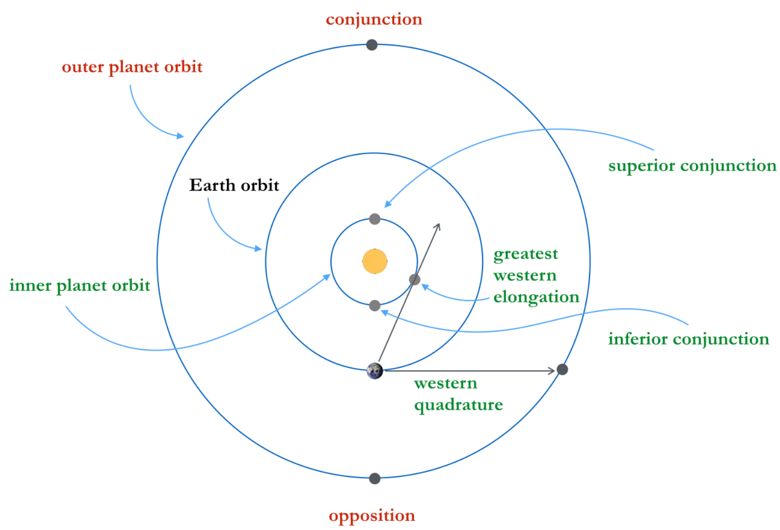


Figure 2.21: Solar system object orientations. Earth image: NASA / Visible Earth https://visibleearth.nasa.gov/view.php?id=57723

Greatest western elongation means that we would see the planet in the morning twilight, before sunrise. There's a similar greatest eastern elongation on the other side of the orbit, describing the largest separation for an inner planet in the evening twilight. Likewise for an outer planet there is a point of eastern quadrature. Next, let's review of a few of the astronomers prominent in the development of the heliocentric model.

Nicolas Copernicus, 1473 – 1543, was a Polish astronomer, mathematician, and doctor of (Catholic Church) canon law, well educated, able to read several languages; he was very much the classic "Renaissance man". From the early years of the 16th century Copernicus was already formulating the heliocentric model that would subsequently come to bear his name. He was reluctant to publish, however, no doubt to some extent because the Catholic Church had adopted the geocentric model. Ultimately his major treatise, *De revolutionibus orbium coelestium*, was published as Copernicus was literally on his deathbed. Andreas Osiander, the Lutheran theologian supervising the printing, added a comment to the beginning of the text, essentially urging the reader to consider the heliocentric model's utility for calculating planetary positions but not to take it seriously as a model of the actual cosmos.

In the above sketch of planet orientations, you can see that Venus (and Mercury) goes between the Earth and Sun. When Venus is between Earth and Sun, we see the dark side of the planet; it's like looking at a New Moon. When Venus is on the far side of the Sun, assuming we could see it, of course, it would be Full. And, it would be smaller in angular size on the sky. That's huge. Galileo (1564 – 1642) looked at Venus with his telescope in ~1610. He saw it go through phases in a manner that demanded that Venus orbit the Sun, not the Earth, really, not simply for purposes of calculations. Here is a sketch that shows how this works:

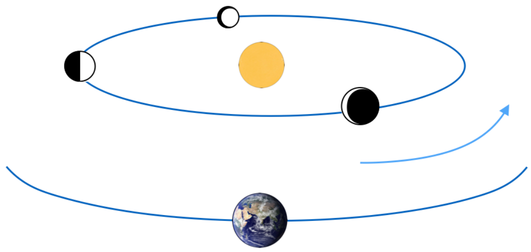


Figure 2.22:
Size and phase of an interior planet as a function of elongation. Earth image: NASA / Visible Earth
https://visibleearth.nasa.gov/view.php?id=57723

In reality the Copernican model was not much of an improvement on the Ptolemaic model when it came to calculating the positions of the planets among the stars because Copernicus was not able to give up the idea that orbits should be circles. That had to wait for another generation and better observations. Tycho Brahe, 1546 – 1601, was a Danish nobleman and arguably the greatest of the last generation of pre-telescopic astronomers. Tycho was a colorful character – he had an artificial nose to replace that which was lost in a duel, a pet moose who, it is said, died after falling down the stairs drunk, as well as the most sophisticated astronomical observatory of his day. Tycho was able to make quite precise measurements of the positions of planets (and comets, as mentioned in the solar system intro) and had the good fortune to observe a supernova (in 1572). In his last years Tycho moved to Prague, where he hired Johannes Kepler as his assistant. When Tycho died – the story goes that he suffered a burst bladder after drinking too much at a royal dinner party – Kepler inherited his data and his royal patron.

Kepler, 1571 – 1630, was deeply religious and convinced that of the underlying "harmony of the spheres". Studying the positions of Mars, Kepler achieved the breakthrough realization that the orbits of planets could be described as ellipses. Mars' orbital eccentricity is ~0.09; that's not much, but it is larger than the eccentricity of Jupiter or Saturn's orbits (Mercury's is larger, but Mercury, so close to the Sun, is tough to observe against the background stars) which undoubtedly helped Kepler see that the positions of Mars could be explained by an ellipse with the Sun at one focus.

Ellipses are described in Chapter 1; here is a recapitulation of Kepler's empirical "laws" of planetary orbits:

1. Planet orbits are ellipses with the Sun at one focus. Today we would say, with the center of mass at one focus.
2. A line connecting that focus to the planet as it orbits sweeps out equal areas in equal times. This turns out to be a statement of conservation of angular momentum.
3. The square of the orbit period is proportional to the cube of the semi-major axis. Once Newton arrives on the scene with a theory of universal gravitation we are able to say

$$P^2 = \frac{4\pi^2}{G(m_1 + m_2)} a^3,$$

where G is the gravitational constant, the *m* are the masses of the Sun and the planet, and *a* is half the length of the major axis of the orbital ellipse and is equal to the average separation between the planet and the Sun. (If you have objects of distinctly different masses it might help to capitalize M_1 to remind yourself which is larger.)

Math note: the derivation of Kepler's first law from Newtonian mechanics mostly requires more advanced mathematics than we are using here. You can, with a small amount of calculus, see roughly how to get the second law. Consider the following orbital geometry, in which we've got a planet moving along its orbit from one point to another in a unit time dt. In that time the angle swept out by a line connecting the Sun and the planet changes by an amount $d\theta$. The area of a triangle is ½ base x height. Assume that the angle $d\theta$ is small, meaning that r doesn't change too much in dt and that the base of the triangle is approximately $rd\theta$. Then the area of the triangle swept out will be $dA = ½ r^2 d\theta$. The angular momentum of an object is the linear momentum times the radius from the pivot point; in this case, our planet's orbital angular momentum will be *mvr*, where *m* is the mass of the planet. The velocity can be expressed as $v = r\, d\theta/dt$. Yes, there would normally be an *r*-component to the velocity as well as a θ-component but for small changes in *r* we can ignore that term.

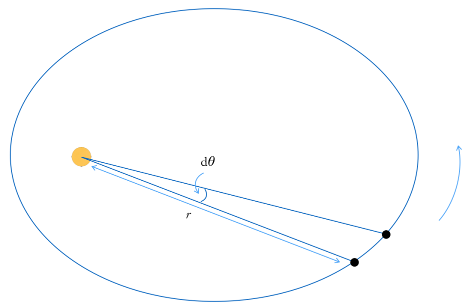


Figure 2.23:
Geometry for Kepler's second law

Conservation of angular momentum says

$m\, r2\, d\theta/dt$ = constant, or

$2\, m\, dA/dt$ = constant.

Unless the planet mass is changing, *dA/dt* must be a constant. In other words, Kepler's second law follows from the principle of conservation of angular momentum. What we observe is that solar system objects move fastest near perihelion and slowest near aphelion (farthest from the Sun).

Another math note. We can get at Kepler's third law, at least for the specific case of circular orbits, by assuming that the forces acting on the two bodies, the Sun and the planet, are balanced. The following sketch shows the Sun and a planet in orbit around their mutual center of mass.

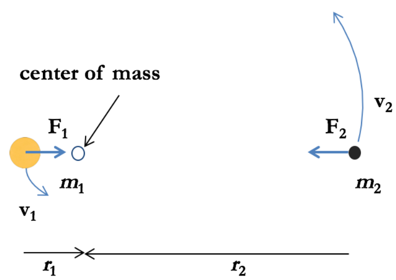


Figure 2.24:
Geometry for center of mass and forces in a rotating system

The center of mass is defined by the relation

$m_1\, r_1 = m_2\, r_2$.

The planet, with less mass, is farther from the center of mass and must have a larger velocity since both objects complete one revolution around the center of mass in the same period, *P*. If we drew the relative orbit of the planet around the Sun, then $r_1 + r_2 = a$, the semi-major axis of the ellipse that describes the planet's orbit. The forces acting on the two objects are equal in magnitude and oppositely directed. The magnitude of a centripetal force is mv^2/r. The centripetal forces must equal the gravitational force acting on the two objects: $F_{\text{grav}} = Gm_1m_2 / a^2$.

We specified circular orbits, so the speed is given by $2\pi\, r / P$. Thus we have

$$F_1 = F_2 \rightarrow 4\pi^2 m_1 r_1 / P^2 = 4\pi^2 m_2 r_2 / P^2 \text{ and}$$

$$r_1 = m_2 a / (m_1 + m_2)$$

which, combined with the expression for the force of gravity, yields Kepler's third law:

$$4\pi^2 m_1 \left[\frac{m_2 a}{m_1 + m_2} \right] / P^2 = \frac{G m_1 m_2}{a^2}$$

or, cancelling the redundant factors of mass,

$$P^2 = 4\pi^2 a^3 / [G(m_1 + m_2)].$$

Yet another math note. In the Introduction chapter you will also find the *vis viva* equation, describing the speed of an object at any given point along its orbit. We can get at that expression by considering the conservation of energy. The total energy of the system described by the two objects in the figure above is the sum of their individual kinetic + potential energies. That total is constant, unless some additional force intrudes:

Total energy = $\frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 - G m_1 m_2 / r$,

where we are using the fact that the gravitational potential energy is the derivative of the gravitational force and r is the instantaneous Sun – planet separation. If the system isn't moving through space, then the linear momenta of the two objects must be equal in magnitude and oppositely directed. In terms of the speeds, this means that

$m_1 v_1 = m_2 v_2$.

Using the total speed, $v = v_1 + v_2$, we can solve the conservation of momentum equation for either of the individual speeds:

$$v_1 = m_2 v / (m_1 + m_2)$$
$$v_2 = m_1 v / (m_1 + m_2).$$

Substitute these into the expression for total energy, cancel excess factors of (m_1+m_2) and factor out the $m_1 m_2$:

$$\text{T.E.} = m_1 m_2 \left[\frac{v^2}{2(m_1 + m_2)} - \frac{G}{r} \right].$$

At this point we need to consider the relative speed at perihelion, where the speed is the highest, to see that the orbit is bound. At perihelion the speed will be totally tangential (no radial component) and (after a bit of calculus) given by

$$v_{\text{perihelion}} = \frac{2\pi a}{P} \sqrt{\frac{(1+e)}{(1-e)}}.$$

Squaring $v_{\text{perihelion}}$ and using Kepler's third law gives us

$$v_{\text{perihelion}}^2 = \frac{4\pi^2 a^2}{\left[\dfrac{4\pi^2 a^3}{G(m_1 + m_2)} \right]} \frac{(1+e)}{(1-e)}.$$

At perihelion, $r = a\,(1 - e)$. Thus

$$v_{\text{perihelion}}^2 = \frac{G(m_1 + m_2)}{r}(1+e) \text{ and}$$

$$\text{T.E.}_{\text{perihelion}} = m_1 m_2 \left[\frac{G(m_1 + m_2) \cdot (1+e)}{r \cdot 2 \cdot (m_1 + m_2)} - \frac{G}{r} \right] = -\frac{G m_1 m_2}{2r}(1-e) = -\frac{G m_1 m_2}{2a}.$$

The fact that the total energy is negative tells us that the orbit is bound. Since T.E. is conserved, it also gives us an expression for the total energy at any point in the orbit. Plug that back in, above, and solve for v^2 and we obtain the final expression (the *vis viva* equation) for the orbital velocity:

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

Two specific cases are of particular use: An infinitely long orbit corresponds to an object having just exactly attained escape speed; in this case, $a = \infty$ and

$$v_{\text{esc}}^2 = 2G(m_1 + m_2) / r.$$

The second interesting case is a circular orbit, in which case r is always the same and is equal to a, giving

$v_{\text{circ}}^2 = G(m_1 + m_2) / r$. (And of course circular velocity is also $v = 2\pi r / P$.)

An additional note on circular velocity: If m_2 is small and $r_1 \approx r$, it's a simple exercise to obtain the approximate circular velocity by equating the centripetal and gravitational accelerations acting on m_2.

$$a_{cent.} = \frac{v^2}{r};\ a_{grav.} = \frac{Gm_1}{r^2};\ \text{setting these equal} \rightarrow v^2 = \frac{Gm_1}{r}.$$

Planet orbits are nearly circular and all lie nearly in the same plane, the ecliptic, which is close to lining up with the Sun's equator. Objects farther out in the solar system (or in the asteroid belt) often have orbits that are more eccentric and substantially inclined to the ecliptic. So far we have characterized orbits by their size (a) and eccentricity (e). Let's consider how to characterize the orientation of a given orbit in space. The practice is to describe the orientation of the orbit by three angles: inclination (i), longitude of the ascending node (Ω), and the argument of the pericenter (ω). Ok, what do those terms mean?!:

- Inclination: the tilt of the plane of the orbit with respect to the ecliptic.
- Ascending node: the point along the orbit where the object "ascends" from the south side of the ecliptic to the north side.
- Longitude: measured eastward (counterclockwise when viewed from the north) along the ecliptic from the point of the vernal equinox (the point in Pisces where the Sun crosses the equator moving northward in March).
- Pericenter: generic term for the end of the major axis that is closest to the center of mass; e.g., perihelion, perigee.
- Argument of the pericenter: the angular distance around the orbit from the ascending node to the pericenter.

These five numbers – a, e, i, Ω, and ω – are called the "orbital elements". Let's look at an object in a retrograde orbit label the angles:

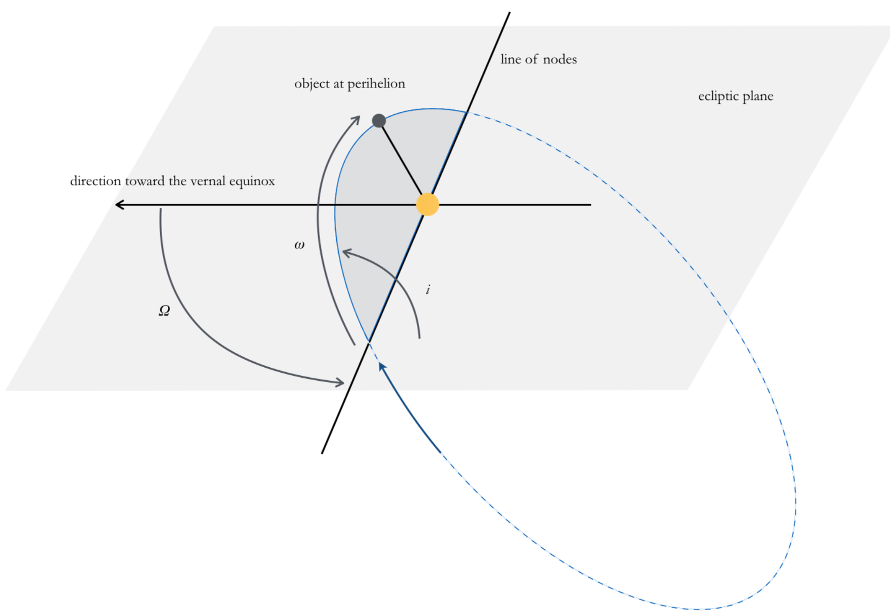


Figure 2.25: Orbital elements for an object in a retrograde orbit

Many comets, such as Halley, orbit the Sun retrograde, in the opposite direction from the orbit of Earth and other planets. We indicate this by an inclination (i) that is more than 90°. The line toward the vernal equinox indicates the direction to 0° ecliptic longitude; in other words, this is the direction from Earth toward the Sun on the day of the vernal equinox. The angle from 0° to the ascending node of the orbit is the angle given by the Ω in this sketch. Those two angles give the plane in which the orbit lies; the third angle, ω, tells us how the long axis of the orbit is oriented within that plane.

Orbits, of course, can also describe the paths of artificial satellites. The Soviet Union put the first artifical satellite, Sputnik 1, into Earth orbit in October 1957. Newton's third law of motion says that forces come in pairs, each force balanced by another that is equal in magnitude and oppositely directed. Barring catastrophic failure on the launch pad, the force shooting the exhaust gases from a rocket engine are balanced by the force shoving the rocket itself skyward. The exhaust gases don't have to push on anything — rocket engines work fine in the emptiness of space.

Pause for a math note: Newton's second law is, more properly, not $\vec{F} = m\vec{a}$ but rather $\vec{F} = {d\vec{p}}/{dt} = m \cdot {d\vec{v}}/{dt} + \vec{v} \cdot {dm}/{dt}$. Usually we ignore the latter term because there is no appreciable change in mass, but without that term rockets don't go to space.

Using the velocity equation, above, we can calculate that a rocket will need to provide a satellite with a speed of ~11 km/sec to escape Earth's gravity. The New Horizons spacecraft which flew past Pluto in the summer of 2015 was launched in 2006 with a speed of 16.3 km/s, and given a further grativational boost by a close encounter with Jupiter en route.

Geocentric orbits, for those spacecraft remaining closer to home, are classified roughly by their altitude. Low-Earth orbits range from ~160 to ~2,000 km above sea level. Much lower than that and the atmosphere provides too much drag on the spacecraft, making its orbit decay too rapidly. Roughly have of all artificial satellites are in low-Earth orbit. The International Space Station, for example, orbits at ~400 km altitude; at that height, its orbit period around Earth is ~93 minutes. The Hubble Space Telescope is a bit higher, ~540 km altitude. Geosynchronous orbits have altitudes of nearly 36,000 km and have orbit periods equal to one sidereal day (23^h56^m). A satellite in a geosynchronous prograde orbit above the equator (i.e., 0° inclination, orbiting in the same direction as the Earth's rotation) is geostationary, remaining at a fixed point in the sky for a ground-based observer. Communications and weather satellites, for which continuous communication with the ground is desireable, are often placed in geostationary orbits. Global Positioning System (GPS) satellites are in between, in medium-Earth orbits at ~20,000 km altitude. There are on the order of 1,000 functioning artificial satellites in Earth orbit these days. But add in dead satellites, spent booster rockets, debris from collisions, i.e., all the other stuff that's up there, everything over about 1 cm diameter, and there are several hundred thousand pieces of material in Earth orbit. Earth orbit is a rather polluted place.

The most efficient means to get a spacecraft from one orbit to another, without using another planet for a gravitational boost, is to put the spacecraft into an orbit that has its perihelion at one orbit and its aphelion at the other. An orbit of this type is called a least-energy orbit or a Hohmann transfer orbit. For example, getting from Earth to Mars on a least-energy orbit, the perihelion of the spacecraft's orbit would be at Earth and the aphelion would be at Mars. Hopefully we launch our spacecraft at the appropriate time so that the spacecraft gets to Mars, not just to Mars' orbit! How long would it take to get to Mars this way? About 8 and a half months.

> The calculation: q = 1 AU and Q = 1.52 AU. Averaging those: (1 + 1.52) / 2 = 1.26 = a for this orbit. Using Kepler's third law we find $P = (1.26^3)^{1/2}$ = 1.414 years. We only need half of that to get from Earth to Mars, or 0.71 years.

The three-body problem, tides, rings, resonances

What we have done so far with orbits has, mostly, dealt with the assumption that we only have two objects, the Sun (or the Earth) and another smaller body. Introducing a third body into the problem makes the mathematics a good deal harder.

At this point, it would be good to read / review the section on Lagrange points in the Introduction chapter. Briefly, Lagrange points are derived by considering the places where the forces on a small test mass due to, e.g., a planet and the Sun, are balanced for a revolving system. At the Earth, we use the L1 and L2 points for spacecraft. L1 is on the sunward side, about 1.5 million km from Earth. A spacecraft placed there will orbit the Sun with a period of one year. In the absence of the Earth, a spacecraft closer to the Sun than 1 AU would have a shorter period; with the Earth present, the spacecraft feels the gravitational attraction of both the Sun and the Earth and the result is an orbit period of one year. L2 is on the anti-sun side of Earth, and a spacecraft here also has an orbit period of one year. L1 and L2 (and L3, on the far side of the Sun) are not very stable and it wouldn't take much of a nudge to displace an object from one of these point, which is probably why we don't find collections of asteroids piling up there.

The points L4 and L5, at 60° ahead and behind the planet in its orbit, are more stable. Objects there, nudged a bit, tend to oscillate (or *librate*) around the Lagrange point. At these points we find the Trojan asteroids. Trojans are, at least as far as we know now, most abundant in the orbit of Jupiter. They are called Trojans because the convention has been to name the Jovian ones for figures from the Trojan War. Mostly the Greeks are at the L4 point and the Trojans are at L5, but there are a couple of "spies": 624 Hektor (diameter ~200 km) is in the L4 crowd and 617 Patroclus orbits with the Trojans at L5. There are over 6,000 catalogued Trojans and possibly as many as $10^{5\text{-}6}$ in the kilometer-size range as yet undiscovered.

The first Earth Trojan was discovered in 2011. Mars and Neptune have several known Trojans and estimates are that Neptune should have many that are too faint to have been detected yet. Venus and Uranus are known to have at least one Trojan. Two of the moons of Saturn, Tethys and Dione, each have two Trojans of their own.

Two of Saturn's small moons demonstrate a related type of three-body orbit. Janus and Epimetheus execute what are called horseshoe orbits. Here is a sketch of what one of the moon's orbits looks like as seen from the second moon. The moon that's initially on the inner track will be moving faster and approach the moon that's initially on the outer track. The two will gravitationally attract each other. The inner moon will gain energy, which causes it to move to the higher orbit; the second moon will lose energy, causing it to drop to the inner orbit.

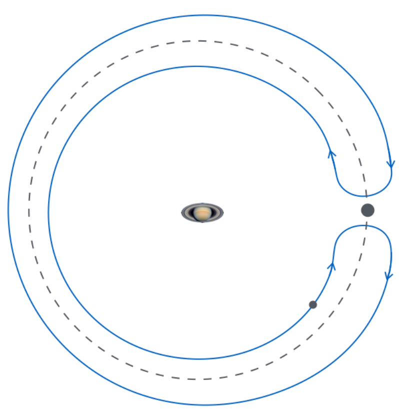


Figure 2.26: Horseshoe orbits
Saturn Hubble Space Telescope image credit: NASA, ESA, and Erich Karkoschka (University of Arizona)
https://photojournal.jpl.nasa.gov/catalog/PIA05982

Eventually the moons will approach again, and again swap orbits. From the perspective of the second moon, the one we're holding stationary in this sketch, the first moon approaches on an inner track, pauses, and recedes on an outer track; much later, the first moon will approach on the outer track, pause, and move away on the inner track. On average the two moons have nearly the same size orbit around Saturn; moons in a configuration such as this are called co-orbitals.

Janus and Epimetheus are not exactly the same mass so the orbital swapping isn't exactly symmetric. Janus is about four times more massive than Epimetheus, with a diameter of ~180-km compared to Epimetheus' 120-ish km. When the two approach each other, every ~4 years, Janus' orbit changes by ~20 km and Epimetheus' orbit changes by ~80 km. They aren't likely to collide – the closest approach still leaves them about 10,000 km apart.

On a related note, here's something to watch for in space movies to see if the writers know their orbital mechanics: Imagine being in a spacecraft, in Earth orbit, behind the International Space Station. You want to catch up to the ISS. Should you fire your rear-directed rockets to speed up? No, because if you do you'll have given your spacecraft more energy, which will move you into a higher orbit with a longer period, and you'll fall farther behind. . .

Tidal effects show up in quite a few places in astronomy, which is why they are described in the Introduction chapter. Now would be a good time to (re)read that section, so that we can extend those ideas to examples that are specific to the solar system, such as planetary rings. The basic idea is that because gravity falls off with distance an object will feel a net tidal force because of the difference in the gravitational force from one side of the object to the other. The classic example is that the force of gravity on the Earth due to the Moon is strongest on the Moon-facing side of the Earth and less on the far side of the Earth. The result is the tidal bulges, as the Earth is pulled into a slightly elongated shape by the tidal force. Because it's a differential force, the tidal force falls off faster than the gravitational force. The tidal force goes as $1/r^3$ rather than the $1/r^2$ of the gravitational force.

Rings around solar system objects happen when the tidal force exceeds the gravitational force holding particles together. In other words, too close to, say, Saturn, and we will have rings and not another moon. Specifically, consider two small masses in orbit around a planet with very slightly different orbital radii r and $r + dr$. The two will feel a mutual gravitational attraction

$$F_{\text{grav}=}Gm^2/(dr)^2,$$

where we are assuming that the two small objects have approximately the same mass.
The differential tidal force will be the difference in the gravity each feels from Saturn,

$$dF_{\text{grav}} = [2GMm/r^3]/dr\ .$$

The Roche limit is the point where these two forces are equal. For any orbit interior to the Roche limit, the two particles will have sufficiently different orbits that they will drift apart. Exterior to the Roche limit, the two particles will be able to stay together as a moon. Setting these two forces equal and solving for r gives

$$r_{\text{Roche}} = (2M/m)^{1/3}dr\ .$$

Reality is a bit more complicated, in that a small moon isn't just held together by the gravitational attraction of two half-moons. How cohesive a moon is depends on what it's made of, and theoreticians have performed various Roche limit calculations based on assuming that a moon is made of liquid water (i.e., zero strength to hold itself together), or is made of ice, or is made of rock. How large the small moon is matters as well. Obviously if you are small enough, you don't break up or else we would never be able to put satellites into Earth orbit. Now let's look at individual planet's ring systems.

Jupiter's rings were discovered in 1979 by the Voyager 1 spacecraft. There are four relatively distinct components: The brightest main ring is fairly thin, on the order of 100 km, and bounded by the orbit of the small moon Adrastea. Interior to that is a thick "halo" ring of tiny dust particles. Beyond the main ring are two "gossamer" rings, associated with the orbits of the small moons Amalthea and Thebe, both composed of micron-sized dust particles. Only the main ring has larger rocky chunks, although how many and how large isn't well constrained. The following is a mosaic of images of Jupiter's rings taken by the Galileo spacecraft. Each successive image is taken at increased sensitivity since the rings get fainter farther from Jupiter. The Sun was behind Jupiter.

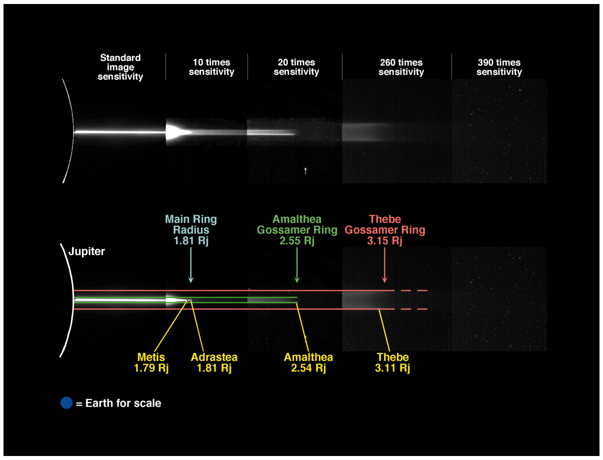


Figure 2.27: Rings of Jupiter

Image credit:
NASA / JPL / Cornell University;
Galileo spacecraft
http://photojournal.jpl.nasa.gov/catalog/PIA01623

Dust or chunks? Let's pause here for a moment to talk about forward and backward scattering. Dust grains preferentially scatter incoming light forward, in the same direction. Larger chunks of rock, ice, metal, whatever, will preferentially reflect light backward. Here's a sketch to illustrate this:

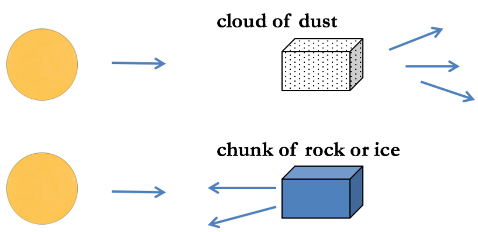


Figure 2.28: Scattering vs. reflection of sunlight

Our Earth-bound perspective doesn't allow us to observe the rings of the outer planets from the far side, to see if there is dust. Spacecraft – the Voyagers, Galileo, Cassini – have given us that view because they can get beyond a planet and look back, toward the Sun. As a rather mundane example, think about a dusty car windshield. If you drive into the sunlight with a dusty windshield, the dust particles will forward scatter the sunlight into the driver's eyes, which is no fun. Bird droppings on the windshield, on the other hand, reflect sunlight back; from inside the car, that spot on the windshield just looks dark and doesn't obstruct the driver's view nearly as much as all that bright dust.

Here is an amazing backlit image of the rings of Saturn.

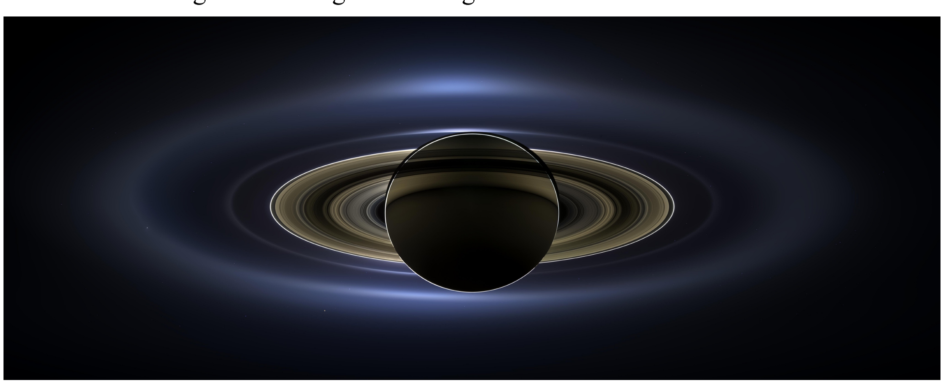


Figure 2.29: Rings of Saturn; NASA / JPL / SSI; Cassini spacecraft;
http://photojournal.jpl.nasa.gov/catalog/PIA17172

There's a pale blue dot in the lower right that is Earth. The brighter dots on the left and lower left are Saturn's moons Enceladus and Tethys. We are more accustomed to the reflected view, such as in the following Hubble Space Telescope image.

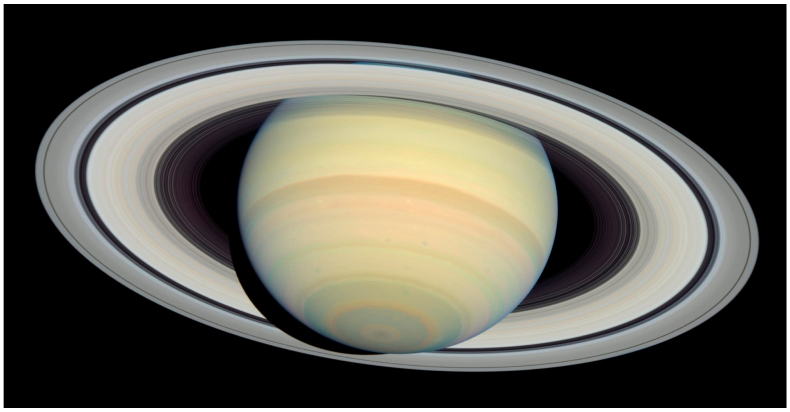


Figure 2.30: Saturn

Hubble Space Telescope 2004 image credit: NASA, ESA, and Erich Karkoschka (University of Arizona) https://photojournal.jpl.nasa.gov/catalog/PIA05982

Comparing these two images, you should be able to see which parts are dustiest, because they will be bright in the backlit image and dark in the lower image. Similarly, the brightest part of the rings in the lower image has relatively large chunks of ice (and a little rock); those particles look dark in the upper, backlit image, because they are reflecting light back toward the Sun and the spacecraft is just seeing the large particles' dark sides.

The regions in the rings were given identifying letters, as shown in the following image. There's also an E ring, very broad but farther out, by the orbit of Enceladus. For scale, the F ring is about 140,000 km from the center of Saturn, and Enceladus's orbit is nearly 240,000 km. The E ring is broad and composed of microscopic particles probably ejected from the geysers on Enceladus. Between the F ring and the E ring is the thin, faint, G ring. It isn't equally bright all the way around Saturn. Ring arcs are not abnormal, and may be associated with localized dust ejected from a small moon, possibly by collisions. Tiny moons have low escape speeds and it isn't hard to splatter dust off of their surfaces. One of the most recent discoveries is a ring out near the orbit of Phoebe, a 200-ish km, retrograde, moon 13 million km out from Saturn. This ring was only discovered in 2009, in images made with the infrared Spitzer Space Telescope. Phoebe and this ring orbit roughly in the place of the ecliptic, rather than in the plane of Saturn's equator.

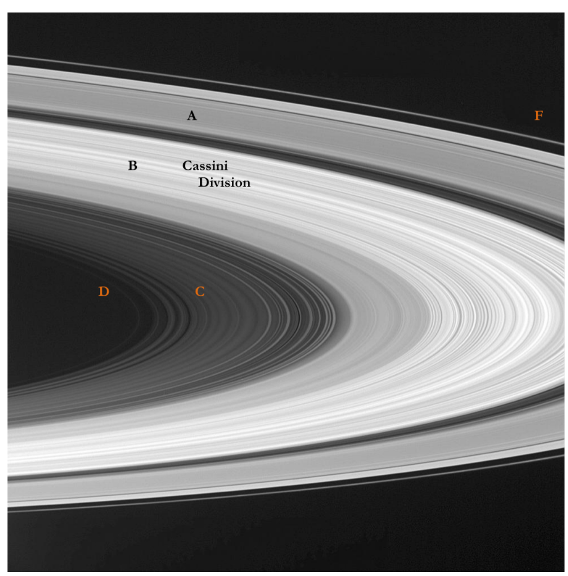


Figure 2.31: Saturn's rings

Credit:
NASA / JPL / SSI; Cassini spacecraft
http://photojournal.jpl.nasa.gov/catalog/PIA06536

It makes sense that some of the outer rings are icy dust kicked off the surface of a small moon or ejected from Enceladus' geysers. It's less clear how or when the inner main rings formed. Were they left-over material from when Saturn formed or were they a moon or a comet that strayed inside the Roche lobe and was, perhaps with the aid of a collision, broken to bits? Have rings been there throughout the age of the solar system or are they more recent? Analysis of some of the last data from the Cassini mission, as it plunged inside the rings in 2107 on its last few orbits prior to impacting Saturn, suggests that the material currently making up the rings hasn't been there for more than a few tens of millions of years. It's possible that rings have formed and dissipated more than once in the history of the solar system.

Uranus' rings were discovered during an occultation in 1977, when Uranus passed in front of a star. At the time Uranus was nearly pole-on to us, so the rings caused extra small dips in the observed light curve as they temporarily blocked the starlight. Additions to this family of rings were discovered by the Voyager 2 spacecraft and Hubble Space Telescope observations. Uranus' rings are very dark – Bond albedo ~0.02 – and tend to be quite narrow. Here is a Hubble Space Telescope image from 2005 showing most of Uranus' main rings.

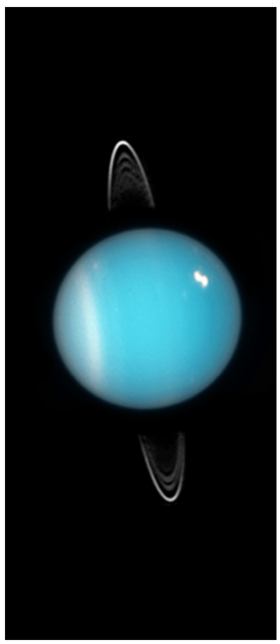


Figure 2.32: Uranus' rings
Credit: NASA / ESA / M. Showalter; HST
http://hubblesite.org/newscenter/archive/releases/2007/32/image/c/

How to constrain narrow rings is an interesting question. Ring particles ought to interact and eventually spread out. One mechanism, that seems to apply to at least one of Uranus' rings as well as Saturn's narrow F ring, is gravity, in the form of nearby shepherd moons. The following image illustrates how this works. It's is based on a sketch in *Rings*, a book by James Elliot and Richard Kerr. Elliot was one of those involved in the discovery of the Uranian ring system. The idea is that the ring lies between the orbits of two small moons. The inner moon, orbiting faster, gets ahead of some small region of ring particles. It tugs on them, backwards as it is catching up and then forwards as it moves ahead. The moon spends relatively more time accelerating the ring particles; the result is that they move into a higher energy orbit, farther out. When that region of ring particles later catches up with the outer moon, that moon tugs them backwards, and the ring particles fall into a lower orbit. The result is to confine the ring particles to a narrow orbit. This is similar to the physics involved in the co-orbital satellites and their horseshoe orbits.

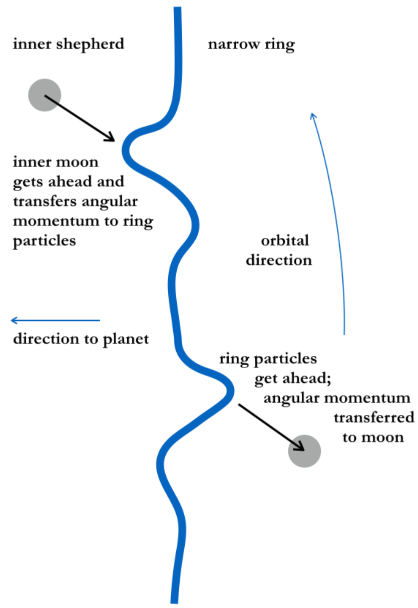


Figure 2.33:
Shepherd moons; adapted from fig 4-2 in Elliot and Kerr *Rings.*

We have observations of some shepherd moons, but not for all narrow rings. It's not clear whether the shepherds are present but too small to have been discovered yet or whether there is some other mechanism that also works to confine rings. The following is a Voyager 2 image of Uranus' moons Cordelia (1986 U7) and Ophelia (1986 U8), shepherds of the epsilon ring. The two moons are small – ~40 km diameter – and dark.

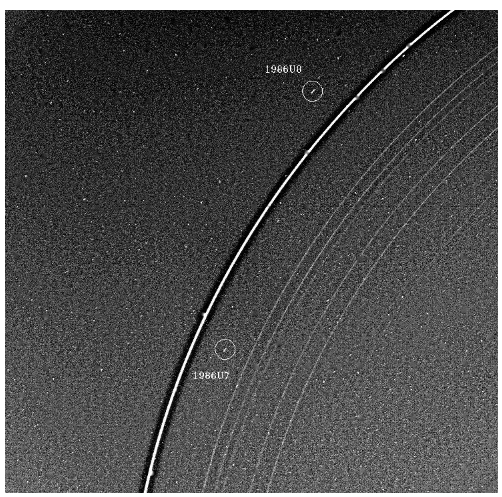


Figure 2.34: Uranian shepherds
Credit: NASA / JPL; Voyager 2
http://photojournal.jpl.nasa.gov/catalog/PIA01976

Neptune's rings are clumpy; ring arcs were discovered in the 1980s, by ground-based observers in the southern hemisphere. Neptune was at that time in the constellation Sagittarius, where, with a declination of ~ –22°, it was much better positioned for southern observatories. These rings are quite dark and relatively dusty. The five main rings have been named for astronomers who played significant roles in the discovery and early studies of Neptune: Galle, Le Verrier, Lassell (discoverer of Triton), Arago (Le Verrier's teacher), and Adams. Here is a pair of Voyager images showing Neptune's rings.

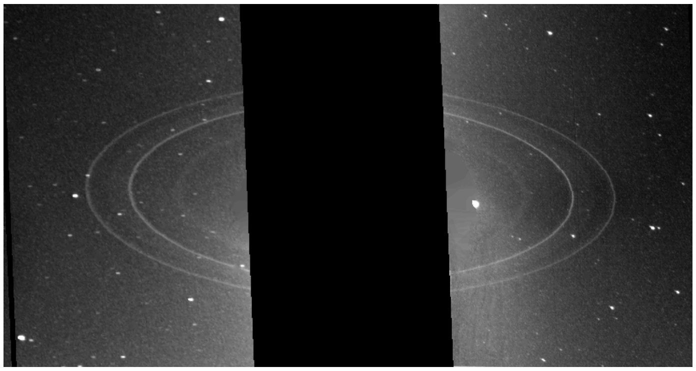


Figure 2.35: Neptune's rings
Credit: NASA / JPL; Voyager 2
http://photojournal.jpl.nasa.gov/catalog/PIA02202

Broad rings often have gaps, such as the prominent Cassini Division separating Saturn's A and B rings. The Cassini Division is named for the Giovanni Cassini, the astronomer who first recorded it, in Paris in 1675, and for whom the Cassini spacecraft is also named. The gap isn't totally empty, but there's definitely a lot less material here than right next door in the A and B rings. One principal reason is Mimas. Ring particles near the inner edge of the Cassini Division have an *orbital resonance* with Mimas; those ring particles orbit Saturn twice for every one revolution of Mimas. The result is that particles at that location get tugged on regularly, in the same place every time, and get nudged out of that resonant orbit.

Gaps also arise if there's a small moon buried within the rings. The Encke Gap in Saturn's A ring is much narrower than the Cassini Division – 325 km vs. 4,800 km – and seems to be due to Pan, an unevenly shaped moon, 35 x 31 x 21 km in size. A very narrow ring within the Encke Gap is made of particles kept in horseshoe-like orbits by Pan (the god of shepherds). Not all of the particles are in well-behaved horseshoe orbits or neatly nudged out of the gap. Some seem to have accreted onto Pan itself, which is shaped a bit like a walnut, with a ridge around its middle. Another moon, Daphnis, only 8 kilometers long, orbits in the Keeler Gap, near the outer edge of the A ring. The Keeler Gap is only 42 kilometers wide. Daphnis' orbit is slightly inclined and as it passes the particles in the A ring, on either edge of the Keeler Gap, it excites waves which, when the illumination angle is right, cast shadows on the surrounding ring. The Cassini spacecraft took the following image not too long before Saturn reached its equinox in August 2009. In other words, the angle of the sunlight was low and the ripples in the ring cast shadows long enough to be noticed. Daphnis itself is a bit hard to see, but it's the little white dot at the base of the long skinny vertical shadow.

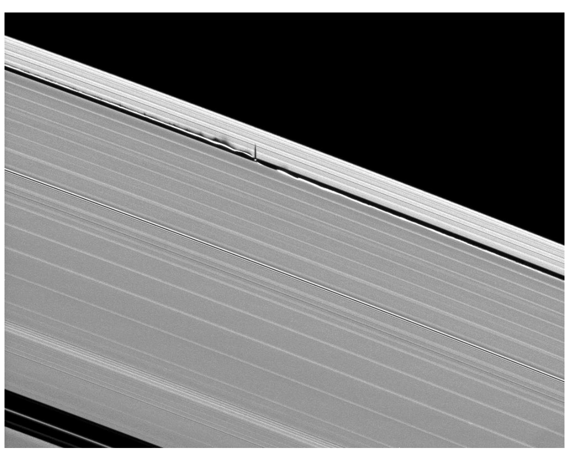


Figure 2.36: Ripples in Saturn's rings

Credit: NASA / JPL / Space Science Institute; Cassini
http://photojournal.jpl.nasa.gov/catalog/PIA11654

Gravitational resonances show up in the asteroid belt, also, in the Kirkwood gaps. Asteroids with orbits at 2.5 AU would have orbit periods in a 3:1 ratio with Jupiter. There are very few asteroids with orbital semi-major axes at 2.5 AU. That doesn't mean there are no objects at 2.5 AU – asteroid orbits are sufficiently eccentric that there are always plenty of objects at 2.5 AU, just very few with orbits of that size. The outer edge of the asteroid belt, at 3.28 AU, corresponds to the 2:1 resonance with Jupiter.

Resonances can be places of stability as well – think of the Jovian Trojan asteroids, in a 1:1 resonance with Jupiter. Pluto is in a 2:3 resonance with Neptune, which is quite stable because whenever Pluto is at perihelion, it is well above the ecliptic and Pluto and Neptune don't get close enough to each other for Pluto's orbit to be significantly perturbed by Neptune. The 3:2 resonance with Jupiter is home to a group of asteroids called the Hildas. Asteroid 153 Hilda has an orbital semi-major axis of 3.973 AU and an eccentricity of 0.141. When it's at aphelion it tends to be either opposite Jupiter or near the L4 or L5 Lagrange points. In other words, like Pluto and Neptune, Hilda tends to avoid getting too close to Jupiter. Below is a sketch of the distribution of asteroids as a function of semi-major axis, based on data from the Minor Planet Center, the primary location for information about all known comets, asteroids, Centaurs, TNOs (Trans-Neptunian Objects), etc., etc. How to tell whether a resonance will be stable and collect objects or whether it will be unstable and objects will get ejected from that orbit is not a simple straightforward math problem.

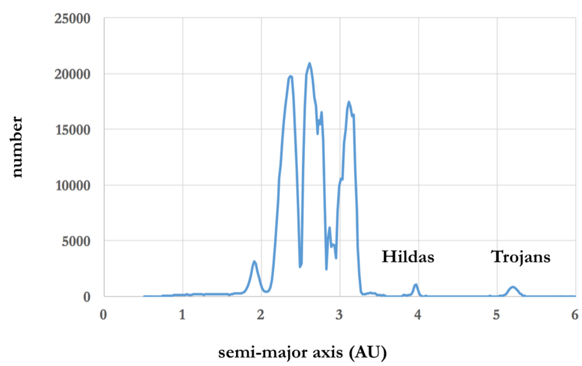


Figure 2.37: Distribution of minor planets in the inner solar system.
Data from the International Astronomical Union's Minor Planet Center; http://www.minorplanetcenter.net/data

Dynamical effects of solar wind and radiation on molecules and small particles

Objects smaller than a kilometer or so are subject to interactions with particles and photons from the Sun in ways that larger, more massive objects aren't.

The solar wind is the term for the sometimes steady, sometimes explosive, outflow of charged particles from the Sun. It is most dramatically apparent in large coronal mass ejections (CME), when a solar prominence disconnects and is blown away from the Sun. These are often associated with bright flares near the Sun's photosphere (the layer from which the majority of the light is emitted). The following set of images from the Solar Dynamics Observatory, a space telescope in orbit at the L1 point, shows a CME directed sideways, not toward the Earth. When eruptions do target Earth, we are likely to observe increased auroral activity ("northern lights").

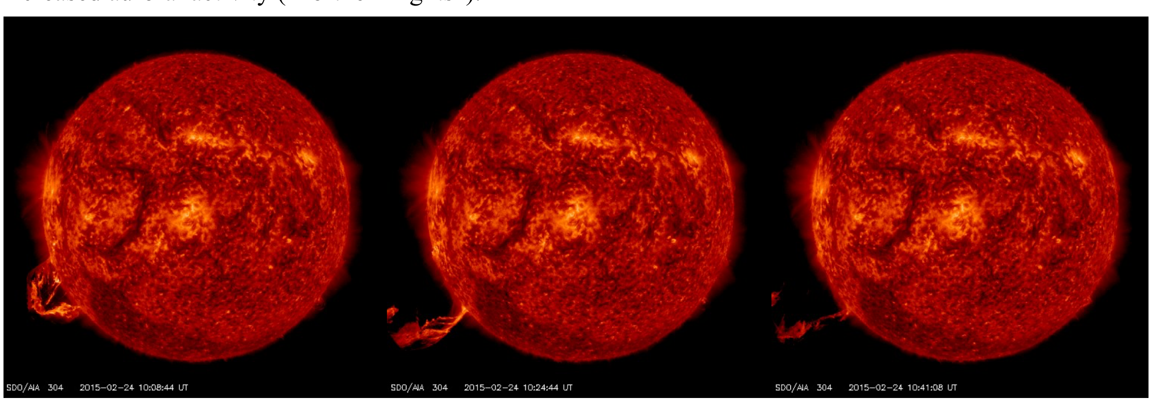


Figure 2.38: Coronal mass ejection; courtesy of NASA/SDO and the AIA, EVE, and HMI science teams.

This eruption occurred on February 24th, 2015, and is shown here in UV light (304 Å) emitted by singly ionized helium atoms. The solar wind is mostly lightweight particles – electrons, protons, alpha particles (helium nuclei) – with energies and densities that vary depending on how energetically the particles were ejected and with distance from the Sun, as well as with where we are in the solar activity cycle. The plasma typically has between 2-3 and 20-30 particles per cm^3, moving with speeds of a few hundred km/sec near Earth. In the absence of a planetary magnetic field, high-energy particles in the solar wind plow straight into the surface of the planet. That's one of the concerns that make long-term human exploration of the Moon or Mars dangerous.

Comet tails are the prime example of gas interacting with the solar wind, noted in our earlier tour of the solar system in the overview chapter. To recapitulate what happens with the ion tail of a comet, solar ultraviolet radiation ionizes some of the gas molecules in the coma and those now charged particles interact with the outflowing solar wind plasma and are blown away from the Sun.

Gas molecules are on the order of a few nanometers in size. Particles that are small but one-two orders of magnitude larger than gas molecules will get blow outward by the radiation pressure of the Sun. Photons have momentum. For an individual photon we can express that momentum as

$$p = E / c = hf / c,,$$

where h is Planck's constant and f (or v) is the frequency of the light. How do we get from momentum to radiation pressure?

1) Force = change in momentum per unit time or $F = dp / dt$.
2) Pressure = force per unit area or $P = F / A$.

Therefore the pressure exerted by photons hitting a surface is equal to the change in the photons' momenta as they hit that surface. How much momentum is transferred to a surface depends on how reflective the surface is – totally reflected light, meaning the photons weren't just stopped but were turned around, transfers twice as much momentum to a surface as we would get in the case of totally absorbed light. Still, we can say for sunlight that the force exerted on a small particle due to the radiation pressure should be proportional to the flux of sunlight times the cross-sectional area of the particle, or

$$F_{\text{rad}} = Q \frac{L_{\text{Sun}}}{4\pi d^2 c} \pi r^2 ,$$

where L_{Sun} is the solar luminosity, in J / sec, and r is the radius of the particle. Note that luminosity is J / sec and flux is J / (sec m^2). Dust grains are fairly dark, so they are likely to absorb most of the incident sunlight. The also are not likely to be nice neat spherical particles, or equally absorbing at all wavelengths, meaning that their effective cross-section for absorbing sunlight is often less than πr^2, perhaps only ½ that much; that's the purpose of the factor Q in the above equation. Will dust particles be stable or will they be pushed outward by the radiation pressure? Set the force due to the radiation equal to the force on the particle due to the gravity from the Sun:

$$Q \frac{L_{\text{Sun}}}{4\pi d^2 c} \pi r^2 = \frac{GM_{\text{Sun}} (\frac{4}{3}\pi r^3 \rho)}{d^2} .$$

Here we are assuming that our particle is roughly spherical for purposes of determining its mass; if it's rock, the density ρ should be about 3,000 kg/m^3. Assume that $Q \sim$ ½. Note the helpful cancellation of the distances – both gravity and sunlight received fall off as the square of the distance. Look up the solar luminosity, mass, gravitational constant, cancel a factor of π r2 from both sides, and solve for the particle size:

$$\frac{1}{2} \cdot \frac{3.85 \cdot 10^{26} \text{ J/s}}{4\pi \cdot 3 \cdot 10^8 \text{ m/s}} = \frac{4 \cdot 6.67 \cdot 10^{-11} \text{kg} \cdot \text{m/s}^2 \cdot 1.99 \cdot 10^{30} \text{kg} \cdot 3{,}000 \text{ kg/m}^3}{3} r$$

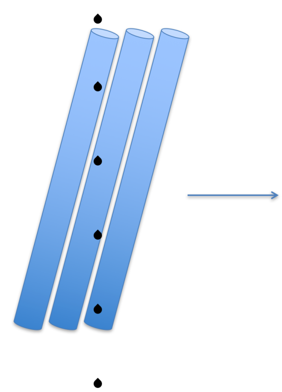
or $r \sim 10^{-7}$ m. In words, particles smaller than about 1/10 micron will be blown outward by solar radiation pressure. The fact that there is still dust in the inner part of the solar system suggests that collisions continue to happen, and new small dust particles continue to be produced.

Radiation pressure doesn't always act to push particles outward, away from the Sun. It can make them spiral *inward*. The tiny particles we considered above were too small to have a leading side / trailing side distinction, but particles on the order of a centimeter or so do. In other words, with micron-sized particles we don't have to consider what direction the sunlight's coming from; with cm-sized particles that matters. In the 1600s astronomers were trying, without yet having success, to measure the parallax of stars. In the new heliocentric cosmos in which the Earth moved in orbit around the Sun, relatively nearer stars should shift back and forth over the year with respect to more distant stars as we viewed them from one side of the Earth's orbit and then the other. What they found first, instead, was the "aberration of starlight", due to the Earth moving into the starlight "falling" at the finite speed of light. As a classic example, consider standing in the rain holding an open length of pipe. Assuming that the wind isn't blowing too strongly, the raindrops will fall straight down the pipe if you hold the pipe vertical. Now start walking forward.

As you move forward, drops that enter the top of the pipe will hit the sides unless you tilt the pipe slightly. The faster you are moving, the more you have to tilt the pipe. Unlike the raindrops, the starlight isn't always coming at us from a direction that's perpendicular to the direction we are moving. But the picture is still pretty similar; i.e., starlight is "falling" at $3{\cdot}10^5$ km/sec, Earth is moving into the starlight at 30 km/sec, and the starlight preferentially hits us from the front.

Figure 2.39 a

To get the starlight to fall straight through a telescope tube, you have to tilt the tube. How much depends on where in the sky the star is; the maximum shift is $\sim \pm 20''$. (That's the $\tan^{-1}$ of the speed ratio, 30 km/s / $3{\cdot}10^5$ km/s.)



Aberration of starlight

Figure 2.39 b

What does this have to do with sunlight hitting cm-sized bits of rock in our solar system? The sunlight is like the falling rain or the incoming starlight in that it will preferentially hit the leading side of small particles. The momentum those particles receive from the sunlight will have a component that is directed oppositely to the particle's motion. The small bits of rock will lose some energy, fall into lower orbits around the Sun and eventually spiral in to the Sun. This is called the Poynting-Robertson effect, first predicted in the early 1900s by Poynting and amended 30 years later by Robertson to take relativistic effects into account. How long it takes for particles to be swept up depends on the size of the particle, where in the solar system it was to begin with, and also how eccentric its initial orbit. Here is a sketch, not to scale!, of a particle in orbit around the Sun, preferentially hitting photons of sunlight on its leading side:

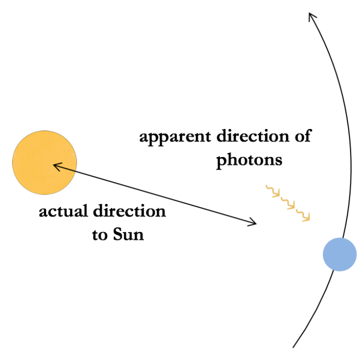


Figure 2.40:
Poynting-Robertson effect

Moving up another step in size, particles of a few to a few tens or hundres of meters are large enough to have distinct daytime and nighttime sides. Rocks this size are warmed by the sunlight on the daytime side; assuming that they rotate in hours (not months), this warm spot will soon be carried around to the evening side and, being distinctly warmer than the dark sky the spot is now seeing, the spot will radiate. Warm rock is close to being a blackbody and the spot will radiate energy at a rate that is proportional to T^4, meaning that objects with lower albedos and/or closer to the Sun will radiate more. Photons have momentum, so the photons radiated away from this chunk of space rock, carrying momentum with them, will act as a kind of jet, applying a thrust to the rock. Whether that thrust is forward or backward depends on which direction the rock is rotating. Most solar system objects rotate prograde, but they don't all. Sometimes, as a result of a collision, a small object's direction of rotation could change, and, as a result, its direction of migration in the solar system, inward or outward, could change. This is called the Yarkovsky effect and it plays a role in modifying the orbits of meteoroids. A meteoroid might start out in the asteroid belt, perhaps as a fragment of an asteroid broken off in a collision; the Yarkovsky effect could modify the meteoroid's orbit enough that a gravitational interaction with Jupiter could perturb the meteoroid's orbit enough to set it on a collision course with Earth. A small asymmetric asteroid could also experience a modification of its rotation rate: if it radiates those photons away from a couple of different patches of its surface that face in roughly opposite directions, then the object will experience a torque which could – over moderately long periods of time, something on the order of hundreds of thousands of years – spin the object up to the point where it could fragment. This is called the YORP effect (Yarkovsky plus three more folks involved in modeling this process); there's even a 100-ish meter-sized asteroid named for it: 54509 YORP has a rapid and increasing rotation rate. Observations of asteroid 6478 Gault (~ 3.7 km diameter) between 2013 and 2019 show episodes of comet-like behavior in which Gault is losing dust and bits of rock. Its rotation speed is roughly 2 hours, which is about as fast as it should be able to rotate without breaking apart. It's possible that the YORP effect plays a role in the formation of small binary asteroids through the break-up of their parent objects.

The effects we have been considering – solar wind, radiation pressure, the Poynting-Robertson effect, the Yarkovsky and YORP effects – are all due to interactions between small objects and the Sun. Gravitational perturbations matter, too, and can have large effects on large objects. When we get to the formation of the solar system we will consider current models in which most of the major planets migrated, in some cases considerably, from their birthplaces to their current locations.

Moment of inertia and real gravitational fields

As a prelude to the chapter on terrestrial planet interiors, let's consider the way mass is distributed inside a moon or planet and how we can measure that distribution. First, let's look at the moment of inertia, which is related to angular momentum. The angular momentum of an object is the linear momentum times the radius from a pivot point. For a moon or planet, imagine coordinate axes *x*-*y*-*z*, centered on the object's center, with *z* being the axis around which the object rotates. The angular momentum, *L*, of the object can be expressed as the sum of the angular momenta of all the particles of which the object is composed. The total angular momentum will only have a *z*-component if the particles aren't rotating around either of the other two axes. Let r_i be the distance from the axis of the ith particle, having mass mi, in the object. The total angular momentum is given by

$$L = L_z = \sum_i r_i m_i v_i = \sum_i r_i^2 m_i \omega_i \ ,$$

where we've changed from linear velocity to the angular velocity, $\omega = r\, v$. For solid objects, the angular velocity is the same for all the particles that make up the object. We can't make that assumption for the Sun or the giant planets, which exhibit differential rotation, where different depths and/or different latitudes rotate at different rates.

Define the moment of inertia, I, as

$$I \equiv \sum_i m_i r_i^2 .$$

This means that we can write the angular momentum as

$$L = I\omega .$$

In comparison with linear motion, where kinetic energy is ½ mv^2, in the case of angular motion, the kinetic energy can be written

$$E_{\text{rotation}} = \tfrac{1}{2} I \omega^2{}_{\text{rotation}} .$$

For a spherical object of mass M and radius R in which the mass is uniformly distributed, we find that

$$I_{\text{uniform}} = 0.4 MR^2 .$$

For purposes of determining gravitational forces we often treat objects as if all the mass were at the center; if that were really the case, then $I_{\text{center}} = 0$. If we could determine the moment of inertia of a moon or planet we could tell whether its mass is relatively more uniformly distributed – I closer to 0.4 $M\,R^2$ – or relatively more centrally condensed – I closer to 0.

Math note: If you've had some calculus, you may have seen the moment of inertia and angular momentum expressed in integral form. It makes sense to use integrals because planets can be treated as if their mass density is smoothly distributed. Here we can write

$$I \equiv \iiint \rho(\vec{r}) r_z^2 d\vec{r} \text{ and}$$

$$\vec{L} = \iiint \rho(\vec{r}) r_z^2 \vec{\omega}_{\text{rotation}} d\vec{r} = I\vec{\omega}_{\text{rotation}} ,$$

where we are being explicit about the fact that the angular momentum and rotation velocity are vectors, having both magnitude and direction.

Before we consider how to tell how the mass is distributed inside a moon or planet, it would help to consider how to describe the way(s) such an object can deviate from spherically symmetric. For instance, neither of the following two objects is round, but they are not-round in different ways. If we spin each around the vertical axis, on the left we get an egg while on the right we get an *oblate* spheroid.

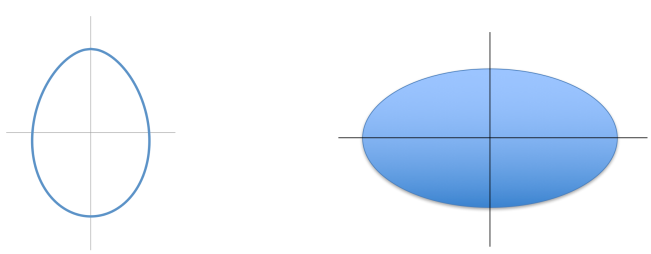


Figure 2.41:
Egg vs. oblate spheroid

(A *prolate* spheroid is the term for an object that is elongated symmetrically along the *z* axis, like a cigar standing on end.) Even if these two objects had the same total mass and that mass were uniformly distributed, it's clear that a small test mass near the equator would feel a different force of gravity on the egg compared to the oblate spheroid. Mass is often not uniformly distributed and a test mass can feel a fairly complicated gravitational force. It can make more sense to describe the gravitational field by the potential energy, E_{Grav}, equal to $-GMm/r$, (or the potential, V, which is the energy per unit mass, $-GM/r$) of the test mass.

Often, not just in this context, we would like to be able to describe a complicated function as a sum of simple functions. You may already be familiar with Fourier series, which use a sum of sines and cosines to describe a complicated pattern of frequency vs. time. An example would be the tones of a musical instrument – "one" note is usually not just the fundamental frequency but also includes higher harmonics. You may or may not have encountered spherical harmonics, which use a more complicated sum of functions to describe complicated patterns on spheres. In terms of the gravitational field, the fundamental, or zeroth harmonic, is the spherically symmetric term we get from assuming that all of a planet's mass is concentrated at the center. The oblate spheroid, above, although a bit exaggerated, is a relevant model for planets such as Jupiter or Saturn, which are largely fluid and which spin fast enough to bulge out along the equator. "Solid" objects such as Earth are less, but still a bit, oblate. Describing being squished in at both poles requires the second harmonic. (An unrealistically exaggerated egg, or maybe a water balloon being squeezed at one end and bulging out at the other, i.e., squished at one pole only, would be described by the first harmonic). The second harmonic is usually denoted "J_2". Using spherical harmonics with an oblate planet we could express the gravitational potential as

$$V(r) = -\frac{GM}{r}\left[1 - \left(\frac{a}{r}\right)^2 J_2 P_2(\cos\theta)\right],$$

where a is the planet's equatorial radius, J_2 describes how oblate the specific planet is, and the expression $P_2(\cos\theta)$ is called the second Legendre polynomial. That's the spherical harmonic function. $P_2(\cos\theta)$ is not too complicated. It's = ½(3 $\cos^2\theta - 1$). The expressions for the higher order harmonics, which we might want to describe more complicated deviations of mass distributions from spherically symmetric, do get a bit messy.

Now let's get back to the question of how to tell how the mass is distributed inside a planet. One way is to look at how large J_2 is, or the, related, relative difference between the equatorial and polar radii. Consider the following two shapes, and imagine rotating them around the z axis:

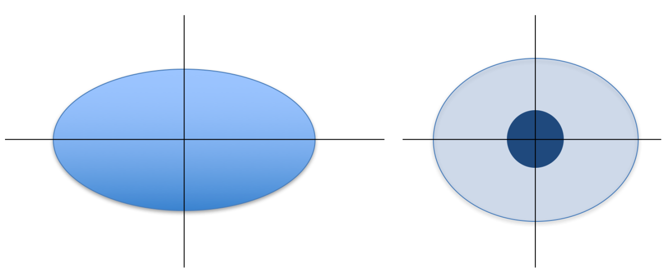


Figure 2.42:
Oblate spheroids, varying mass distributions

A rotating object in which the mass is more uniformly distributed will bulge out more at the equator than one that is more centrally condensed. If we have good enough measurements we may be able to see this effect by measuring the ratio of radii $f = (a - c)/a$, where a is the equatorial radius and c is the polar radius. For example, Saturn's equatorial radius is 60,268 km and its polar axis is only 54,364 km; for Saturn, the ratio $f = 0.098$, or nearly 10%. Given a specified rotation rate for a planet, you should be able to see that it should be possible to predict how oblate a planet would be for different possible mass distributions. In other words, there's a relationship between the oblateness ratio f , the second harmonic term J_2, and the planet's rotation rate. With a bit of math, which we are not going to do here, it can be shown that

$$f = \frac{3}{2}J_2 + \frac{a^3\omega^2}{2GM},$$

where ω is the angular speed of the planet. You can probably anticipate that J_2 is related to the relative moments of inertia around different axes through the planet. A bit more math gives the following relationship between J_2 and these moments of inertia:

$$J_2 = \frac{C - A}{Ma^2}.$$

C is the moment of inertial measured from the spin axis and A is the moment of inertia measured from an axis going through the equator. In other words, putting all these pieces together, observations of the oblateness and the rotation speed give us an indication of how centrally concentrated the mass inside a planet is.

We may also be able to detect this difference in mass distribution by the effect it has on a spacecraft passing a known distance from the planet's center of mass. More mass out further from the rotation axis will accelerate the spacecraft more. Actually going to the planet or moon and measuring the acceleration on a spacecraft helps improve the estimations of mass distribution derived from Earth-based observations.

Real moons and planets have complicated mass distributions that require using higher harmonics in addition to J_2. The GRACE mission (Gravity Recovery And Climate Experiment), a joint project of NASA and the German Aerospace Center (DLR), is a pair of satellites launched into low Earth orbit in 2002 designed to measure the Earth's gravity field and how it varies with time. It is able, for instance, to measure the mass loss in the ice sheets in Greenland and Antarctica. The following figure shows GRACE measurements of the recent average gravity field anomalies of Earth; in other words, how much does the gravity field differ from a perfectly smooth ellipsoid. The color scale shows the range of gravitational acceleration (around the average 9.8 m/s2) in units of mGal; 1 Gal = 1 cm/s2. This unit is named for Galileo.

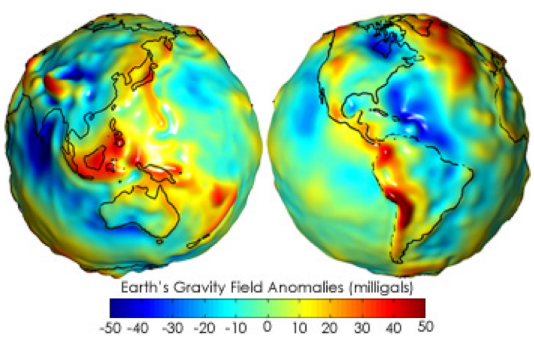


Figure 2.43: GRACE observations of Earth's gravity field anomalies

Credit: NASA Earth Observatory
http://earthobservatory.nasa.gov/Features/GRACE/page3.php

Why do we care about the distribution of mass? For the Earth, events such as major subduction zone earthquakes or the melting of ice sheets, mentioned above, change the distribution of mass within the planet. For planetary bodies whose surfaces we haven't or can't visit, the distribution of mass, coupled with the heat flow from the surface, provide the major clues to the structure of the moon or planet. As an example, in the overview chapter you read that some of the moons in the outer solar system appear to have subsurface oceans. One of the pieces of evidence that supports the idea of a subsurface ocean is measuring the distribution of mass inside the moon. Our Moon has an interesting distribution of mass: the center of its geometric figure isn't at the same location as its center of mass. Why not? What does this tell us about the history of the Moon? In chapter 5 we will examine the interiors of moons and terrestrial planets.

Sample problems.

1. Pluto and Charon are separated by 19,571 km; their masses are $1.303 \cdot 10^{22}$ and $1.586 \cdot 10^{21}$ kg, respectively. Pluto's diameter is 2,377 km. Verify that the center of mass between Pluto and Charon is not inside Pluto. Hint: $m_1 r_1 = m_2 r_2$ and $r_1 + r_2 = r$; start by eliminating r_2.

2. Calculate the synodic period of Jupiter as seen from Earth. The relevant equation is: $\left|\frac{1}{P_1} - \frac{1}{P_2}\right| = \frac{1}{S}$.

In this equation, P_1 and P_2 are two sidereal periods, i.e., periods with respect to the stars, and $1/S$ is the "beat frequency"; S itself is the synodic period, the time it takes for three objects (e.g., here, Sun – Earth – Jupiter) to come back into alignment. Jupiter's orbital period is 11.86 years.

3. Harder problem: Consider a solar eclipse where the Moon is at its closest perigee distance (~356,400 km), going straight across the center of the disk of the Sun (the Sun's radius is ~ 696,000 km), and casting its shadow on the Earth's equator; let the Earth be at its aphelion distance from the Sun (~152,100,000 km). Assume that the surface of the Earth where the shadow hits is flat and that the Sun and Moon are directly overhead. Estimate how long totality will last. Hints: This is a multi-step problem; some things you might want to consider are: How far is the Moon from the Earth and how fast is it moving? What is the Earth's rotation period and what is the ground speed of the Earth's equator? How large is the shadow of the Moon at the surface of the Earth? Sketching a side-on view and identifying various similar triangles might be very helpful. Another hint: don't assume that the vertex of the shadow cone is any place special, such as the surface or center of the Earth.

4. Same set up as in problem 3 except let the Moon be at its average apogee distance (~405,400 km). Where will the tip of the shadow cone be with respect to the surface of the Earth?

5. Given two solar eclipses with the same diameter shadow cone, would totality last longer at the equator or in Seattle? Hint: think about the ground speed of the Earth.

6. As noted above, asteroid 153 Hilda has an orbital semi-major axis of 3.973 AU and an eccentricity of 0.141; the Hilda family of asteroids are in a 3:2 resonance with Jupiter (a = 5.20 AU). Calculate the ratio of the orbit period of Hilda to that of Jupiter. Hints: this is a Kepler's third law question and ratios are our friends — you don't need to calculate the period of either object to answer this question.

7. Suppose that you wanted to get a spacecraft from an asteroid at 2.3 AU from the Sun to a second asteroid 3.5 AU from the Sun on a least-energy orbit.
   a) How long would it take the spacecraft to get from one asteroid to the second?
   b) Sketch the orbits of the two asteroids and the spacecraft.

8. Venus revolves around the Sun in 0.615 years and has a rotation period = −243 days. How long is the solar day on Venus? Hint: don't forget the minus sign on the rotation period.

9. Suppose Mercury is at aphelion and the Earth at its average distance from the Sun *and* that Mercury is at greatest elongation. How many degrees is it away from the Sun? You can look up Mercury's aphelion but you should convince yourself that you can calculate it from a and e.

10. Reading carefully?
   a) How is it that there could be 5 solar eclipses in one calendar year?
   b) If it's First Quarter Moon today, what phase will the Moon be tomorrow?

c) Why do some of the gaps in Saturn's rings look bright when viewed from behind?
d) What is an ascending node?
e) What is an oblate spheroid?
f) What is the aberration of starlight?
g) Is a comet orbit with an inclination of 135° retrograde or prograde?

Answers to selected questions are on the next page:

1. The distance from the center of Pluto to the center of mass = r_1 = 2124 km, which is more than the radius of Pluto. Thus the center of mass is out in space between the two objects, not inside Pluto.

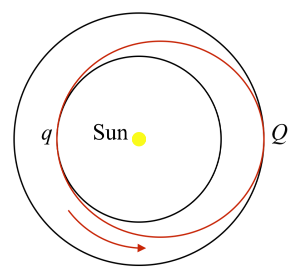
2. 13 months.

3. Moon speed is ~ 1.1 km/sec; Earth ground speed at equator is ~ 0.465 km/sec. The shadow diameter at the Earth's surface is ~270 km across. Totality thus lasts ~ 7 minutes.

4. This eclipse would definitely be annular.

5. Longer at the equator.

7. a) 2.5 years. b)



8. 117 days

9. 25 degrees