THE GREAT COURSES®

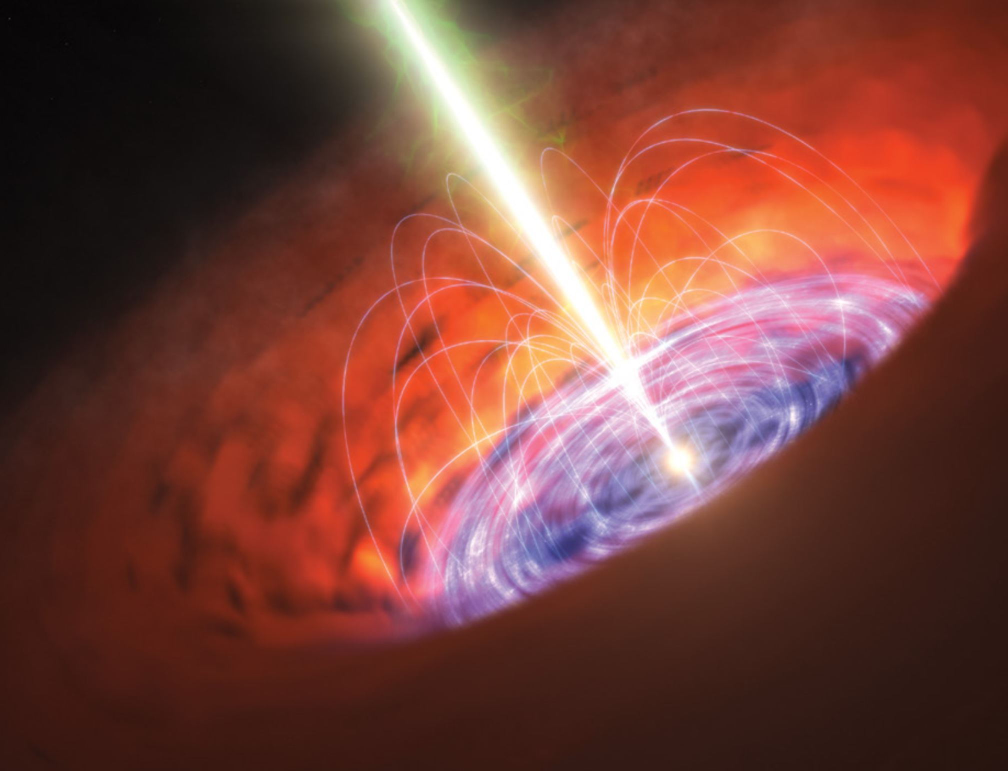
Topic
Science & Mathematics

Subtopic
Astronomy

Introduction to Astrophysics

Course Guidebook

Professor Joshua N. Winn
Princeton University



Published by

THE GREAT COURSES

Corporate Headquarters

4840 Westfields Boulevard | Suite 500 | Chantilly, Virginia | 20151-2299

[PHONE] 1.800.832.2412 | [FAX] 703.378.3819 | [WEB] www.thegreatcourses.com

Printed in the United States of America

JOSHUA N. WINN, PHD

Professor of Astrophysical Sciences
Princeton University

Joshua N. Winn is a Professor of Astrophysical Sciences at Princeton University. He studied physics as an undergraduate at the Massachusetts Institute of Technology (MIT), spent a year as a Fulbright Scholar at Cambridge University, and then returned to MIT to pursue a doctorate. While in graduate school, Professor Winn worked in medical physics, condensed matter physics, and astrophysics and wrote for the science section of *The Economist*. After earning his PhD, he held fellowships from the National Science Foundation and NASA at the Harvard-Smithsonian Center for Astrophysics. He was on the MIT Department of Physics faculty for 10 years before joining the faculty of Princeton.

Professor Winn's research goals are to explore the properties of planets around other stars, understand how planets form and evolve, and make progress on the age-old question of whether there are other planets capable of supporting life. His research group uses optical and infrared telescopes to study exoplanetary systems, especially those in which the star and planet eclipse each other.

Professor Winn was a participating scientist in NASA's Kepler mission, which had as its goal the detection of earthlike planets, and he is a coinvestigator in another NASA mission called the Transiting Exoplanet Survey Satellite, which was launched in 2018. He has authored or coauthored approximately 200 scientific articles, mainly on the topic of exoplanetary science. Over the years, Professor Winn and his research group have pursued topics in stellar astronomy, planetary dynamics, radio interferometry, gravitational lensing, and photonic bandgap materials.

Professor Winn has taught numerous subjects in physics and astronomy to both undergraduate and graduate students. He has received the MIT Buechner Faculty Teaching Prize and the MIT School of Science Graduate Teaching Prize for excellence in teaching.

Professor Winn's other Great Course is *The Search for Exoplanets: What Astronomers Know*. ★

TABLE OF CONTENTS

INTRODUCTION

GUIDES

SUPPLEMENTARY MATERIAL

To go back to the page you came from, **PRESS Alt + ←** on a PC or **⌘ + ←** on a Mac. On a tablet, use the bookmarks panel.

INTRODUCTION TO **ASTROPHYSICS**

Astronomy is a fascinating subject because the universe is full of such wonders as black holes, exploding stars, and colliding galaxies and because new discoveries are being made at a rapid pace. While it is possible to appreciate astronomy with images and qualitative descriptions, the goal of this course is to gain access to the deeper level of beauty and understanding that astrophysics—the application of the laws of physics to comprehend celestial phenomena—can provide. This not only gives a greater appreciation for the wonders of the universe, but also allows for the perception of hidden regularities and connections between phenomena. For example, the relationship between a white dwarf and a neutron star is analogous to the relationship between an atom and an atomic nucleus.

This course conveys the quantitative foundations of astrophysics with the hope of both satisfying and stimulating curiosity about the subject. The most important prerequisites are the desire to understand deeply, the capacity and patience for learning new things, and a sense of wonder. To gain the most from the course, a good background in freshman-level classical mechanics and calculus is needed; logarithms, trigonometry, and vectors are employed throughout. After completing this course, you will have a firmer grip on the universe and an enhanced ability to solve problems in the physical sciences.

Astrophysics spans more orders of magnitude in space, time, and mass than any other science, and the first 3 lectures of this course provide a unifying structure to help comprehend the vast range of scales, based on orders of magnitude and logarithmic charts. The first lecture zooms out from spatial scales of human beings to the entire observable universe. The second lecture zooms in to the fundamental particles and the 4 fundamental forces of nature. The third lecture is about how astrophysicists establish the locations of objects in 3 dimensions, despite being stuck in an arbitrary location within just one of countless galaxies.

The next topic is gravity. Lectures 4 through 7 apply the law of gravity to understand the motion of planets, the destructive power of tidal forces, and the existence of black holes. A feature of this section is a detailed examination of the relationship between Kepler's laws of planetary motion and Newton's laws of motion and gravity, a topic usually reserved for more advanced courses.

Attention then turns to a different force of nature: electromagnetism. Lecture 8 is on photons, the basic unit of electromagnetic radiation. The properties of photons are compared with those of ordinary particles, and many important formulas are introduced. Lecture 9 provides an immediate application of these concepts to understand the basic properties of the planets in the solar system. The next 3 lectures are about telescopes. They discuss the fundamental purpose of telescopes and the differences between telescopes in the radio, optical, and x-ray domains of the electromagnetic spectrum. Lecture 12 takes a deep dive into spectroscopy, the main way to learn about the physical conditions of a star, planet, nebula, or galaxy.

Then begins a sequence of lectures about stars and their planets in which the following questions are addressed: How are the properties of stars determined? What has been discovered about planetary systems around other stars? Why do stars shine, and how long do they last? What are the conditions like at the center of a star? What happens when a star runs out of fuel? How can the existence of stars that are millions of times denser than the Earth be explained?

After such detail is spent on understanding stars, they are destroyed in lecture 19, which is about supernovas and their causes. Then comes a highlight of the course, in lecture 20, about gravitational waves. The seminal first detection of colliding black holes is examined in detail, starting with the original data and culminating in a calculation of the masses of the black holes and their distance from Earth. Even as recently as 2015, this lecture could not have been written.

The last 4 lectures zoom out to gain a perspective on galaxies and the universe as a whole. Lecture 21 not only features dazzling images of galaxies—the orchids of the universe—but also introduces the mind-bending astrophysical concept of the galaxy as a collisionless fluid of stars. The topic of galaxies is developed further in lecture 22, including active galaxies, in which material is funneling into a central black hole, and the mystery of dark matter. Finally, lectures 23 and 24 present the quantitative basis of the modern creation story: the big bang. The course ends at the frontier of astrophysics and particle physics, with the discovery of what may turn out to be an entirely new force of nature. ★

Lecture 1

ZOOMING OUT TO DISTANT GALAXIES

The words "astrophysics" and "astronomy" are basically interchangeable these days, but there is a subtle intellectual distinction. Astrophysics is the application of the laws of physics to understand celestial phenomena. Occasionally, we even discover a new law of physics by studying what's out there. In contrast, astronomy can be defined as the careful observation of heavenly bodies—a cultural activity dating back thousands of years that only gradually became scientific. The ancient Babylonians, the Chinese, and the Mayans were all accomplished astronomers, but they weren't astrophysicists. Compared to astrophysics, no other science spans such a vast range of scales—from nanometers to billions of light-years and from the radiation of a single electron to the output of trillions of suns.

Astrophysics began in the 17th century with Isaac Newton, who explained the motions of the planets with his shiny new equations relating force, mass, acceleration, and gravity.

The actual word "astrophysics" is more recent. It's from the mid-19th century, after the invention of photography and spectroscopy. These techniques allowed us to go beyond looking through telescopes with our eyes; now we could make more objective records, detect fainter sources, and connect our observations to laboratory experiments with light, heat, and atoms.

PUTTING THE UNIVERSE INTO PERSPECTIVE

- It's difficult to put the whole universe into perspective. Even if we scale everything down by a factor of a billion, the nearest star to the Sun would be 25,000 miles away, and our next-door neighbor galaxy would be 20 billion miles away. No matter how much we try to scale things down to a manageable size, we still get mind-boggling numbers. The problem is there's no one scale factor that will put all the phenomena from the cosmos to the microworld into a mentally comprehensible map.

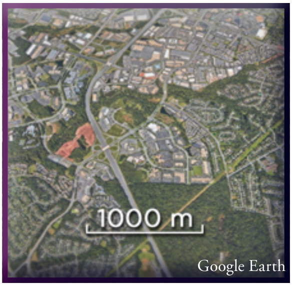
- If we start with a map of a building near Washington DC, for example, and a scale bar representing 100 meters, then when we expand our field of view by a factor of 10—making the scale bar 1000 meters, or 1 kilometer—we can take in the whole city.

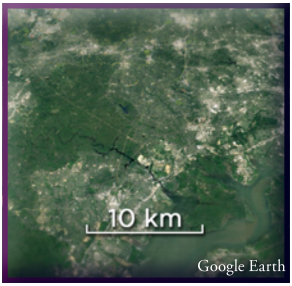
- If we expand our field of view by another factor of 10, we start to see regional features, such as the Chesapeake Bay. And by taking another step, the scale bar becomes 100 kilometers, and we can see the entire mid-Atlantic Seaboard.

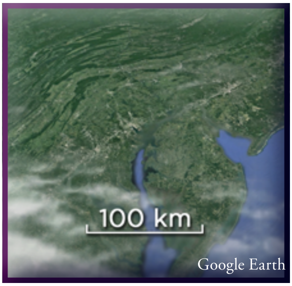
- Expand another factor of 10 and we can see the entire Earth, hanging in empty space. At this point, we've zoomed out from hundreds to millions of meters.

- This brings up the issue of units of measurement. The standard metric unit for length is meters, including millimeters, kilometers, and so on. The scale bar we have just used is 1 million meters long, or 1 megameter. Another way to write that is with scientific notation: 10^6 meters, because 10 to the sixth power (1,000,000) is 1 million.

- But when we're thinking about entire planets, meters are not very convenient; it's better to measure things in units of the radius of the Earth. One Earth radius is defined as 6378 kilometers. That way, we can say that the planet Neptune has a radius of about 4 Earth radii, and Jupiter's is about 11. These numbers are much easier to comprehend than however many millions of meters. That's why the Earth radius is a handy unit; it's written as $R_{\oplus}$, where $\oplus$ is the astronomical symbol for the Earth.







- The next useful unit that we'll need is for the size of stars. The Sun's radius is about 700 million meters, or a little more than 100 times bigger than Earth. The solar radius, the unit of choice when dealing with stars, is written as $R_\odot$, where $\odot$ is the astronomical symbol for the Sun.

- Once we get to a scale of 10^{13} meters, most of the other planets come into view. We've reached the scale of planetary systems, for which the traditional unit is the radius of Earth's orbit around the Sun—a unit called the astronomical unit (AU). It's about 215 solar radii, or 150 billion meters. With the astronomical unit, the solar system can easily be described. Mercury is about 2/5 of an AU from the Sun; Jupiter is out at 5.2 AU.

- When we expand the scale again, beyond the solar system, we find ourselves in empty space for quite a while, until we get to 10^{16} meters, at which point some of the neighboring stars come into view. A good unit to use on this scale is the light-year, or the distance light travels in 1 year, which is just short of 10^{16} meters. For example, the nearest star, Proxima Centauri, is 4.2 light-years away.

- In practice, astrophysicists don't use light-years. Instead, the preferred unit is called the parsec, and it's about 3.3 light-years. The typical distance between stars is 1 or 2 parsecs.

- From here, we need to zoom out 4 more factors of 10—4 more orders of magnitude—until the architecture of the Milky Way Galaxy comes into view, at around 10^{20} meters. At this stage, we just keep using parsecs, but with metric prefixes, such as "kilo-" for 1000. The diameter of a typical spiral galaxy is 10 or 20 kiloparsecs.

- It takes a few more orders of magnitude to start seeing neighboring galaxies. The typical spacing between galaxies is a few megaparsecs, or millions of parsecs.

- After another step, the galaxies group together to form clusters of galaxies, joined by what look like filaments, or webs of galaxies. And when we keep increasing the scale bar all the way to 10^{26} meters, the universe starts to look like random static, with nowhere different from anywhere else. The natural scale at this stage is the gigaparsec, or billions of parsecs.

- That's the end of the line—the largest spatial scales about which we have any direct knowledge. By zooming out 26 orders of magnitude, we have a view of the entire observable universe.

SCALE	UNIT	EXPRESSED IN METERS	EXPRESSED IN THE PREVIOUS UNIT
Planets	$R_\oplus$	6.4×10^6 m	—
Stars	$R_\odot$	7.0×10^8 m	109 $R_\oplus$
Planetary Systems	AU	1.5×10^{11} m	215 $R_\odot$
Between Stars	pc	3.0×10^{16} m	206,265 AU
Galaxies	kpc	3.0×10^{19} m	1000 pc
Between Galaxies	Mpc	3.0×10^{22} m	1000 kpc
Observable Universe	Gpc	3.0×10^{25} m	1000 Mpc

LOGARITHMIC MAPS AND CHARTS

- Another tactic that astrophysicists use to cope with all of these orders of magnitude is by making logarithmic maps. Taking the logarithm of a number means expressing the number as a power of 10 and then plucking out the exponent. For example, 1000 is 10 to the third power, written as 10^3, so the logarithm of 1000 is 3. The log of 1 million is 6.

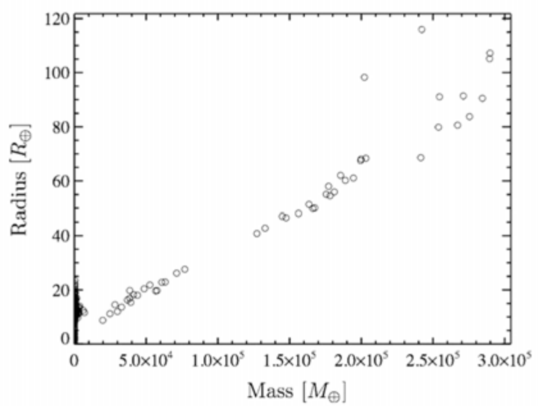
- This also works for numbers smaller than 10. The number 1 is equal to 10 to the 0^{th} power, written as 10^0, so the log of 1 is 0; 1/10 is 10 to the –1 power, written as 10^{-1}, so the log of 1/10 is –1; and so on.

- A logarithmic map is an ordinary map based on a single scale factor. For example, 1 inch on the map might be 1 kilometer in real life. But on a logarithmic map, the scale factor changes when moving from one end to the other. The first inch might correspond to 1 meter in real life, but then the second inch is 10 meters, then 100 meters, 1000 meters, then 10^4, then 10^5, and so on. Mathematically, with every inch, the logarithm of the scale factor is increased by 1 unit.

- Besides maps, there are other logarithmic charts, such as logarithmic time lines as well as more abstract logarithmic charts that help make sense of things that range over many orders of magnitude.

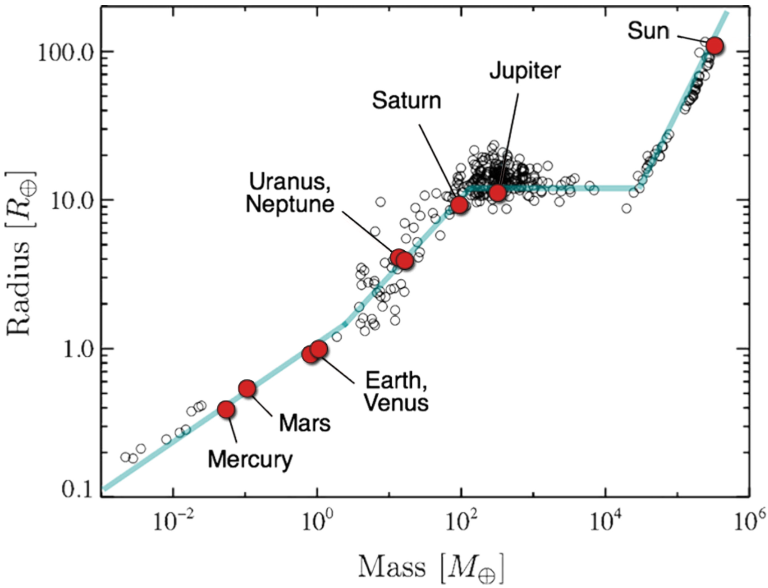
- For example, our galaxy is full of objects ranging widely in mass and size. Among other things, there are asteroids, moons, planets, and stars. Let's say that we go around our galaxy and measure the mass and radius of everything smaller than the Sun. To compare all these things, we can make a chart of mass versus radius, with mass on the horizontal axis and radius on the vertical axis.

- Each data point shows the mass (in units of Earth masses) and radius (in units of Earth radii) of a single object. This chart shows a relationship between radius and mass: The more massive the object, the bigger the radius, which makes sense.



- But there's a problem with this chart. Because we need to make the axes range high enough to encompass the very largest objects—millions of Earth masses—the more numerous, smaller objects end up crammed in close to the origin and the details are difficult to see.

- If we remake this chart with logarithmic axes, the horizontal axis is still telling us the mass, but now each tick mark represents a factor of 10. Likewise, the vertical axis still tells us the radius, but on a logarithmic scale.



- This logarithmic chart shows all the data clearly and, even better, some patterns that were hidden in the ordinary chart. There are 4 different groups, differing in the relationship between mass and radius. In each of these 4 zones, we can fit the data, at least approximately, with a straight line.

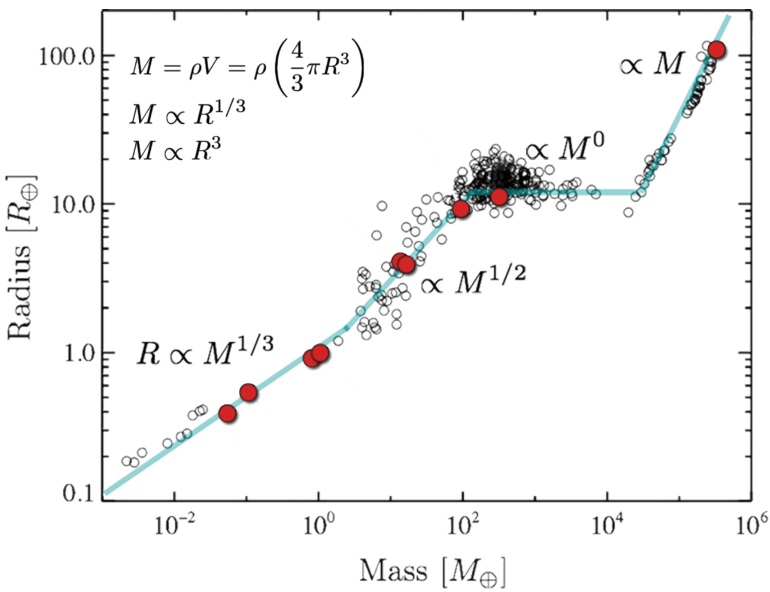
- On a regular x-y chart, a straight line means that $y = ax + b$, where a is the slope of the line and b is a constant, the y-intercept. That's a linear relationship. But on a logarithmic chart, a straight line means that there's a linear relationship between the logs of the variables: $\log x = a \log y + b$.

- In this case, the log of the radius (R) equals a times the log of the mass plus a constant: $\log R = a \log M + b$.

- Solving for R results in the following: $R \propto M^a$.

- This kind of relationship is called a power law, where one variable is proportional to another raised to some power. And evidently, when Mother Nature creates objects in our galaxy, she uses 4 different power law relationships between radius and mass.

- Let's measure the slope—the value of a—in each of the 4 zones.



- For the lowest-mass objects, the slope is about 1/3. That means $R \propto M^{1/3}$. This low-mass regime is closest to the one where we have some direct experience: small things, such as rocks and boulders. And the relation between radius and mass of everyday objects depends on the density of whatever material the object is made of. We can understand objects that have masses less than 1 Earth mass; they behave like rocks.

- In the second zone, the slope is about 1/2, so $R \propto M^{1/2}$. This means that the more massive objects have bigger radii than we would expect if they all had the same density. The more massive objects are less dense. The most massive objects in this zone are a lot less dense than rock; this makes sense because these are gaseous planets.

- In the third zone, the slope is 0. The size hardly changes at all, even when the mass is increased by a factor of 100. In everyday life, when we pack more mass onto a ball, the ball gets bigger; apparently this is not the case for balls between 100 and 10,000 Earth masses. The more massive objects are much denser than the less massive versions. Part of the reason these objects are increasingly dense is gravitational compression: They are so massive that their own gravity compresses them to higher densities than usual. The other part of the explanation is an effect called quantum degeneracy pressure. The objects in this zone are sometimes called Jovian planets, though toward the higher-mass end, the traditional term is brown dwarfs.

- For the highest-mass objects, the slope is about 1, which means that radius is proportional to mass. These are stars—objects for which gravitational compression is so strong that nuclear fusion ignites at the center, creating lots of heat and pressure. This same nuclear fusion also produces the light that stars are famous for; it's what makes stars shine.

READINGS

Astronomy Picture of the Day, https://apod.nasa.gov/apod/astropix.html.

Choudhuri, *Astrophysics for Physicists*, chap. 1.

Chown, *Solar System*.

Eames Office, "Powers of Ten™," https://www.youtube.com/watch?v=0fKBhvDjuy0.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 1.

Khan Academy, "Introduction to Logarithms," https://www.khanacademy.org/math/algebra2/exponential-and-logarithmic-functions/.

Obreschkow, "Cosmic Eye," https://www.youtube.com/watch?v=jfSNxVqprvM.

Scharf, *The Zoomable Universe*.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 2

ZOOMING IN TO FUNDAMENTAL PARTICLES

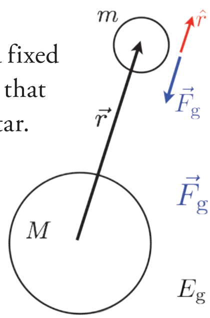
Usually, when something is called "astronomical," it means it's really big. But just as crucial for astrophysics are the orders of magnitude smaller than human scales, because the smallest and largest scales of the universe are deeply connected. In this lecture, you will be exposed to the realm of fundamental particles, including electrons, protons, neutrons, and neutrinos. In addition, you will learn about the 4 fundamental forces of nature: gravity, electromagnetism, the strong nuclear force, and the weak nuclear force.

GRAVITY

- Gravity—probably the most familiar of the 4 fundamental forces of nature—is what keeps us pinned to the surface of the Earth.

- Every mass attracts every other mass, according to Newton's law of gravity, which says that the force is proportional to the product of the 2 masses and inversely proportional to the square of the distance between them. The constant of proportionality is G, Newton's gravitational constant, which has a value of 6.7×10^{-11} N m^2 kg^{-2}.

- Let's suppose that we have a big mass, M, held fixed at the origin of our coordinate system and that m is free to move—like a planet orbiting a star. This means that m will feel a pull toward the origin.



$$\vec{F}_\text{g} = -\frac{GMm}{r^2}\,\hat{r}$$

$$E_\text{g} = -\frac{GMm}{r}$$

- To convey the direction of the force, we'll use vector notation. A vector is a quantity with a magnitude and direction. We put an arrow (→) over the F to remind us it's a vector. And, by convention, a "hat" (^) on top means it's a unit vector, or a vector with magnitude of 1, so all it's doing is specifying a direction. $\hat{r}$ points in the direction of increasing r—that is, away from the origin. But the force is toward the origin, which is why there is a minus sign. The acceleration vector points in the opposite direction as $\hat{r}$.

- The potential energy associated with the gravitational force also varies as the product of masses, but it goes inversely with r as opposed to r^2. Again, it's negative. But does this make sense? If we let m fall toward the origin, r shrinks, and according to the formula, the potential energy becomes more negative, which implies positive energy must be showing up somewhere else, because the total energy is conserved. And that does make sense: The kinetic energy, $1/2mv^2$, is increasing as the mass accelerates toward the origin. The gravitational potential energy is being converted into kinetic energy.

Logarithmically Zooming In

On the human scale, things are measured in meters. Zooming in to a tenth of a meter, we center on the human face, and as we keep narrowing our field of view to a hundredth of a meter, 10^{-2}, we stare into the human's eye.

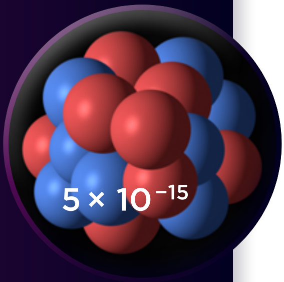
Another factor of 10, to the millimeter scale, and we can fit through the pupil of the eye and dive inside. At 10^{-4} meters, we can see the blood vessels in the retina, and by the time we hit 10^{-5}, we can see individual blood cells.

Once we reach 10^{-6} meters, a millionth of a meter—called a micron—we can see individual bacteria, each one a few microns across. We've zoomed in to the size of the wavelength of light. Light is a wave, an oscillating pattern of electric and magnetic fields—but it's hard to tell this on human scales, because the wavelength is only about half a micron. On this scale, though, light waves bend and spread, like water waves, and it's impossible to focus them sharply. That's the phenomenon of diffraction.

After another few orders of magnitude, at 10^{-8} meters, we start to see that the water that surrounds us is not a continuous fluid. It's made of individual molecules.

When we zoom in to 10^{-9} meters—that is, the scale of nanometers—molecules don't look solid. Instead, they look fuzzy, and they're in constant motion, jiggling and vibrating. They're getting knocked around by other molecules. The energy of all those random motions is what we perceive as heat on human scales. The hotter the material, the more vigorously the molecules are bouncing around.

Zooming in closer, we see the individual atoms that make up molecules. For example, a single water molecule is made of 2 hydrogen atoms and 1 oxygen atom. Like any atom, oxygen has a nucleus, which has a positive electrical charge, and is surrounded by orbiting electrons, which have a negative electrical charge. Because of the opposite charges, the nucleus and electrons are attracted to each other.

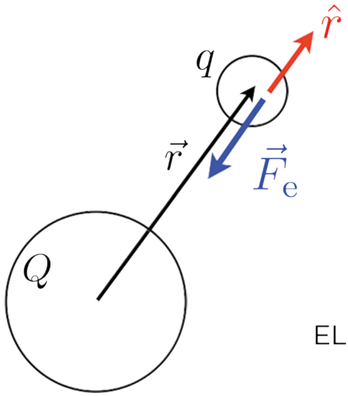


Before we can make out any details of the nucleus, we need to go 4 orders of magnitude below the atomic scale. The diameter of the oxygen nucleus is about 5×10^{-15} meters, or 5 femtometers. At this scale, if the nucleus were a marble, the electron cloud would be the size of a football field. We can see now that the nucleus is actually a cluster of 16 little marbles: 8 are protons and 8 are neutrons.

Zooming further, inside the proton, things become very hectic. There are quarks within a sea of particles called gluons, and everything is in motion, with particles appearing and disappearing.

ELECTROMAGNETISM

- The second fundamental force of nature is electromagnetism. The relevant equation here is Coulomb's law, which says that the electrical force goes as the product of charges divided by r^2 and is very similar in form to Newton's law of gravity. For the proportionality constant, we'll use the Greek letter eta (η). Numerically, η is 9×10^9 N m^2/C^2, where the Coulomb (C) is the standard unit of charge. In those units, the electron and the proton both have a charge of magnitude 1.6×10^{-19}, which can be represented as e.



$$\eta = 9.0 \times 10^{9}\ \mathrm{N\,m^2\,C^{-2}}$$

COULOMB'S LAW

$$\vec{F}_{\mathrm{e}} = +\frac{\eta\, Q q}{r^2}\,\hat{r}$$

COULOMB ENERGY

$$E_{\mathrm{e}} = \frac{\eta\, Q q}{r}$$

ELECTRON OR PROTON CHARGE

$$e = 1.6 \times 10^{-19}\ \mathrm{C}$$

- Notice that the force law has a plus sign this time, not a minus sign. When the product of charges is positive—that is, when they're both the same sign—the force is repulsive, pushing the charges apart. When the charges have opposite signs, like an electron and a proton, they attract.

- The Coulomb energy is the potential energy associated with the electric attraction or repulsion. As in the case of gravity, it varies as $1/r$.

- The Coulomb force explains why the electrons of an atom are attracted to the nucleus. But there must be something else going on, because why don't the electrons fall all the way down onto the nucleus, neutralize it, and come to rest?

- We might ask the same question about the Earth: If it's attracted to the Sun, why doesn't it fall in and burn up? The answer in this case is that the Earth has a nonzero angular momentum—a sideways velocity—and the gravitational acceleration just keeps turning its velocity vector around in a circle.

- When we look closely at an atom, we might expect to see the electrons whirling around the nucleus, like a miniature solar system, but we don't. Instead, the electrons look indistinct; there's an electron cloud surrounding the nucleus. That's because electrons, like all fundamental particles, obey the rules of quantum theory, the counterintuitive laws of motion and interaction. These rules are more exact and fundamental than Newton's laws of motion.

- Quantum theory says that when we measure the location of an electron, or any fundamental particle, we get a specific answer. But when we're not measuring it—when we're not forcing the question of where it is—the electron spreads out into a cloud. And there's no way to predict exactly where we'll find it when we do measure it. All we can say is we're likely to find it somewhere in this cloud, or wave function.

The cloud is called a wave function because the equation that governs the size and shape of the cloud—how it moves and interacts with other clouds—resembles the equation for ordinary waves. And like regular waves, the wave function can take the form of a pattern moving through space with a certain speed. It can even interfere with other wave functions, producing fringes, like when water waves overlap.

- In the case of an electron near the nucleus of an atom, the wave function isn't moving; it's trapped by the electrical attraction to the nucleus. It's like a sound wave reverberating inside an organ pipe or the vibrating surface of a drum. And the wave function obeys Heisenberg's uncertainty principle: If you try to pin down a particle's location, by trapping the wave function in a tiny volume, the particle's momentum—mass times velocity—becomes more uncertain.

HEISENBERG'S UNCERTAINTY PRINCIPLE

$$\Delta x \, \Delta p \geq \frac{\hbar}{2}$$

$$\hbar = \frac{h}{2\pi} \longleftarrow \underset{\text{PLANCK'S CONSTANT}}{6.6 \times 10^{-34}\ \mathrm{J} \cdot \mathrm{s}}$$

- In this mathematical relationship, Δx is the spatial extent of the cloud and Δp is the extent of the momentum cloud (the range in the possible values of momentum the particle might have if you measure it).

- You can't make both Δx and Δp as small as you might want; their product is always at least $\hbar/2$, a fundamental constant of nature. The h is Planck's constant, 6.6×10^{-34} joule-seconds, and the small bar through it is shorthand for $h/2\pi$.

- The uncertainty principle explains why atoms are stable. Even if you drop an electron directly onto a proton, with 0 angular momentum, it doesn't fall down and come to rest. That would imply that Δx and Δp are both 0—and this can't be. Instead, the wave function strikes a balance between D*x* and D*p*.

- The proton exists as a wave function, too, but there's a big difference. Even though the proton and electron have charges of equal magnitude, the proton is much more massive—by a factor of 1800. This ends up causing the proton's wave function to be much smaller than the electron's.

- Neutrons are nearly the same size and mass as protons but without any electrical charge. They're neutral. But if the protons in a nucleus have positive charge and the neutrons are neutral, then there aren't any negative charges, so what's holding this cluster of "marbles" together? Shouldn't the protons repel each other and fly apart?

THE STRONG NUCLEAR FORCE

- This brings us to the third fundamental force of nature: the strong nuclear force. This is a very short-range force that acts between nucleons—protons and neutrons. It's a complicated force with no simple equation. It depends on how many nucleons are present, which kinds, whether they're spinning, and other things. And it only acts over femtometers. Beyond that, it's negligible.

- Think of it this way: In a stable nucleus, all the "marbles" are coated with a layer of "glue" strong enough to withstand the electric repulsion. That's the strong force. The strong force is also why the "marbles" are rigid: The force is attractive up until the point of contact, but then it becomes repulsive, so it's very difficult to compress a nucleus.

THE WEAK NUCLEAR FORCE

- Neutrinos—neutral particles that have a much smaller mass than the electron by at least a factor of a million—interact mainly through the fourth fundamental force of nature: the weak nuclear force.

- The weak force is a short-range force, like the strong force, but it's not like any sort of "glue." In fact, it's kind of a stretch to call it a force; it's more like a special power nucleons have to change identities. A neutron can change into a proton, or vice versa. For example, a neutron sitting all by itself will spontaneously turn into a proton within about 10 minutes.

- The total electrical charge should be conserved, so the new proton's positive charge has to be balanced by negative charge somewhere else. What happens is the weak force conjures up an electron along with the proton and they sail away in nearly opposite directions.

- You'd expect them to be exactly opposite, because in addition to charge, momentum has to be conserved. The initial momentum of the stationary neutron was 0, so you'd think the proton and electron would have equal and opposite momenta. But when you measure them, they're not exactly opposite. The reason is that the weak force also produces a neutrino that sails away at nearly the speed of light, carrying just enough momentum so it all adds up to 0.

COMPARING FORCES

- What sets the nuclear forces apart is you only notice them on femtometer scales. In contrast, gravity and electromagnetism are long-range forces, acting on all scales, and their force laws look similar: They both go as $1/r^2$. But there are major differences between gravity and electromagnetism, starting with the fact that electromagnetism is much stronger.

- Say we have 2 protons separated by some distance r. What's the ratio between the force of electric repulsion and the force of gravitational attraction?

- To find out, we divide the Coulomb force by the gravitational force. The r^2 terms cancel, and when we plug in the numerical values of all the constants, we find a ratio of 10^{36} power. In other words, the electrical repulsion is unimaginably stronger than the gravitational attraction.

$$9.0 \times 10^{9}\ \mathrm{N\,m^2\,C^{-2}} \qquad\qquad 1.6 \times 10^{-19}\ \mathrm{C}$$

$$\frac{F_{\mathrm{e}}}{F_{\mathrm{g}}} = \frac{\frac{\eta e^2}{r^2}}{\frac{G m_p^2}{r^2}} = \frac{\eta e^2}{G m_p^2} \approx 10^{36}$$

$$6.7 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}} \qquad\qquad 1.7 \times 10^{-27}\ \mathrm{kg}$$

- But if gravity is really so pathetic, why is it the most familiar force of nature? The reason is that gravity is always attractive; it's never repulsive. There's no such thing as negative mass.

- Electromagnetism is different. Here, particles can be positive or negative, and when they merge together, the result is neutral. It doesn't feel any electric force. So, what happens is that all the positive and negative charges in the universe attract one another. They quickly find each other, getting pulled together with tremendous force, forming tiny structures with no net charge: atoms. This is why the incredible strength of the electric force is hidden from us.

Richard Feynman compared the situation of the proton and electron pulling on each other inside the atom to a pair of Olympic arm wrestlers pulling on each other's arms with tremendous force. They're equally matched; their clenched hands aren't moving. From far away, you might not even be aware of their intense effort.

- Once neutral atoms form, all that's left of electric forces are the slight imbalances that arise because the negative charge, the electron cloud, is more spread out than the positive charge, the nucleus.

- In fact, what we perceive as everyday forces—our feet pushing on the ground, our hands pulling on a rope, our knuckles knocking on a door—are all complex manifestations of the residual forces that are left over from the combination of electromagnetism and quantum theory.

- Gravity, on the other hand, never gets cancelled. That's why, when we zoom out to astronomical scales, gravity is the dominant force. That's why gravity—even though it's weak—sculpts the properties of planets, stars, and galaxies.

- Electromagnetism is more than just attraction and repulsion. There are magnetic fields, which come from moving charges, and there is electromagnetic radiation, which comes from accelerating charges. Whenever you accelerate a charge—speed it up, slow it down, whirl it around—it radiates. It takes some of its own energy and flings it outward at the speed of light. The radiated energy takes the form of photons.

This chart of distance scales from the previous lecture now includes 3 key units discovered during this lecture's zoom-in: the micron (same order of magnitude as the wavelength of visible light), the Bohr radius (the atomic scale), and the femtometer (the nuclear scale).

SCALE	UNIT	EXPRESSED IN METERS	EXPRESSED IN THE PREVIOUS UNIT
Nucleus	fm	10^{-15} m	—
Atom	a_0	5×10^{-11} m	52,000 fm
Visible Light	μm	10^{-6} m	18,900 a_0
Human	m	1 m	10^6 μm
Planets	$R_\oplus$	6.4×10^6 m	6.4×10^6 m
Stars	$R_\odot$	7.0×10^8 m	109 $R_\oplus$
Planetary Systems	AU	1.5×10^{11} m	215 $R_\odot$
Between Stars	pc	3.0×10^{16} m	206,265 AU
Galaxies	kpc	3.0×10^{19} m	1000 pc
Between Galaxies	Mpc	3.0×10^{22} m	1000 kpc
Observable Universe	Gpc	3.0×10^{25} m	1000 Mpc

READINGS

Astronomy Picture of the Day, https://apod.nasa.gov/apod/astropix.html.

Choudhuri, *Astrophysics for Physicists*, chap. 1.

Chown, *Solar System*.

Eames Office, "Powers of Ten™," https://www.youtube.com/watch?v=0fKBhvDjuy0.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 1.

Obreschkow, "Cosmic Eye," https://www.youtube.com/watch?v=jfSNxVqprvM.

Scharf, *The Zoomable Universe*.

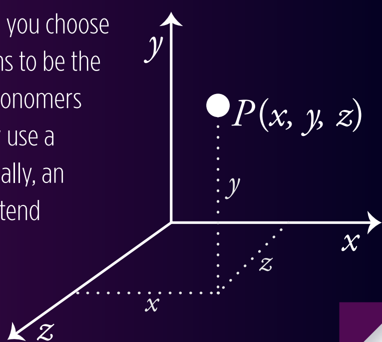
Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 3

MAKING MAPS OF THE COSMOS

If we want to explore the whole universe, we'd better have a good map. This lecture is about how astronomers locate objects in the universe.

If you want to specify the location of an object in 3 dimensions, you use Cartesian coordinates; you choose an origin and pick 3 perpendicular directions to be the *x*, *y*, and *z* axes. Instead of this system, astronomers use latitude, longitude, and elevation. They use a spherical polar coordinate system—specifically, an Earth-centered coordinate system—and extend the concepts of latitude, longitude, and elevation up into the heavens.



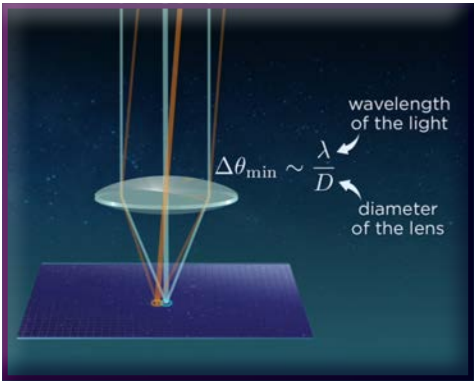
Imagine a giant transparent sphere that is centered on the Earth and marked with grid lines of latitude and longitude. The latitude lines tell us how far we are from the celestial equator—the projection of the Earth's equator up into the sky—and the longitude lines tell us how far east or west we are from the celestial equivalent of the prime meridian. That way, when we look at a distant star, we can read off the star's angular coordinates by seeing where it appears relative to the grid. That leaves only the third dimension: the distance to the star, the celestial equivalent of elevation, which is much trickier to measure.

ANGULAR COORDINATES

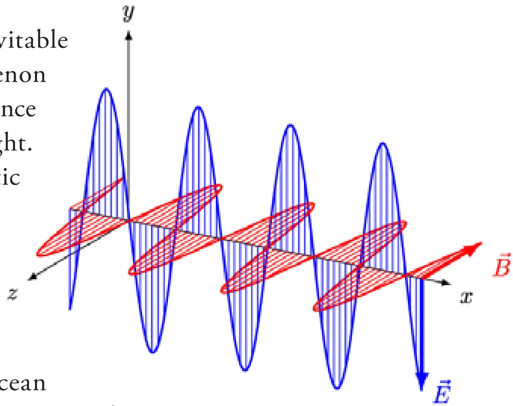
- Say there are 2 stars that happen to be located along nearly the same line of sight from the Earth, so they appear close together on the celestial sphere. If they're too close, they blend together and appear as a single point of light, rather than 2. What determines whether we can perceive the double star?

- Generally, it depends on how good the telescope is, but we can be more specific than that. This question pinpoints the basic dilemma of astronomy—that all we have is light. With few exceptions, our only source of knowledge is the electromagnetic radiation that happens to hit the spinning ball of rock we live on. So, we need to understand the physics of light.

- Imagine a telescope as a big lens pointed at a star straight overhead that focuses the starlight into a tight spot on a camera. If there's another star in a slightly different direction, then ideally, the lens focuses its light onto a different spot in the image, which shows 2 dots: star 1 and star 2.



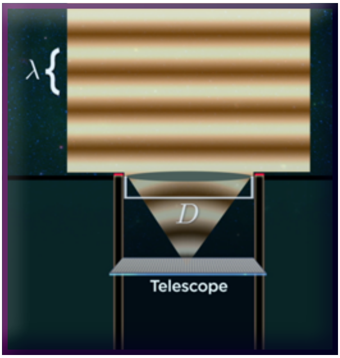
- The problem, though, is we can't focus light into as small a point as we might like. The stars blend together when the angle between them, $\Delta\theta$, is of order λ/D, where λ is the wavelength of light and D is the diameter of the lens—or mirror, or whatever we're using to collect and focus light.



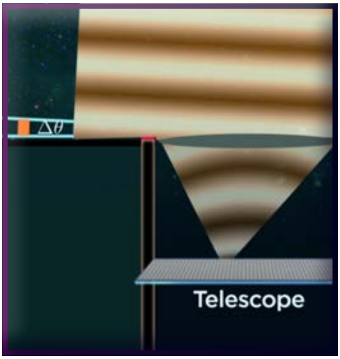
- The reason for this inevitable blurring is the phenomenon called diffraction, a consequence of the wave nature of light. Light is an electromagnetic wave, a traveling pattern of oscillating electric and magnetic fields.

- So, we can imagine the light from star 1 as an ocean wave, a traveling pattern of crests and troughs of electromagnetic energy, with a wavelength—a separation between crests—of λ. This wave passes through the diameter D of the telescope and then a lens or mirror responds to that pattern of energy by redirecting it toward a camera.

- The energy gets redirected to a different position on the camera if the starlight comes in from a different angle, tilted by $\Delta\theta$. But if $\Delta\theta$ is tiny, how could the lens or mirror possibly tell the difference? The wave energy is smeared out, with a spatial extent of λ. So, the telescope still sees a crest filling the opening. The optical system responds essentially the same way as before, directing the energy to the same spot on the detector.



- As $\Delta\theta$ increases, at what point do the waves from star 2 start to look different from star 1? The answer is when you no longer have a crest extending across the opening. The tilt is large enough that there's a crest at one end and a trough at the other end. For this minimum value of $\Delta\theta$, it's at least possible for the optics to distinguish the 2 waves.



- Because the distance between a crest and a trough is half of λ,

$$\tan \Delta\theta_{\min} = \frac{\lambda/2}{D} .$$

- In practice, $\Delta\theta_{\min}$ is a tiny number. For visible light, λ is about half a micron while D is usually a meter or more. So, we can use the small-angle approximation: Whenever an angle is small, the tangent is nearly equal to the angle itself, expressed in radians. In this case, this means that

$$\Delta\theta_{\min} \approx \frac{\lambda}{2D} .$$

- That's the minimum angular separation we can reliably measure with a telescope of diameter D.

- Keep in mind that this is just an order-of-magnitude relationship, based on the rough idea that we need to tilt the waves enough to get a shift of order λ across the telescope opening. To get this exactly right, we'd need to calculate the diffraction pattern of light after it passes through a circular hole of diameter D.

- What that more complex calculation shows is that the starlight gets focused into a blob surrounded by a pattern of rings, and the diameter of the blob is $1.22\lambda/D$. So, if 2 stars are closer than that, their blobs merge together. That's why the usual definition of this diffraction limit is $1.22\lambda/D$.

λ/D is a dimensionless number—a length divided by a length. That means the angle is measured in natural units, or radians.

In ordinary life, we measure angles in degrees, with a right angle being 90° and 360° going around the whole circle. Using the number 360 is a tradition going at least as far back as the ancient Babylonians, who used a base-60 counting system. That's also why we divide 1 degree into 60 arc minutes and 1 arc minute into 60 arc seconds.

It's simpler to calculate in radians, the system in which a right angle is $\pi/2$ and the whole circle is 2π.

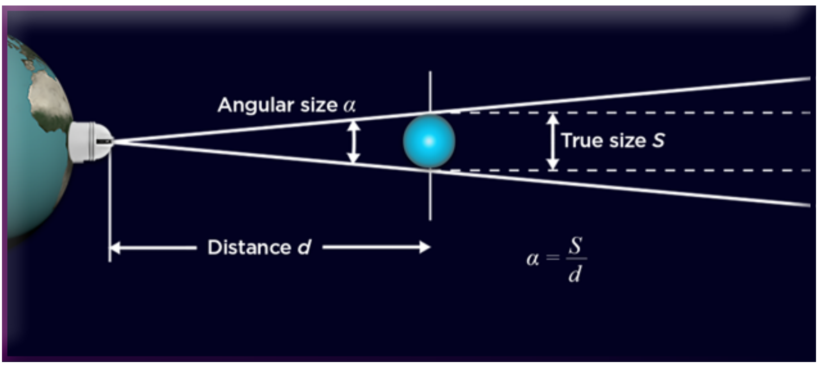
- Because the diffraction limit is proportional to λ/D, the way to improve our angular resolution is to increase D: build a bigger telescope. And that works, to a point. But in practice, there are other reasons why our images are blurry besides diffraction. Maybe our lens isn't polished perfectly or our mirror has defects. And then there's the constantly fluctuating atmosphere, which scrambles the directions of light rays at the level of an arc second, even at our best mountaintop observatories. So, even with a large telescope, we usually cannot achieve the ultimate diffraction limit. That's one reason why we launch telescopes into space, above the atmosphere.

MEASURING THE DISTANCE FROM EARTH

- Let's return to the original problem of locating things in space. The really tough problem—historically the most difficult problem in astronomy—is measuring that third dimension: the distance from Earth.

- Imagine we discover some new galaxy. We measure its angular extent on the celestial sphere. With only that information, we can't tell if the galaxy is relatively small and nearby or huge and billions of light-years away.

- So, we're on the Earth and an object is located a distance d from Earth. The object has a true size of S and an angular size of α—that is, the rays arriving from opposite sides of the object have an angle of α between them.

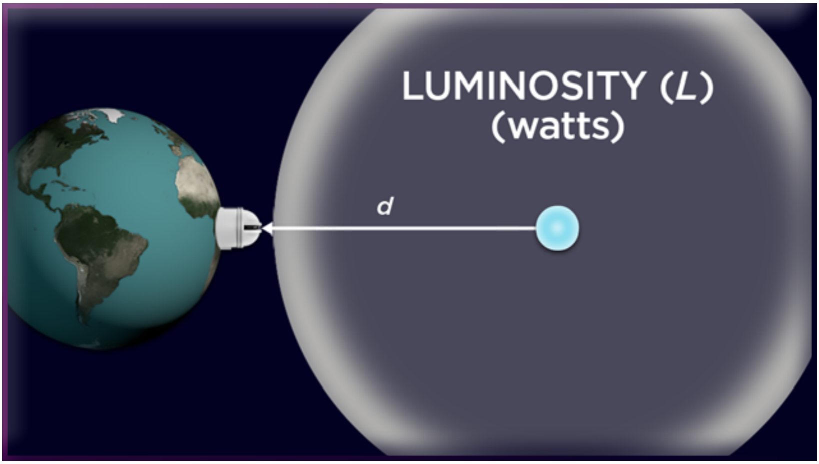


- We use the small-angle approximation again, because in practice α is tiny (maybe just a few microradians). That means, $\alpha \approx S/d$, or, equivalently, $S = \alpha d$. So, if all we know is α, we can't figure out S.

- A similar situation arises when we measure the brightness of a source. For a given brightness, we can't tell if the source is intrinsically luminous and far away or if it's actually intrinsically faint and happens to be nearby.

- Suppose the luminosity of the source is L. That's the power—the energy per unit time—that the luminous object is pouring out into space. We can measure L in watts, for example, and all that power spreads out as the radiation goes farther from the source.

- The Earth is far away, at a distance d. By the time the light reaches Earth, it's been spread out over a huge sphere of radius d.



- We can't measure L directly. Instead, we have a telescope with a certain collecting area, and we measure the power received by the telescope. We then calculate the power per unit area, which is called the flux, F.

- The following equation, representing the flux-luminosity relation, is an example of an inverse square law: The flux goes down as the inverse square of the distance.

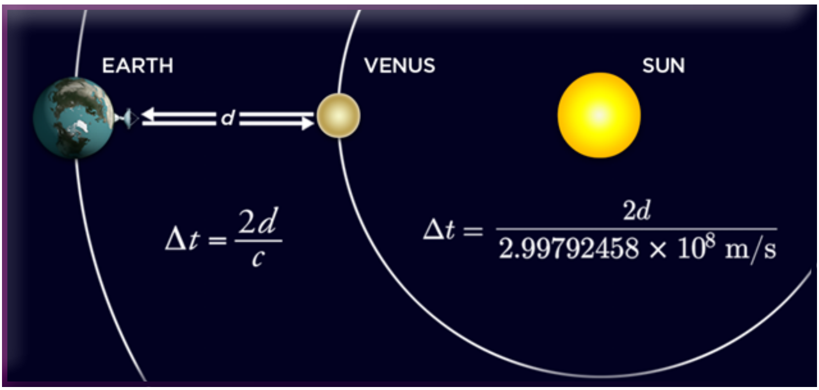
$$F = \frac{L}{4\pi d^2}$$

- If we measure F, we need to know d in order to deduce L. And we need to know L, the true luminosity, if we want to figure out what's physically going on to produce that radiation.

- Measuring the distance to an astronomical object is a crucial problem we need to solve. And we do so in 4 different steps: radar, parallax, standard candles, and standard explosions.

RADAR

- Build a giant radio transmitter, aim at a nearby planet, and fire. If you hit the target and your receiver is sensitive enough, you can detect the echo—the reflected radio waves. The echo is delayed by a time interval $\Delta t = 2d/c$, where $2d$ is the round-trip distance and c is the speed of radio waves (that is, the speed of light). And because we know the speed of light, we can calculate d.



- With the world's biggest transmitters, we can measure the distances to Mercury, Venus, Mars, and even some asteroids. That allows us to make maps of the solar system with a precision of a few parts in 10 billion, an astonishing level of detail.

- But unfortunately, this method is limited to relatively nearby objects. You can show that the amplitude of the echo falls off like $1/d^4$. To go beyond the solar system, we need other techniques. We move on to the second step of the 4-step solution: parallax.

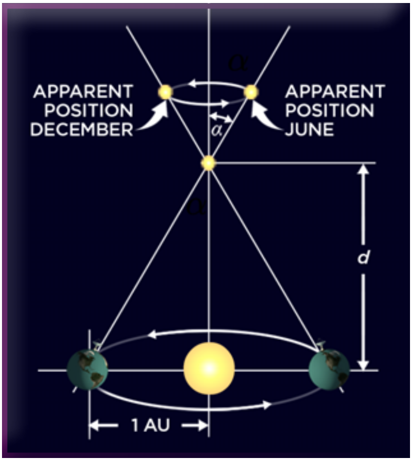
PARALLAX

- Parallax is based on simple geometry. Hold out your arm and raise a finger. Next, close your left eye and look at your finger and the scene in the background. Then, switch eyes: Open your left and close your right eye. It looks like your finger jumped! That's because your right eye views from a slightly different angle, so it sees your finger projected against a different part of the background scene.

- That's parallax. If you measure that shift in angle, as well as the distance between your eyes, you could use trigonometry to calculate the distance to your finger.

- In astronomy, we take advantage of the Earth's motion around the Sun. We take a picture of a nearby star and then wait 6 months for the Earth to go halfway around and take another picture. That's like closing your left eye and opening the right. The nearby star, like your finger, will appear to have shifted in position relative to the more numerous background stars.

- Let's do the math. Here's the Earth going around the Sun in a nearly circular orbit, with a radius of 1 AU. A star is at distance d. We need to point our telescope in a certain direction to see the star. But 6 months later, we need to point in a slightly different direction. In fact, as the year progresses, the star will seem to move in a little circle on the celestial sphere, with an angular radius of α.



- Using the small-angle approximation, $\alpha = 1\ \text{AU}/d$, or $d = 1\ \text{AU}/\alpha$.

- And because we already know the value of the astronomical unit very precisely from radar ranging, whenever we measure α, we can calculate d.

$$\alpha_{\text{arcsec}} = 206{,}265\,\alpha$$

$$d = \frac{206{,}265\ \text{AU}}{\alpha_{\text{arcsec}}}$$

$$1\ \text{pc} \equiv 206{,}265\ \text{AU}$$

$$d = \frac{1\ \text{pc}}{\alpha_{\text{arcsec}}}$$

- For simplicity, the star is shown directly above the plane of Earth's orbit. In general, though, the star will be off to the side somewhere. That doesn't change the basic idea. It just means that the star will appear to move in an ellipse, rather than a circle, and the parallax angle is the semimajor axis of the ellipse. And if the star is right on the ecliptic—the projection of the Earth's orbit onto the celestial sphere—it'll go back and forth along a straight line.

As is the case with the diffraction limit, the equation $\alpha = 1\ \text{AU} / d$ works when α is expressed in radians. But what if we want to use arc seconds?

One radian works out to be 206,265 arc seconds, so if we're expressing α in arc seconds, the right side of the parallax equation becomes 206,265 AU / d.

The tradition at this point is to define a new unit of distance, the parallax second, or parsec, equal to 206,265 AU. That way, the numbers are easier:

$$d = 1\ \text{parsec} / \alpha \text{ in arc seconds.}$$

The parsec is a handy unit for measuring the distances between stars, and it happens to have the same order of magnitude as the light-year.

$$1\ \text{parsec} \approx 3.3\ \text{light-years}$$

- Parallax is by far our reliable method for measuring distances to stars. But as the distance gets larger, eventually the parallax angle becomes too small to measure—if for no other reason than the diffraction limit.

- Right now, our best parallax measurements come from a space telescope called Gaia, launched by the European Space Agency in 2013. It measured parallaxes as small as 0.0001 of an arc second, good enough to make maps of the galaxy out to 10,000 parsecs, or 10 kiloparsecs.

- That's impressive. But to go beyond our galaxy—and there's a lot beyond our galaxy—we need to take another step in the quest to measure distances.

STANDARD CANDLES

- Our 2 best ways to measure more distant objects both rely on the flux-luminosity relation derived earlier: $F = L/4\pi d^2$, where L is the power an object emits and F is the power per unit area measured by Earthlings. We can measure F, but we can't figure out L unless we also know the distance, d.

- But suppose there were some light source out there for which we already knew L. In that case, we could calculate the distance by rearranging the flux-luminosity equation.

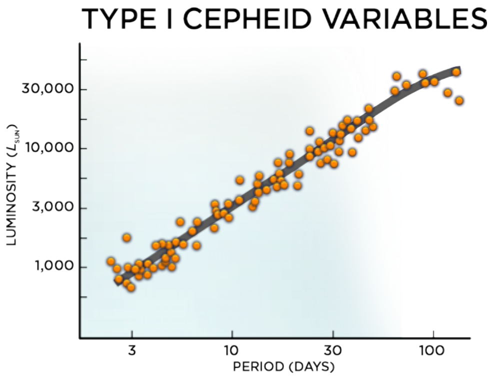
$$F = \frac{L}{4\pi d^2}$$

$$d = \sqrt{\frac{L}{4\pi F}}$$

- RS Puppis is an example of a category of stars called Cepheid variables—so called because the first known example was in the constellation Cepheus and because they vary in brightness. They pulse, getting brighter and fainter, in an endless cycle, with a period of typically a few weeks.

- The average luminosity of a Cepheid can be predicted accurately from the period of the pulsations, as discovered by Henrietta Leavitt in 1912. Stars that pulse more slowly are intrinsically more luminous.

- One reason we know this to be true is that some Cepheids are close enough for parallax measurements, so we can determine their luminosities. And among that collection, we observe that L is linked to the pulsation period, P. A schematic chart of luminosity versus period shows this increasing relationship.



- So, if we spot a Cepheid a megaparsec away, in some other galaxy, we can't measure its parallax, but we can measure the pulsation period. We just monitor the flux and see how long it takes to rise and fall. Then, we use the period-luminosity relationship to determine L and then calculate d, which is the distance to the galaxy where the Cepheid resides.

Cepheid variables were what first allowed us to zoom out beyond the Milky Way.

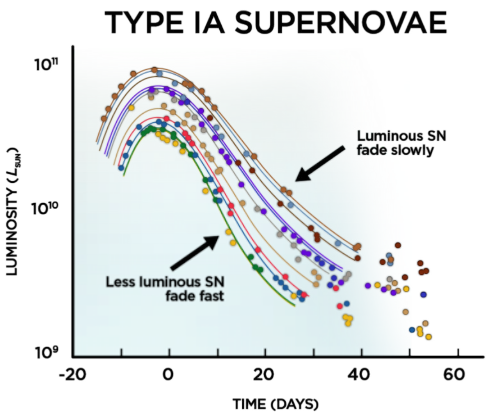
- Cepheids are examples of what are called standard candles—sources for which we somehow know the luminosity. Even though Cepheids have been used for more than a century to map out our galactic neighborhood, we don't know exactly why they are such good standard candles.

- With our best telescopes, Cepheids can be seen out to a distance of around 50 megaparsecs. But if we want to go farther—to reach out to gigaparsecs—we need to use standard explosions.

STANDARD EXPLOSIONS

- In the 1980s, astronomers realized that a certain category of exploding stars, or supernovas, produce fireballs that all have nearly the same peak luminosity. They're called Type Ia supernovas, and they all explode with nearly the same energy. They're predictable enough so that if you measure the color and duration of the afterglow, you can determine its luminosity to within a few percent. We know this because we've spotted Type Ia supernovas in nearby galaxies that also have Cepheids in them.

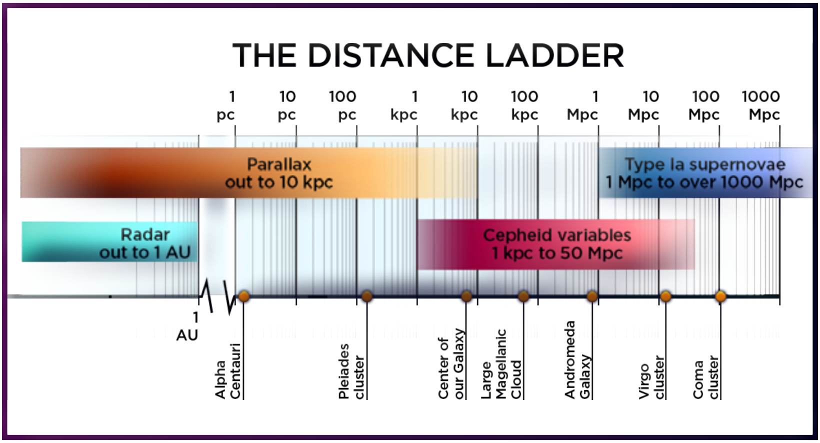
- Here's a chart showing data for some nearby Type Ia supernovas. The horizontal axis is time (in days) and the vertical axis shows the measured luminosity of the explosion. Notice that all the supernovas rise to about the same level, with differences that correlate with the duration: The faster the explosion fades, the lower the peak luminosity is.



- When we measure the rise and fall of flux from a really distant Type Ia supernova, we can match the observed duration of the event to one on this chart and then read off the peak luminosity.

- The great advantage of using Type Ia supernovas as standard explosions is that they're as bright as 5 billion Suns, bright enough to see even when they happen in galaxies that are extremely far away.

- We don't know for sure what causes Type Ia supernovas. They're almost certainly exploding white dwarfs, but the trigger for the explosion remains a topic of active research. What is clear, though, is that we can use them to measure cosmological distances. They're the last step in our quest.



READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 1.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 4.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

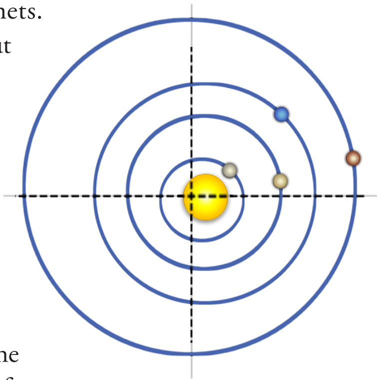
THE PHYSICS DEMONSTRATION IN THE SKY

To the naked eye, the planets of the solar system are points of light that move from night to night relative to a fixed background of stars. It was this celestial physics demonstration that led Isaac Newton to his law of gravity. The planets move in response to the Sun's gravity, without any friction. And each one is a different distance from the Sun, so by observing them, you can figure out how gravity depends on distance. The goal of this lecture—and the next one—is to solve this physics problem in the sky.

> The name "planet" comes from an ancient Greek word for "wanderer," because the planets wander through the constellations.

KEPLER'S FIRST LAW

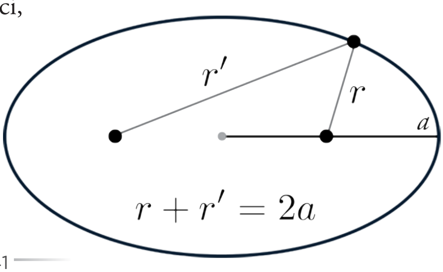
- The 17th-century astronomer Johannes Kepler pointed out 3 patterns in the motion of the planets. Kepler's first law is that the planets trace out ellipses as they go around the Sun. The orbits look like circles, but they're not; they're slightly flattened into ovals. And the Sun is not at the center.



- Before getting to the physics of why the orbits are ellipses, let's address the geometry of ellipses. Let's start simple, though, with a circle. Mathematically, a circle is defined as the set of all points that are the same distance from some chosen center.

> Wouldn't it be simpler if the orbits of planets were circles? The ancient Greeks certainly thought so. And later, even when the data got better and proved to be inconsistent with circular motion, theoreticians didn't abandon circles—they doubled down on them. They had the planets move in circles, the centers of which were themselves moving in circles. These are the "epicycles" that became the basis for the Ptolemaic model for the solar system, which prevailed up until Kepler.

- An ellipse has 2 focus points, or foci, and for all the points on the ellipse, the distance to the first focus plus the distance to the second focus is a constant. And that constant is equal to the length of the long axis of the ellipse, the major axis.



- If a circle has a radius of a, the equivalent for an ellipse is the radius along the major axis, called the semimajor axis, which can also be labeled a. With an ellipse, we also have the distance between the center and either focus, which can be whatever we want, as long as it's smaller than a. Tradition dictates that we express that distance as ae, where e—or eccentricity—is a number smaller than 1.

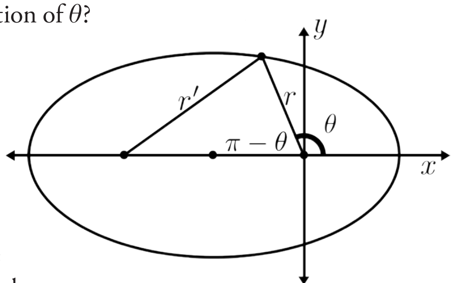
- When e is 0, the foci coincide at the center and we have a circle of radius a. As e gets larger—and closer to 1—the foci separate and we get a more elongated ellipse.

- The area of a circle is πa^2. For an ellipse, area is

$$\text{area} = \pi a^2 \sqrt{1 - e^2}.$$

- Next, we need the mathematical equation for an ellipse in polar coordinates. We start by introducing a coordinate system. We'll put the origin at one of the foci—for example, the one on the right—and we'll lay down x and y axes along the major and minor axes. To specify the points on the ellipse, we use polar coordinates: r is the distance from the origin and θ is the angle measured counterclockwise from the x axis.

- What is the equation for r as a function of θ?

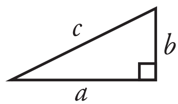


- To find the equation, we start with the fact that at any point, the sum of distances to the foci is equal to the length of the major axis, $2a$. We can write that as $r + r' = 2a$, where r is the distance to the focus at the origin and r' is the distance to the other focus.

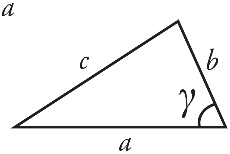
- But we want the equation to be purely in terms of r and θ, not r'. To get rid of the r', we use the law of cosines.

- The Pythagorean theorem says $c^2 = a^2 + b^2$, where a, b, and c are the lengths of the sides of a right triangle. The law of cosines is the generalization to any triangle, with γ being the angle across from the c side.

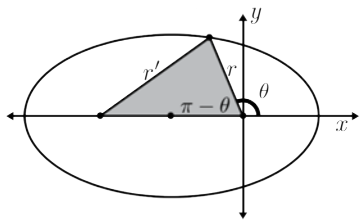
$$c^2 = a^2 + b^2$$



$$c^2 = a^2 + b^2 - 2ab\cos\gamma$$



- Let's apply the law of cosines to this triangle, with r′ as the c side.



$$\begin{aligned}
(r')^2 &= r^2 + (2ae)^2 - 2(r)(2ae)\cos(\pi - \theta) \\
(r')^2 &= r^2 + (2ae)^2 + 4are\cos\theta \\
(2a - r)^2 &= r^2 + 4a^2e^2 + 4are\cos\theta \\
4a^2 + r^2 - 4ar &= r^2 + 4a^2e^2 + 4are\cos\theta \\
a^2 - ar &= a^2e^2 + are\cos\theta \\
a^2 - a^2e^2 &= ar + are\cos\theta \\
a(1 - e^2) &= r(1 + e\cos\theta) \\
r(\theta) &= \frac{a(1 - e^2)}{1 + e\cos\theta}
\end{aligned}$$

$$\boldsymbol{r(\theta) = \frac{a(1 - e^2)}{1 + e\cos\theta}}$$

That's our equation!

- What Kepler noticed—his first law—is that all the planets move in ellipses, with the Sun not at the center but at one of the foci.

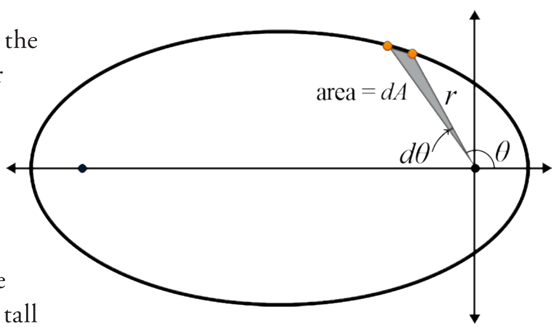
- Each planet has its own semimajor axis and eccentricity. For the Earth, a is about 93 million miles, or 150 million kilometers, or 1 AU. The Earth's orbital eccentricity is 0.017, which is quite small. All the planets in the solar system have small eccentricities. That's why it took so long to notice that the orbits are not circles. Mercury has the most eccentric orbit, with $e = 0.21$.

- Some of the planets around other stars that have been discovered over the last few decades have larger eccentricities, some even larger than 0.9. These incredibly elongated orbits were one of the big surprises of exoplanetary science.

KEPLER'S SECOND LAW

- Kepler's second law is about how fast the planets move. When they're close to the Sun, they move faster, in a specific way. As the planet moves, the line joining the planet and the star—the planet's radius vector—sweeps out area at a steady rate.

- When the planet is near the star, the radius is shorter but the planet moves faster, so the swept-out sector has a large angular width; when the planet is far away, the radius is longer but the planet is slower, making a tall and skinny sector. Kepler found that all the sectors have the same area.

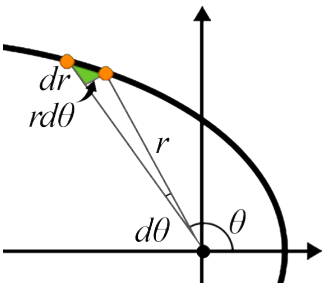


- Let's put the planet at an arbitrary position and say it moves for an infinitesimal time interval dt. It sweeps out a thin sector spanning an angle of $d\theta$, with an area of dA.

- What is the area of the sector? The planet moves in both the radial direction, or the r direction, and the perpendicular direction, or the θ direction—the direction of increasing θ. It's the motion in the θ direction that sweeps out area; purely radial motion doesn't sweep any area.

- In time dt, the planet moves in the θ direction by an amount $rd\theta$, using the small-angle approximation. So, the swept-out sector is basically a skinny right triangle with sides of r and $rd\theta$. The area of that triangle is

$$\frac{1}{2}r \cdot rd\theta.$$



- That leaves out a tiny corner piece of the sector whose dimensions are dr and $rd\theta$, the product of 2 tiny numbers. In the limit of infinitesimal dt, that piece is vanishingly small compared to the rest of the triangle. This means that we can write

$$dA = \frac{1}{2}r \cdot rd\theta.$$

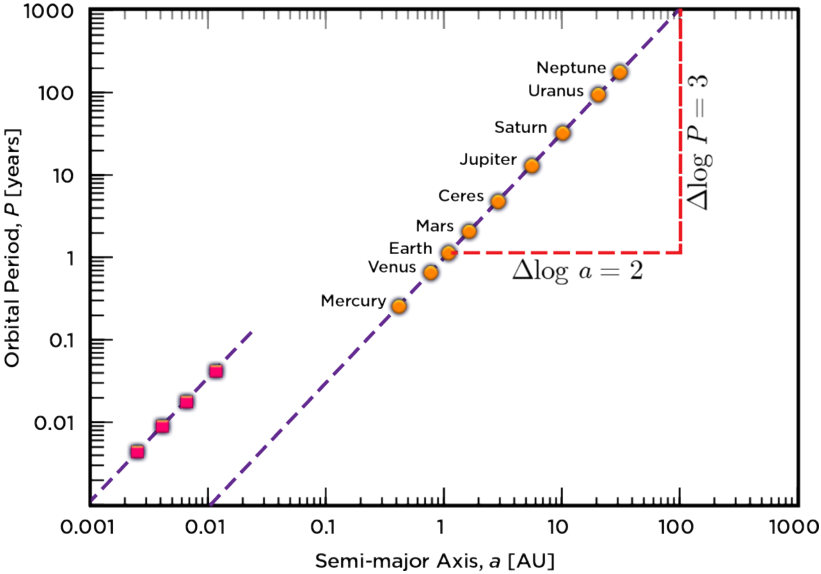
- And dA/dt, the rate at which area is swept out, is $\dfrac{dA}{dt} = \dfrac{1}{2}r^2\dfrac{d\theta}{dt}$.

- If that rate is a constant, as Kepler observed, then $d\theta/dt$ must be proportional to $1/r^2$. In other words, θ advances at a rate that varies as $1/r^2$.

KEPLER'S THIRD LAW

- Kepler's third law is about total time required to go all the way around: the orbital period. The bigger the orbit, the longer the period.

- In this logarithmic chart, the horizontal axis shows the semimajor axis in AU and the vertical axis shows the period in Earth years. So, the point representing Earth is at 1 AU and 1 year, and the other points are for the other planets.



- It's striking—they all fall on a single straight line! It has a slope of 3/2; if we move 2 units to the right, the line goes up 3 units. Because this is a log plot, that means

$$\log P = \frac{3}{2} \log a + \text{constant}.$$

- This in turn means that $P \propto a^{3/2}$. That's Kepler's third law.

- Unlike Kepler, we now know that other planets have orbiting bodies, or moons. The data points for the 4 biggest moons of Jupiter all lie on a straight line, too—with the same slope of 3/2. But interestingly, it's not the same line as the one defined by the planets. It's shifted up, to a longer period for a given semimajor axis.

- Why is P proportional to $a^{3/2}$, and why do the moons of Jupiter have a larger proportionality constant?

NEWTON'S LAWS OF MOTION AND GRAVITY

- Kepler didn't understand why his laws—which should really be called patterns—hold. That task fell to Isaac Newton.

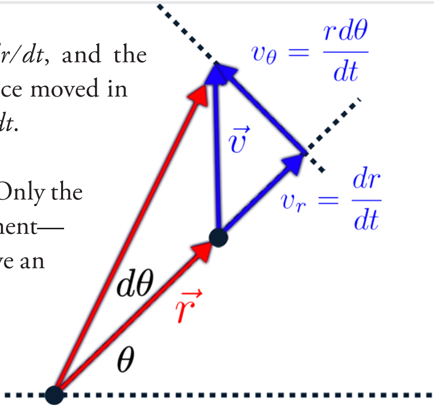
- Newton's law of motion is that the force acting on a body equals the mass of that body times its acceleration:

$$\vec{F} = m\vec{a}, \text{ or } \vec{F} = m\frac{d\vec{v}}{dt}.$$

- Newton's law of gravity is $\vec{F}_{\mathrm{g}} = -\frac{GMm}{r^2}\hat{r}$, where M is the Sun's mass and m is the planet's mass.

- How do these laws relate to Kepler's laws? Kepler's second law is the most fundamental, so let's start there. The key concept is the conservation of angular momentum.

- But before getting to angular momentum, let's consider momentum and velocity. Momentum is mass times velocity, $p = mv$, and the velocity has 2 components. In time dt, the velocity takes the planet from one position to another, changing both r and θ, so the velocity has a radial component—toward or away from the origin—and an angular component in the perpendicular direction.



- The radial component, v_r, is equal to dr/dt, and the angular component, v_θ, is $rd\theta$—the distance moved in the direction of increasing θ—divided by dt.

- Angular momentum, L, is defined as rmv_θ. Only the angular component—the sideways component—of the velocity matters. And because we have an equation for v_θ, we can write

$$L = r \cdot mv_\theta = mr^2 \frac{d\theta}{dt}.$$

- In vector language, $\vec{L} = \vec{r} \times m\vec{v}$.

- The cross product is the way to pick out only the perpendicular component; it has a magnitude of r times the component of mv that's perpendicular to r.

- In some circumstances, angular momentum is conserved. It doesn't change with time, even if the body is changing in other ways. The classic example is the twirling figure skater who pulls in her arms, effectively reducing her r—which means her $d\theta/dt$ must increase to compensate. That's why she twirls faster.

- Angular momentum is conserved whenever there's no net torque—no force in the θ direction. This is certainly true for the planets; the only force is gravity, which is in the radial direction, toward the Sun.

- So, as a planet goes around, even though r and v are always changing, $r^2 \frac{d\theta}{dt}$ is a constant and therefore

$$\frac{d\theta}{dt} \propto \frac{1}{r^2}.$$

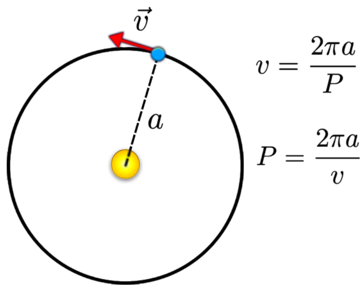
- That's Kepler's second law! This shows that Kepler's second law is a consequence of the conservation of angular momentum. Just as the ice skater twirls faster when she pulls in her arms, the planets twirl faster when they approach the Sun.

- This is an important result, with implications beyond planetary motion. It helps explain why material speeds up as it spirals into a black hole, why a star spins faster when it contracts in size, and why a young star is surrounded by a spinning disk of material within which the planets are formed.

- In Kepler's third law, why is the orbital period proportional to the 3/2 power of the semimajor axis?

- We'll answer this question in 2 stages: First, we'll prove Kepler's third law for a circular orbit, and then, in the next lecture, we'll prove it for the general case of elliptical orbits.

- Imagine a planet moving in a circle of radius a with some constant speed v. Over a full orbital period, P, the planet travels all the way around the circle. Therefore, v must equal the circumference of the circle, $2\pi a$, divided by P. Or, equivalently, $P = 2\pi a/v$.

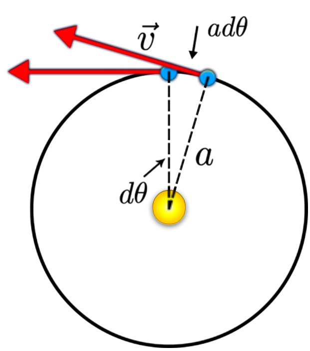


- One reason why P increases with a is that the circumference of the circle gets bigger. There's a longer way to go.

- In addition, when a is larger, v is lower; the planet moves more slowly because the gravitational attraction is weaker. This increases P even more so that at the end of the day P goes as $a^{3/2}$.

- In a time dt, the planet advances by a small angle $d\theta$, which corresponds to an arc length of $ad\theta$. So,

$$v = a\frac{d\theta}{dt}.$$



- During that same time interval, the velocity vector rotates by the same angle $d\theta$. The change in the velocity vector is $vd\theta$, so the magnitude of the acceleration, the rate of change of velocity, must be $\frac{v\,d\theta}{dt}$.

- Let's combine the equations by solving the first one for $d\theta/dt$ and then inserting the answer, v/a, into the second equation. This gives

$$\text{acceleration} = \frac{v^2}{a}.$$

- We just proved that to keep a body moving at speed v in a circle of radius a, an inward acceleration—a centripetal acceleration—of v^2/a needs to be supplied.

- In the case of a planet, that acceleration is provided by the Sun's gravitational force, $\frac{GM}{a^2}$.

- We set that equal to v^2/a: $\frac{GM}{a^2} = \frac{v^2}{a}$.

- Then, we solve for v, finding

$$\frac{GM}{a} = v^2$$

$$v = \sqrt{\frac{GM}{a}}.$$

- We can insert this into our previous expression for the period, and we find that P is proportional to $a^{3/2}$.

$$P = \frac{2\pi a}{v} \longleftarrow v = \sqrt{\frac{GM}{a}}$$

$$P = 2\pi a\sqrt{\frac{a}{GM}}$$

$$P = \frac{2\pi}{\sqrt{GM}}\; a^{3/2}$$

- We also find that the proportionality constant is $\dfrac{2\pi}{\sqrt{GM}}$.
- It decreases with the mass of the attracting body. That's why there was a vertical offset between the data for the planets and for the moons of Jupiter. The Jovian moons have a longer period for a given a because Jupiter is less massive than the Sun.

Kepler's third law is the most reliable way we have to measure the mass of just about anything in astrophysics. The basic idea is that to measure an object's mass, we need to watch other things moving in response to its gravity. It works for stars, planets, black holes, neutron stars, and entire galaxies. It even works—in a sense—for measuring the mass of the entire universe.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 2.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 2.

Ryden and Peterson, *Foundations of Astrophysics*, chap. 3.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 5

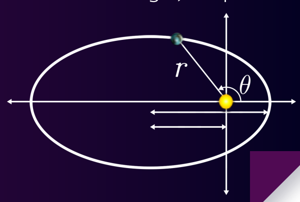
NEWTON'S HARDEST PROBLEM

Imagine that the year is 1660. The law of gravity is unknown. You've just read Johannes Kepler's books and puzzled over the 3 patterns he observed in the motion of the planets. Can you use those patterns to figure out the law of gravity? This is a tough problem that took Isaac Newton years to solve. But you have an advantage: calculus.

Kepler's Laws in Equation Form

Kepler's law says that the orbits of the planets are ellipses with the Sun at one focus, so if we use a polar coordinate system with the Sun at the origin, the path of the planet, $r(\theta)$ is

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



This is the equation for an ellipse that we derived in the previous lecture, in terms of the semimajor axis, *a*, and the eccentricity, *e*.

Kepler's second law says that the line from the Sun to the planet sweeps out area at a steady rate. In the previous lecture, we showed that this implies that

$$\frac{1}{2}r^2\frac{d\theta}{dt}$$

is a constant—a certain area per unit of time—specific to each planet. For the Earth, the numerical value is $\pi(\mathrm{AU})^2$ per year, because the Earth's orbit is approximately a circle of radius 1, which has a total area of π.

More generally, $\frac{1}{2}r^2\frac{d\theta}{dt}$ is equal to the area of the ellipse divided by the orbital period, *P*:

$$\frac{d}{dt}(\text{area}) = \frac{1}{2}r^2\frac{d\theta}{dt} = \frac{\pi a^2\sqrt{1-e^2}}{P}.$$

Kepler's third law says that *P* is proportional to $a^{3/2}$.

$$P \propto a^{3/2}$$

THE VELOCITY VECTOR

- We already know from laboratory experiments that force equals mass times acceleration, but we don't yet know the equation for the force of gravity. To obtain a clue, we need to calculate the acceleration of a planet that obeys Kepler's laws.

- To calculate acceleration, first we need to know the planet's position as a function of time. Then, we'll take the time derivative to get the velocity. Then, we'll take another time derivative to get the acceleration.

- Kepler's first law tells us the position, but not as a function of time—it's a function of angle, θ. All the time information is the second and third laws. So, we need to combine the equations somehow.

- Another problem is that we wrote Kepler's first law in polar coordinates, but with vectors, such as acceleration, it's easier to take derivatives in Cartesian coordinates, x and y. So, let's convert to Cartesian coordinates.

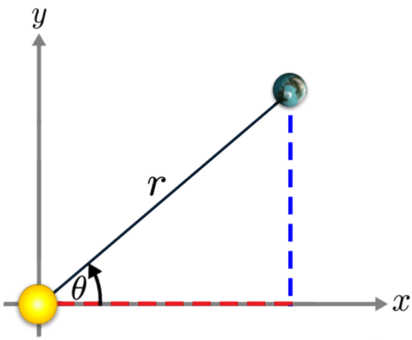
- In general, when the polar coordinates are r and θ, the x coordinate is $r\cos\theta$, and $y = r\sin\theta$.

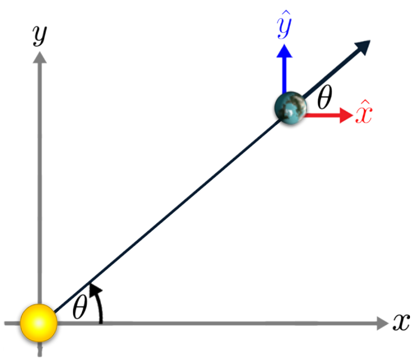
- So, for our planet,

$$x = a(1-e^2)\frac{\cos\theta}{1+e\cos\theta}$$

- And we get a similar equation for y:

$$y = a(1-e^2)\frac{\sin\theta}{1+e\cos\theta}$$





- We can do the same thing with unit vectors. In the polar coordinate system, $\hat{r}$ is a vector of length 1 pointing in the direction of increasing r. That means $\hat{r}$ changes orientation as the planet moves around; it always points away from the origin. At any point, though, we can write $\hat{r}$ as

$$\hat{r} = \hat{x}\cos\theta + \hat{y}\sin\theta.$$

- Now let's calculate the velocity by taking the time derivative of x and y. Because they're written as functions of θ, not time, we need to use the chain rule. v_x, the x component of velocity, is dx/dt, which is $dx/d\theta$ times $d\theta/dt$. To get $dx/d\theta$, we use standard tools of calculus. Because x has functions of θ in the top and bottom of the expression, we use the quotient rule. We take the derivative of the top times the bottom, minus the top times the derivative of the bottom, over the bottom squared. And we can simplify a bit.

$$v_x = \frac{dx}{dt} = \boxed{\frac{dx}{d\theta}} \cdot \frac{d\theta}{dt}$$

$$x = a(1-e^2)\frac{\cos\theta}{1+e\cos\theta}$$

$$\frac{dx}{d\theta} = a(1-e^2)\left[\frac{-\sin\theta(1+e\cos\theta) - \cos\theta(-e\sin\theta)}{(1+e\cos\theta)^2}\right]$$

$$= -a(1-e^2)\frac{\sin\theta}{(1+e\cos\theta)^2}$$

- For $d\theta/dt$, we need Kepler's second and third laws, the ones relating to time. Let's consolidate them by writing the P in the second law in terms of a using the third law, which says that P equals some constant, K, times $a^{3/2}$. Because the second law has a 1/2 on the left side and a π and a square root on the right side, we can cancel out the 1/2 and the π and fit the K under the square root and write the third law as follows.

$$v_x = \frac{dx}{dt} = \frac{dx}{d\theta} \cdot \boxed{\frac{d\theta}{dt}}$$

KEPLER'S 2ND LAW $$\frac{1}{2}r^2\frac{d\theta}{dt} = \frac{\pi a^2\sqrt{1-e^2}}{P}$$

KEPLER'S 3RD LAW $$P \propto a^{3/2} = \frac{2\pi}{\sqrt{K}}a^{3/2}$$

$$r^2\frac{d\theta}{dt} = \sqrt{Ka(1-e^2)}$$

$$\frac{d\theta}{dt} = \frac{\sqrt{Ka(1-e^2)}}{r^2}$$

- Now we have the ingredients we need to calculate the velocity. We plug in the expressions we just derived, which leads to an equation in terms of r and θ. To put everything in terms of just one variable, θ, we insert the ellipse equation for $r(\theta)$ and simplify.

$$v_x = \frac{dx}{dt} = \boxed{\frac{dx}{d\theta}} \cdot \boxed{\frac{d\theta}{dt}}$$

$$= \boxed{-a(1-e^2)\frac{\sin\theta}{(1+e\cos\theta)^2}} \cdot \boxed{\frac{\sqrt{Ka(1-e^2)}}{r^2}}$$

$$= -a(1-e^2)\frac{\sin\theta}{(1+e\cos\theta)^2} \cdot \sqrt{Ka(1-e^2)}\left[\frac{1+e\cos\theta}{a(1-e^2)}\right]^2$$

$$= -\sqrt{\frac{K}{a(1-e^2)}}\sin\theta$$

- The factor in front of $\sin\theta$ is a constant—it doesn't depend on θ or time—and it has units of velocity. To make the equation look even simpler, let's name that constant v_0 and simplify.

$$\sqrt{\frac{K}{a(1-e^2)}} \longrightarrow v_0$$

$$v_x = -v_0 \sin\theta$$

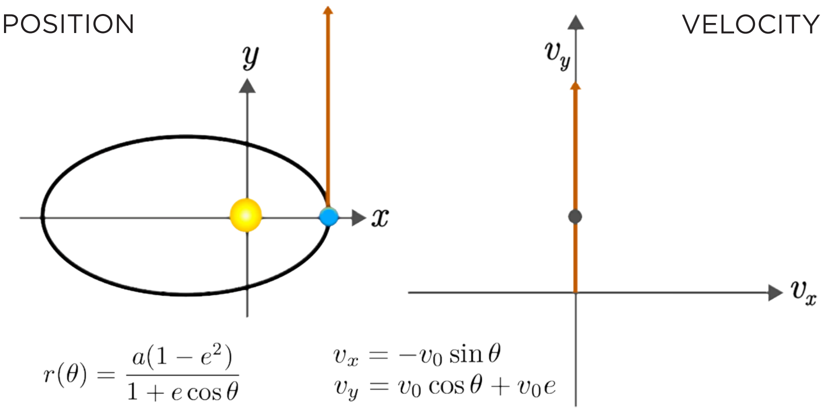
- That leaves the other component of velocity, v_y, which we calculate as $dy/d\theta$ times $d\theta/dt$. The steps are similar to the ones for v_x.

$$v_y = \frac{dy}{dt} = \frac{dy}{d\theta} \cdot \frac{d\theta}{dt}$$

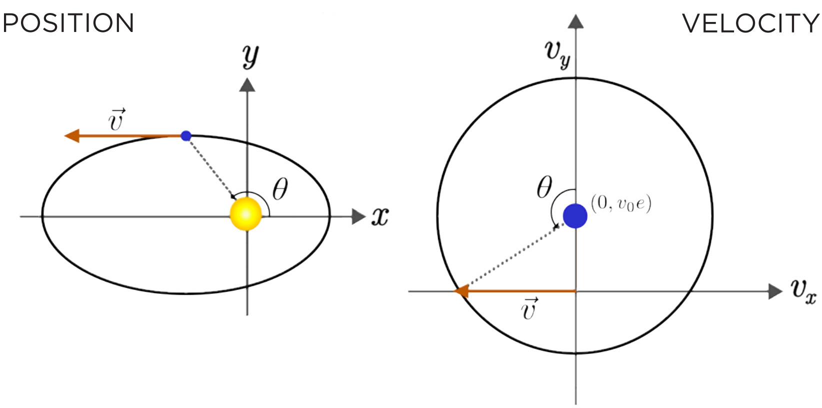
$$= \cdots = \sqrt{\frac{K}{a(1-e^2)}}(\cos\theta + e)$$

$$v_y = v_0\cos\theta + v_0 e$$

- What does all this mean? Let's find out by tracking the planet's velocity vector over a full orbit. We'll plot v_x on the horizontal axis and v_y on the vertical axis. This kind of chart is called velocity space; each point specifies a velocity, rather than a position.



- The equations tell us that v_x starts at 0 and v_y starts at $v_0 + e$. Then, as θ increases, v_x goes negative and v_y shrinks. When we keep going, we find that the tip of the velocity vector moves in a circle!



- We can prove it algebraically, too, by showing that our equations imply the equation for a circle in velocity space with radius v_0 and centered at the point $(0, ev_0)$.

$$\left.\begin{aligned} v_x &= -v_0 \sin\theta \\ (v_y - ev_0)^2 &= v_0^2 \cos^2\theta \end{aligned}\right\} v_x^2 + (v_y - ev_0)^2 = v_0^2$$

- While a planet moves in an ellipse, its velocity vector traces out a circle.

THE ACCELERATION VECTOR

- The x component of acceleration, a_x, is dv_x/dt, which we can use the chain rule, as we did previously, to write as $\frac{dv_x}{d\theta}\frac{d\theta}{dt}$.
- Substituting for v_0 and $d\theta/dt$,

$$= -\sqrt{\frac{K}{a(1-e^2)}}\cos\theta\,\frac{\sqrt{Ka(1-e^2)}}{r^2} = -\frac{K}{r^2}\cos\theta.$$

- The y component works the same way. We take the θ derivative to get $\sin\theta$ and then plug in $d\theta/dt$, leading to

$$a_y = \frac{dv_y}{dt} = \frac{dv_y}{d\theta}\frac{d\theta}{dt} = -v_0 \sin\theta\,\frac{d\theta}{dt}$$

$$= -\sqrt{\frac{K}{a(1-e^2)}}\sin\theta\,\frac{\sqrt{Ka(1-e^2)}}{r^2} = -\frac{K}{r^2}\sin\theta.$$

- In vector notation, we just learned that acceleration is

$$\vec{a} = -\frac{K}{r^2}\left(\hat{x}\cos\theta + \hat{y}\sin\theta\right) = -\frac{K}{r^2}\hat{r}.$$

- It all hangs together, if the Sun is pulling the planet toward it with a force whose strength varies as $1/r^2$. We just "discovered" the law of gravity by following Kepler's 3 clues.

- We now see that the constant K that appeared in Kepler's third law sets the overall strength of the Sun's gravitational force. If we further assume that the force is proportional to the mass of the attracting body, we can write K as $GM_\odot$, where G is a fundamental constant of nature.

$$-\frac{GM_\odot}{r^2}\hat{r}$$

THE CONSERVATION LAWS

- Now let's return to a modern stance, in which we already know the law of gravity and want to understand some other aspects of planetary motion. Specifically, let's consider the 2 big conservation laws: the conservation of angular momentum, L, and energy, E.

- Both L and E remain constant throughout a planet's elliptical orbit, even while the planet is moving and changing speed. So, we should be able to derive expressions for L and E purely in terms of constants—G, M, m, a, and e.

- First, let's do this for angular momentum. In general, $L = mrv_\theta$. Remember, only the angular component of the velocity matters. Because $v_\theta = r\frac{d\theta}{dt}$, we can also write L as $mr^2\frac{d\theta}{dt}$.

- And previously in this lecture, we consolidated Kepler's second and third laws into one equation:

$$r^2\frac{d\theta}{dt} = \sqrt{Ka(1-e^2)}.$$

- If we multiply this by m and substitute GM for K, we arrive at a new formula for angular momentum that is purely in terms of constants:

$$L = m\sqrt{GMa(1-e^2)}.$$

As an immediate application of this new formula for angular momentum, we can take care of a piece of unfinished business: Kepler's third law. In the previous lecture, we proved it for a circular orbit. Now we can prove it for the general case of an elliptical orbit.

KEPLER'S 2ND LAW $$\frac{d}{dt}(\text{area}) = \frac{1}{2}r^2\frac{d\theta}{dt} = \frac{\pi a^2\sqrt{1-e^2}}{P}$$

$$L = mr^2\frac{d\theta}{dt}$$

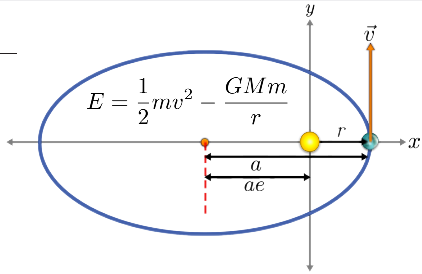
$$\frac{L}{2m} = \frac{\pi a^2\sqrt{1-e^2}}{P}$$

$$\frac{m\sqrt{GMa(1-e^2)}}{2m} = \frac{\pi a^2\sqrt{1-e^2}}{P}$$

$$\frac{\sqrt{GMa}}{2} = \frac{\pi a^2}{P}$$

KEPLER'S 3RD LAW $$P = \frac{2\pi}{\sqrt{GM}}a^{3/2}$$

- Energy—the other conserved quantity—has 2 parts: kinetic, $1/2mv^2$, and potential, $-GMm/r$.



- Their sum must be equal to some combination of the constants G, M, m, a, and e. Let's figure out what it is.

- Because energy is constant, we can calculate it at any point in the planet's orbit and get the same answer, so let's make life simple by choosing $\theta = 0$. That's when the planet makes its closest approach to the Sun and $r = a(1 - e)$.

- We can figure out the velocity with another application of our new angular momentum formula. In general, $L = mrv_\perp$.

- Here, at $\theta = 0$, $v_\perp$ is simply v, because at that point, the velocity vector is totally perpendicular to the radius vector: r is in the x direction and v is in the y direction. So, at $\theta = 0$, $L = ma(1 - e)v$.

- We solve for v and plug in our new expression for L.

$$v = \frac{L}{ma(1-e)} = \frac{m\sqrt{GMa(1-e^2)}}{ma(1-e)}$$

- Then, we insert that expression for v into the energy equation and simplify. The algebra leads to a cascade of cancellations and a result that's refreshingly simple.

$$E = \frac{1}{2}m\frac{GMm^2a(1-e^2)}{m^2a^2(1-e)^2} - \frac{GMm}{a(1-e)}$$

$$E = \frac{GMm}{2a}\left[\frac{(1+e)(1-e)}{(1-e)(1-e)} - \frac{2}{1-e}\right]$$

$$E = \frac{GMm}{2a}\left[\frac{1+e-2}{1-e}\right]$$

$$E = -\frac{GMm}{2a}$$

- All the terms related to eccentricity cancelled out—it turns out that energy depends only on the semimajor axis of the ellipse, not its eccentricity. If you have a nearly circular orbit with radius 1 AU, like the Earth's, and you compare it to a planet on a highly elliptical orbit, with $a = 1$ AU and $e = 0.9$, they'll both have the same energy.

- They'll also have the same orbital period (1 year) because Kepler's third law says P depends on a, but not on e. The planet on the elliptical orbit whips around the Sun near its closest approach, and moves more slowly when it's far away, and the 2 effects cancel each other exactly to give the same period as the Earth. It's an interesting coincidence.

A GRAPHICAL APPROACH TO UNDERSTANDING ORBITS

- Imagine a particle of mass m that is gravitationally attracted to a larger mass, M. We give our particle some initial position and velocity, which in turn corresponds to some values of angular momentum, L, and energy, E.

- As before, the energy is

$$\frac{1}{2}mv^2 - \frac{GMm}{r}.$$

- And the v^2 has 2 components, radial and angular: ${v_r}^2$ and ${v_\theta}^2$.

$$E = \frac{1}{2}m(v_r^2 + v_\theta^2) - \frac{GMm}{r}$$

- Let's now bring in the conservation of angular momentum.

$$L = mrv_\theta \longrightarrow v_\theta = \frac{L}{mr}$$

- So, we can rewrite the energy equation as

$$E = \frac{1}{2}mv_r^2 + \frac{L^2}{2mr^2} - \frac{GMm}{r}.$$

- That's an interesting way to write it, because the second term is purely a function of r, making it look sort of like the potential energy, even though it's really part of the kinetic energy. That's the basis of a neat trick: We define an effective potential energy, U_{eff}, equal to the highlighted term below.

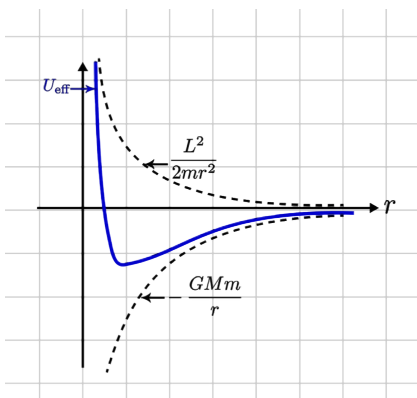
$$E = \frac{1}{2}mv_r^2 + \boxed{\frac{L^2}{2mr^2} - \frac{GMm}{r}.}$$

- That way, we can write

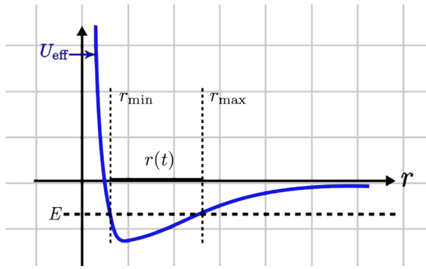
$$E = \frac{1}{2}mv_r^2 + U_{\text{eff}}(r).$$

- The reason this helps is because now the energy equation only depends on r and v_r, which is the time derivative of r. So, even though we live in a 3-dimensional world, the motion of the particle is governed by a single-variable equation! That's what makes it easy to understand graphically.

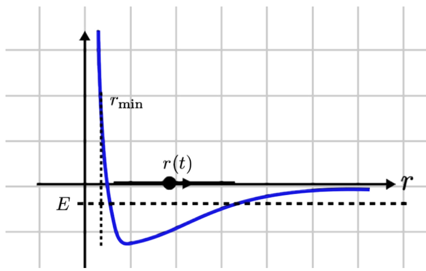
- Let's plot U_{eff} as a function of r. For small r, the $1/r^2$ is dominant, and it's positive, so U_{eff} shoots up to infinity. For large r, the $1/r$ is dominant, and it's negative, so as r grows, the potential dives down to negative values and rises toward 0 as r goes to infinity. It makes a bowl shape. The exact shape of the bowl depends on L, how much angular momentum we give the particle.



- The trajectory of the particle depends on E, how much total energy we give it. First, let's consider the case in which E is negative—the negative potential energy dominates the positive kinetic energy. We'll plot E as a dashed horizontal line.



- Because the difference between E and U_{eff} is equal to $1/2mv_r^2$, which is always a positive number, the particle's radius (r) must be confined to the region where E is bigger than V—that is, where the dashed line is higher than the solid line: inside the bowl of the effective potential.



- Furthermore, at locations where $E - U_{\text{eff}}$ is large, that means v_r is large, too, so the particle is moving quickly in the radial direction. Wherever U_{eff} gets close to E, the particle must be slowing down. When the lines cross, v_r is 0 and r is momentarily staying still.

- All this means that the particle's radial motion can be understood qualitatively by imagining we drop a marble in this bowl, starting at one of the intersection points. The marble starts at rest, rolls to the bottom and speeds up, rolls up to the same height on the other side, stops briefly, then drops down again, and keeps oscillating.

- Likewise, the r value of our particle will grow, then shrink, and then grow again, as it's whirling around. That makes sense: We already know that the particle follows an ellipse, with a distance to the origin that gets bigger and smaller as it goes around.

- And if we happen to put the particle right at the lowest point in this bowl, it will just stay there. That corresponds to a circular orbit, with an unchanging radius.

- This graph can teach us other things, too. For example, we've just seen that for a given angular momentum, a circular orbit has the minimum possible energy—the low point in the bowl. Whenever you drain energy out of an orbit, with friction or some other process that leaves angular momentum alone, the orbit will circularize.

- In addition, we see that it's impossible for the particle to ever reach $r = 0$. That's because of the first term in the effective potential, $L^2/2mr^2$, which makes an infinitely high barrier, guarding the origin. The only exception would be if L, the angular momentum, is exactly 0. Then, there's no barrier.

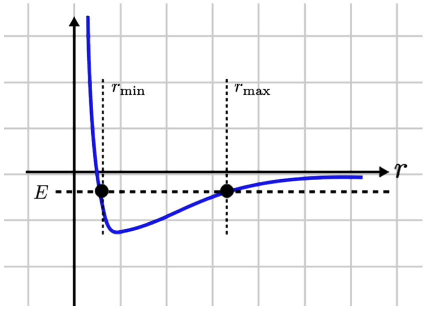
- In plain language, to make a direct hit on the origin, you need to be dropped straight in, with no sideways motion. If you have any angular momentum at all, you'll orbit the attractor—you won't hit it.

- When the total energy is positive, rather than negative, the dashed line intersects the solid line only once, near the center. So, the particle approaches the origin and then turns around and flies away, slowing down but never returning.

- That's an unbounded trajectory—like what happened in 2017, when an interstellar asteroid (which was later named 'Oumuamua) approached the Sun at high speed and with a lot of energy. It flew in from parsecs away, and its radial coordinate, r, shrunk to a quarter of an AU before the Sun deflected it in a different direction and it flew away. It was the first time anyone had detected an asteroid from some other star system encountering our own.

- The plot of the effective potential tells us what's happening to the radial coordinate, but it doesn't tell us what's happening to its θ coordinate. We need keep in mind that while r is changing, the particle is also moving in the perpendicular direction.

- What's happening to the θ coordinate for a bound orbit with negative total energy? The conservation of energy tells us that there will be some minimum and maximum radius for the particle that is set by the intersection points between the energy (the dashed line) and the effective potential (the solid curve).



- And the conservation of angular momentum tells us that the perpendicular velocity is L/mr, so it whirls faster when r is small and slower when it's farther out. That's Kepler's second law—which holds for any central force, not just gravity.

- So, you can play this trick—defining an effective 1-dimensional potential—for any central force law, whether the force goes as $1/r^3$, or $\sqrt{r}$, or whatever.

- In general, the particle whirls around, going from the minimum to the maximum radius and back again, in accordance with Kepler's second law. The trajectory makes a beautiful pattern called a rosette orbit that fills in the space between the minimum and maximum distance.

- But for the special case of the inverse square law, there's a remarkable coincidence: The trajectory comes around and repeats, making an ellipse. Just about any other force law—any other power of r—leads to infinitely looping rosettes instead of a fixed geometric shape.

- Another exception is if the particle is attached to the origin with an ideal spring, with force proportional to r; then, its trajectory is also an ellipse, but in that case, the origin is the center of the ellipse instead of the focus.

- Why is it that the actual force law chosen by Mother Nature is one of the exceptional cases that gives ellipses? It turns out that this coincidence is related to the fact that for the specific case of the $1/r^2$ law, there is a third conserved quantity besides energy and angular momentum. Here's the equation for this additional constant of motion.

$$\vec{e} = \frac{\vec{v} \times \vec{L}}{GMm} - \hat{r}$$

- The equation takes the planet's velocity vector, crosses it with the angular momentum vector, divides by GMm, and then subtracts the $\hat{r}$ unit vector. The result is called the eccentricity vector.

- The time derivative of this quantity is 0. The eccentricity vector is constant in time, even while v and $\hat{r}$ are changing throughout the orbit.

- The magnitude turns out to be the orbital eccentricity, and the direction of the vector specifies the orientation of the ellipse—it points along the major axis. Working that out is another way to prove that Newton's law of gravity implies Kepler's first law (as opposed to what we did, which was demonstrate that Kepler's laws imply an inward acceleration going as $1/r^2$).

- Whenever there's a conserved quantity, such as energy or angular momentum, there's a corresponding symmetry in nature—a sense in which nature is mathematically simpler than it could have been. This is called Noether's theorem, after Emmy Noether.

Energy is conserved because the laws of physics don't change with time. Angular momentum is conserved whenever the situation has rotational symmetry.

- It turns out that the equations governing the motion of a particle under the force of gravity from another particle are mathematically equivalent—through a complicated change of variables—to the equations for a particle moving freely, without any force, on the surface of a 4-dimensional sphere. And it's the perfect symmetry of that 4-dimensional sphere that leads to the conservation law for the eccentricity vector.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 2.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 2.

Ryden and Peterson, *Foundations of Astrophysics*, chap. 3.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 6

TIDAL FORCES

This lecture will address 3 questions. First, the major rings around planets are all in the range of about 2 to 2.5 times the radius of the planet. Why is that? Second, all of the giant planets have moons. Why are the large moons spherical while the smaller ones can have irregular shapes? Third, all the large, spherical moons are far from the planet—well outside the rings—whereas the small moons are found all over, including in and among the rings. Why are the large moons only found in wide orbits? To answer these questions, our discussion of gravitational orbits needs to be expanded.

UNDERSTANDING RINGS AND MOONS

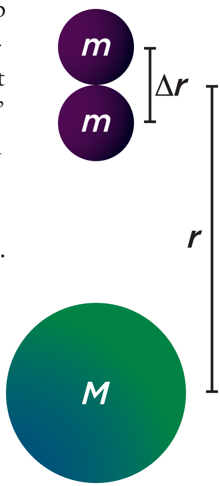
- So far, we have assumed that the orbiting body is a point mass—an infinitesimally small mathematical point. We've ignored the fact that a planet, or a moon, is a real object with a nonzero size.

- This is important because the force of gravity depends on distance; it gets weaker as you get farther away from the attracting mass. For example, the side of the Earth facing the Sun—the dayside—is pulled harder than the nightside. The differences in gravitational forces from one part of a body to another are called tidal forces.

- Suppose we have a planet of mass M that has a moon with a small mass, m, whirling around in a circular orbit of radius r.

- Actually, let's start with an even simpler case. If we drop the moon, starting from rest, it will accelerate downward and crash into the planet. It doesn't orbit because we didn't give it any angular momentum.

- Now let's give the moon a nonzero size. We could make it a sphere, but the math would get too hairy. To keep things simple, let's just take one step beyond the point-mass approximation. We'll model the moon not as 1 point mass, but as 2 point masses—2 rigid spherical "rocks," each of mass m, with their centers separated by some small distance Δr.



- When we let go of the rocks, they both fall onto the planet. They do not stay together as they fall. The inner rock is closer to the planet, so it feels a stronger gravitational force than the outer one, leading to a larger acceleration and causing it to pull away from the outer rock. Our "moon" breaks apart as it falls.

- In general, the magnitude of the force from the planet is

$$F(r) = \frac{GMm}{r^2}.$$

- Let's calculate the difference in force, ΔF, experienced by the rocks. Because Δr is very small compared to r,

$$\Delta F = F_{\text{out}} - F_{\text{in}} \approx \frac{dF}{dr}\Delta r = -\frac{2GMm}{r^3}\Delta r.$$

- The minus sign means that the force weakens with distance.

- We've learned that the part of the "moon" closest to the planet is pulled harder by an amount proportional to Δr, the size of the moon, and M, the mass of the planet and inversely proportional to the cube of the moon's orbital distance. Those are the hallmarks of tidal forces: They grow with the size of the body—the mass of the attractor—and fall off as the third power of distance.

- If we want to keep our "moon" intact, we need to supply a force to counteract ΔF: the gravitational force between the rocks, which are attracted to each other with a "self" gravitational force of

$$F_{\text{self}} = \frac{Gm^2}{(\Delta r)^2}.$$

- If we want them to stay together as they fall, the magnitude of ΔF must be smaller than F_{self}. That leads to an inequality that can be simplified as follows.

$$|\Delta F| < F_{\text{self}}$$

$$\frac{2GMm}{r^3}\Delta r < \frac{Gm^2}{(\Delta r)^2}$$

$$\frac{2M}{r^3} < \frac{m}{(\Delta r)^3}$$

- If m is big enough and Δr is small enough—that is, if the rocks are massive and closely packed—we can satisfy this inequality and they hold together as they fall.

- But there's a $1/r^3$ on the left side. As time goes on and r gets smaller, the left side grows rapidly and, at some point, overwhelms the right side. Let's call that minimum orbital distance r_{min}, which we can solve for by setting the 2 sides equal to each other. Inside of r_{min}, the moon will break apart.

$$\frac{2M}{r_{\text{min}}^3} = \frac{m}{(\Delta r)^3}$$

$$r_{\text{min}} = \left(\frac{2M}{m}\right)^{1/3} \Delta r \;\approx 1.26 \left(\frac{M}{m}\right)^{1/3} \Delta r \text{ (two rocks)}$$

- Our model of a moon as 2 point masses is not very realistic. You can do a similar calculation for a model in which the moon is a big blob of fluid that can deform and flow in response to tidal forces. That takes more mathematical horsepower, but the result is similar. In fact, it's the same as our equation, but with Δr representing the average radius of the moon, and the 1.26 is replaced by 2.44.

$$r_{\text{min}} \approx 2.44 \left(\frac{M}{m}\right)^{1/3} \Delta r \text{ (fluid body)}$$

- It's more traditional to write the equation in terms of the densities and sizes of the bodies, rather than masses. We can replace the planet mass, M, with volume times density, and we can do the same for the moon, m. With those substitutions and a touch of algebra, the minimum radius comes out to be 2.44 times the planet radius, times the cube root of the density ratio.

$$r_{\text{min}} \approx 2.44 \left(\frac{M}{m}\right)^{1/3} \Delta r \quad \begin{cases} M = \dfrac{4\pi R^3}{3}\rho_M \\[2ex] m = \dfrac{4\pi(\Delta r)^3}{3}\rho_m \end{cases}$$

$$r_{\text{min}} \approx 2.44 \left[\frac{R^3 \rho_M}{(\Delta r)^3 \rho_m}\right]^{1/3} \Delta r$$

$$r_{\text{min}} \approx 2.44\, R \left(\frac{\rho_M}{\rho_m}\right)^{1/3} \qquad \text{The Roche Limit}$$

- This was first worked out by Édouard Roche, so the minimum distance is known as the Roche limit: the distance within which tidal forces overcome the gravitational binding force of a moon modeled as an idealized fluid body.

- What if the moon is orbiting, instead of just falling in? Is there still a Roche limit? Yes. It's the same, and in that case, r_{min} refers to the minimum orbital distance.

- Suppose a moon made from a bundle of rocks is going in a circular orbit. The rocks closer to the star have slightly smaller orbital distances, so, by Kepler's third law, they have shorter orbital periods. The outer rocks have longer periods.

- Therefore, unless the moon's self-gravity is strong enough, the rocks drift apart over time, with the inner ones moving ahead of the outer ones. The moon gets shorn into pieces and strung out into an arc around the star. Eventually, the arc reaches all the way around the planet, making a ring. This may be where planetary rings come from!

- The number 2.44 is a pretty good match to the observed sizes of the rings of the giant planets, which range up to 2 or 2.5 times the radius of the planet. The density ratio is always of order unity, because the densities of the moons and planets are of the same order of magnitude, a few grams per cubic centimeter. So, the numbers fit the story.

All the planets have a Roche limit, including Earth. The Moon's mean density is around 3 grams per cubic centimeter, typical of rocks. The Earth's is higher, around 5.5, because the Earth's stronger gravity compresses its interior and because the Earth has more iron in its core. Given those numbers, the Roche limit comes out to about 3 Earth radii. Our Moon is at a distance of 60 Earth radii, so it's not in any danger of tidal destruction.

- We derived the Roche limit by setting the tidal force, which tries to pull the body apart, equal to the attractive gravitational force trying to hold it together. But there are other ways for a body to hold together besides gravity. There are also chemical or material forces that give rocks their rigidity. The silicon atoms in a rock aren't held in place by gravity; they're stuck together with chemical bonds, which are ultimately electromagnetic forces at the atomic level.

- The Roche limit is only relevant for objects that are held together mainly by gravity. And we shouldn't take the factor of 2.44 too seriously; that's for the ideal case of a frictionless fluid. Material forces allow a body to come closer than this official limit.

What happens when a body violates the Roche limit?

In 1994, a comet named Shoemaker-Levy-9 crashed into Jupiter. By the time it hit Jupiter, tidal forces had broken it into fragments, each of which punctured Jupiter's clouds in a different place, making a series of brown spots.

The comet broke apart because it was a loose conglomeration of rocks and chunks of ice; there wasn't much holding the chunks together besides gravity.

- Why can we find little moons, smaller than about 500 kilometers, nestled right within the rings that are inside the Roche limit? It must be because those objects are held together mainly by material forces, not gravity. In fact, you can tell if an astronomical object is gravitationally bound by just looking at it. If gravity is the dominant force, it'll be a sphere. That's because each piece is attracted to every other piece. Left to its own devices, gravity draws everything inward toward the center of mass, smoothing out any lumps to make a perfect sphere.

- In contrast, chemical and material forces are very local; they act only between neighboring molecules or surfaces in direct contact. So, a body held together by those forces can be any shape—an egg, a potato, a person—depending on the history of how the pieces came together.

- But if we make the object bigger and bigger, gravity becomes more important and eventually dominates over chemical and material forces.

- Let's do an order-of-magnitude calculation to estimate how big a body has to be for gravity to mold the shape into a sphere.

- Suppose we have a rock with a characteristic size of *R* and a mass of *M*. Now let's make it slightly larger by adding a single silicon atom. Which is more important: the molecular forces that bind the silicon to the minerals on the surface or the gravitational attraction of the atom to the entire mass of rock?

- It'll be easiest to compare the relevant amounts of energy. The energy levels of electrons in atoms are always on the of order of a few electron volts. The energy scale for material forces tends to be an order of magnitude lower, because rocks aren't perfect crystals—they're ragged collections of crystals, and the interactions between them are weaker. So, let's say the energy released is, on average, 0.1 eV per silicon atom.

- Meanwhile, the gravitational potential energy that's released when we add mass *m* is of order GMm/R. For gravity to be more important, we need that to be much larger than 0.1 eV. That gives us a condition on the rock's mass over radius, M/R. We're wondering about the critical size, so let's write *M* as volume times density and solve for *R*.

$$E_{\text{chem}} \sim 0.1\,\text{eV} \qquad E_{\text{grav}} \sim \frac{GMm}{R}$$

$$E_{\text{grav}} \gg E_{\text{chem}} \longrightarrow \frac{GMm}{R} \gg 0.1\,\text{eV}$$

$$\frac{M}{R} \gg \frac{0.1\,\text{eV}}{Gm} \qquad M = \frac{4\pi R^3}{3}\rho$$

$$\frac{4\pi}{3}R^2\rho \gg \frac{0.1\,\text{eV}}{Gm}$$

$$R \gg \sqrt{\frac{3}{4\pi}\cdot\frac{0.1\,\text{eV}}{Gm\rho}} \approx 600\,\text{km}$$

1.6×10^{-20} J

3000 kg m^{-3}

6.7×10^{-11} N m^2 kg^{-2}

4.7×10^{-26} kg

- The density of rock is around 3 grams per cubic centimeter, or 3000 kilograms per cubic meter, and the mass of a silicon atom is about 28 proton masses, or 4.7×10^{-26} kilograms. Plugging in those numbers along with the constants leads to a critical radius of about 600 kilometers.

- Based on this calculation, we would expect objects much larger than that to be sculpted into spheres by gravity, while much smaller objects can have irregular shapes. And this is what we observe among the moons of Saturn.

Why are the major rings of the giant planets all within about 2.5 planetary radii? Because that's the approximate location of the Roche limit.

Why are the large moons spherical? Because they're big enough for gravity to dominate over material forces.

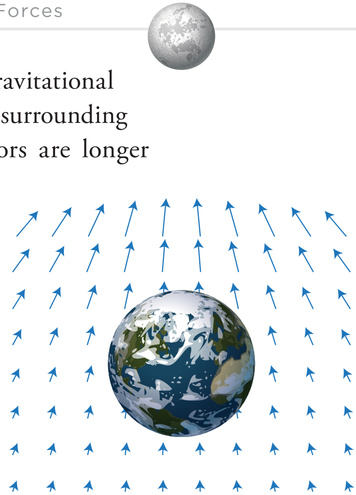
Why do the large moons orbit well outside of the rings? Because otherwise tidal forces would break them into smaller pieces.

OCEAN TIDES

- In addition to helping us understand rings and moons, tides are also relevant to planets, stars, black holes, and entire galaxies. A more down-to-earth example of tidal forces is ocean tides.

- The Earth's gravity pulls on the Moon, and the Moon's gravity pulls on the Earth. That means the Moon exerts tidal forces that, left unopposed, would tear the Earth apart by squeezing it along the direction to the Moon. The Earth's gravity prevents that from happening. But there's more to it than that.

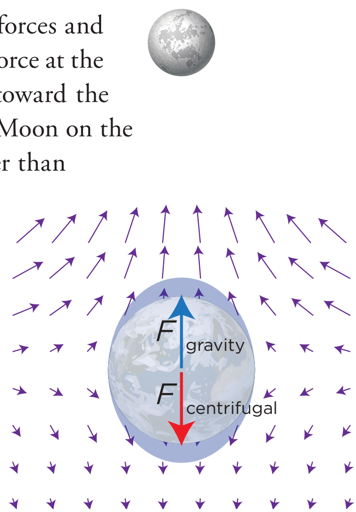
- The vectors in this image represent the gravitational force of the Moon at different points in space surrounding the Earth. Close to the Moon, the vectors are longer because the force is stronger. There's also some variation in direction, because all the vectors point straight at the center of the Moon.



- But importantly, the Moon is orbiting the Earth. To see the Earth and Moon sitting still, as they are in this image, we must be in a frame of reference that's rotating along with the orbit, once a month.

- We're allowed to do physics in rotating frames, but the price we pay is that we must insert a fictitious force: the centrifugal force. In this case, the centrifugal force on the Earth points away from the Moon, with a strength such that at the center of the Earth, the centrifugal force cancels out the gravitational force exactly. That's why the Earth is sitting still in this frame of reference.

- Let's add the centrifugal and gravitational forces and replot the net force vectors. There's no net force at the center of the Earth. The net force points toward the Moon on the near side and away from the Moon on the far side, where the centrifugal force is larger than the gravitational attraction. And in between there are sideways forces.

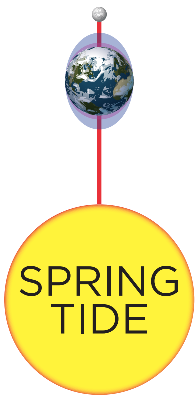


- Imagine that the Earth is a frictionless sphere surrounded by a thin layer of water. What would happen to the water? It would feel these net forces and flow around the surface to form 2 bulges, one on the near side and one on the far side.

- Then, if the frictionless Earth were to rotate, sliding underneath the layer of water, an observer on the surface would see the ocean rise in height, then fall, rise, and fall again over the course of a full day. That's why we observe 2 high tides and 2 low tides in 1 day.

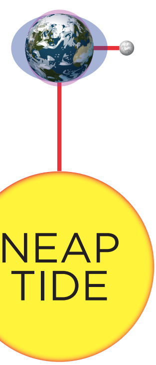
- The Earth is not a frictionless sphere. There's lots of friction, and there are continents, underwater mountains, and all kinds of things that make the picture more complicated. That's why we need tide tables.

- The tidal forces from the Sun are also significant. That's why the maximum height of the tide varies with the phase of the Moon. When the Sun, Moon, and Earth are along a line, the Sun and Moon work together and produce unusually high tides called spring tides. When the Sun and Moon are at right angles, the contrast between high and low tides is reduced, and they are called neap tides.



- From the relative heights of the spring and neap tides, we can determine that the Sun's tidal forces are not quite as strong as the Moon's—they're only about half as strong.

- Tidal forces are proportional to the mass of the attracting body over the cube of the distance, so the ratio of solar to lunar tidal forces is equal to the Sun's mass divided by the Moon's mass, times the cube of the distance to the Moon over the distance to the Sun.



$$\frac{\Delta F_{\text{sun}}}{\Delta F_{\text{moon}}} = \frac{M_{\text{sun}}}{M_{\text{moon}}} \left(\frac{d_{\text{moon}}}{d_{\text{sun}}} \right)^3$$

- Suppose we don't already know the masses and orbital distances. Can we learn something interesting about the Sun, or Moon, by observing spring and neap tides? Yes, we can—if we also know about total solar eclipses. The stunning thing about total eclipses is the Moon blots out the Sun almost exactly, rim to rim. They have the same angular radius in the sky.

- Because the angular radius, $\Delta\theta$, is equal to the true radius divided by distance, the observation of total eclipses tells us that

$$\Delta\theta = \frac{R_{\text{sun}}}{d_{\text{sun}}} = \frac{R_{\text{moon}}}{d_{\text{moon}}}.$$

- Or, equivalently, the ratio of distances is equal to the ratio of radii. So, in our tidal force equation, we can replace the cube of the distance ratio by the cube of the radius ratio.

$$\frac{\Delta F_{\text{sun}}}{\Delta F_{\text{moon}}} = \frac{M_{\text{sun}}}{M_{\text{moon}}} \left(\frac{R_{\text{moon}}}{R_{\text{sun}}}\right)^3$$

- This is the ratio of the average densities of the Sun and Moon!

- The fact that the Sun's tidal force is about half the Moon's tells us that the Sun's average density is half that of the Moon. The Moon looks like a rock, so its density is about 3 grams per cubic centimeter, from which we can deduce that the Sun's average density is around 1.5 grams per cubic centimeter.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 19.

Ryden and Peterson, *Foundations of Astrophysics*, chap. 4.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

QUIZ

LECTURES 1–6

1 In our neighborhood of the Milky Way, the typical spacing between stars is 1 pc, and the relative velocities between neighboring stars are of order 20 km/sec. Imagine a scaled-down model in which stars are replaced by grains of sand (1 mm). What is the typical spacing between the grains of sand, and how fast are they moving? [LECTURE 1]

CLICK to navigate.

2 Your weight on a planet is the product of your mass and the planet's surface gravitational acceleration, GM/R^2. Suppose you visit different planets that all have the same density. Write a proportionality to express how your weight (W) depends on planet mass (M). [LECTURE 1]

To go back to the page you came from, **PRESS Alt + ←** on a PC or **⌘ + ←** on a Mac.
On a tablet, use the bookmarks panel.

3 Make a list that compares the 4 fundamental forces of nature. Think of at least one way in which each force has a direct impact on everyday life. [LECTURE 2]

4 Suppose you start out with the mass of an electron and double your mass every day. How many days would elapse before you have the mass of the entire Milky Way Galaxy (approximately 1 trillion solar masses)? [LECTURE 2]

5 Try to measure the angular resolution of your eye in arc seconds. One way is to prick 2 holes in a piece of cardboard a few millimeters apart and illuminate them from the back with a flashlight. Then, in a large, dark room, see how far away you can still tell that there are 2 holes instead of one. [LECTURE 3]

6 The flux we receive from a star is proportional to $1/r^2$. But the flux we receive from a distant asteroid is approximately proportional to $1/r^4$. Why? What is the key difference? [LECTURE 3]

7 Think about the problem of establishing Kepler's laws of planetary motion based on naked-eye observations of the night sky. What would be some of the major difficulties? [LECTURE 4]

8 Neptune's moon Triton has an orbital period of 5.877 days and a semimajor axis of 354,759 km. Calculate the mass of Neptune in units of Earth masses. [LECTURE 4]

9 If a spacecraft in a circular orbit fires its rocket, boosting the speed of the spacecraft in the forward direction, how does the total orbital energy change? How does the angular momentum change? What are the resulting changes to the shape of the orbit? [LECTURE 5]

10 Halley's comet has an orbital period of 75.32 years and an orbital eccentricity of 0.96714. How close does Halley's comet come to the Sun in AU? [LECTURE 5]

11 Some have argued that the definition of a planet should include the requirement that the shape is a sphere. What are the merits and faults of this definition? Think about which objects in the solar system would be classified as planets and how this definition could apply to planets around other stars. [LECTURE 6]

12 Express the Roche limit as a minimum orbital period, rather than a minimum orbital distance. You should find that the minimum orbital period is independent of the properties of the central body. [LECTURE 6]

CLICK to navigate.

Go to page 336 for solutions.

Lecture 7

BLACK HOLES

In the popular imagination, black holes are like vacuum cleaners, sucking up everything around them. That's not quite true—or at least it's no truer than it is for the Sun. The Sun doesn't suck in all the planets. To fall into the Sun, or a black hole, you need to have nearly 0 angular momentum—hardly any transverse velocity—so that you follow a radial trajectory all the way down to $r = 0$. Black holes do swallow things, but that only happens when there's some way for orbiting material to get rid of its angular momentum.

The Gravitational Field of a Spherical Object

Newton's law of gravity, $F = GMm/r^2$, gives the magnitude of the force between 2 masses. Specifically, it's the force between 2 point masses—idealized mathematical points with nonzero mass but zero size.

When we allow the body feeling the force to have a nonzero size, it feels tidal forces, differential gravitational forces that tend to pull the body apart. But what happens when we allow the body producing the force—such as the Earth—to have a nonzero size?

The Earth isn't a point mass; it's a sphere. When we're standing on the surface, our bodies are gravitationally attracted to every cubic centimeter of material in the entire Earth, pulling all at once. And the rock just below our feet is much closer to us, and pulls harder, than the same amount of rock on the other side of the Earth.

So, you might think it would be a mess to calculate the net force from the entire Earth; you'd have to add the forces exerted by each cubic centimeter, taking into account the huge variations in distance and direction from place to place.

But Isaac Newton taught us that when we're outside a spherically symmetric object, we can calculate the gravitational force by pretending it's a point mass endowed with the total mass of the real object.

What happens if we're inside an object with a spherically symmetric mass distribution? Let's say we dig a tunnel in the Earth down to a radius r from the center. It turns out the net gravitational force at the bottom of the tunnel can be calculated by pretending that all the mass down deeper, with radius less than r, is replaced by a point mass at the center and by ignoring all the exterior mass! This is because although the nearby atoms are pulling harder, there are more atoms farther away pulling in the other direction, and the effects cancel each other.

Therefore, if we were inside the Earth, we'd feel a gravitational force, which we can calculate by pretending all the interior mass is concentrated at the center. If the Earth had a constant density, ρ, the total interior mass would be

$$M_{\text{int}} = \frac{4\pi}{3} r^3 \rho.$$

The gravitational force is proportional to the interior mass divided by r^2, so it goes as r^3/r^2—it's proportional to r. As we dig further down, the force declines in strength, and at the center of the Earth, there's no gravitational force at all.

$$F = \frac{GM_{\text{int}}m}{r^2} = \frac{G4\pi r^3 \rho m}{3r^2} \propto r$$

THE MASS AND RADIUS OF A BLACK HOLE

- A black hole has no surface; it's a real-life point mass. Its mass is concentrated into a single point in space, so as you approach, the gravitational force gets stronger, without bound—all the way down to $r = 0$, where it becomes infinite.

- A black hole is an infinitely deep gravitational pit. And the pit has slippery sides; if you get too close, you inevitably fall in.

- Let's say we're cruising around the galaxy and find ourselves a distance of r_0 from a black hole. In a panic, we turn away from the hole and fire our thrusters, burning all the fuel in a desperate burst and giving the spaceship a velocity of v_0 directed away from the hole. Will we escape?

- We can figure this out based on the conservation of energy. The initial energy of the spaceship is the kinetic energy plus the gravitational potential energy, where M is the mass of the black hole and m is the mass of the ship.

$$E = \frac{1}{2}mv_0^2 - \frac{GMm}{r_0}$$

- At some later time, we make it out to some larger distance, r_1, with a slower speed, v_1, because the black hole's gravity has been slowing us down. So, the energy is

$$\frac{1}{2}mv_1^2 - \frac{GMm}{r_1}$$

which we can set equal to the initial energy.

$$\begin{aligned} E &= \frac{1}{2}mv_0^2 - \frac{GMm}{r_0} \\ &= \frac{1}{2}mv_1^2 - \frac{GMm}{r_1} \longrightarrow 0 \end{aligned}$$

- In the case where we just barely escape, r_1 approaches infinity and the potential energy approaches 0. In addition, v_1 approaches 0, because we had just enough energy to make it out, with no leftover kinetic energy. So, the total energy in this case must be 0, giving us a simple equation, which we can solve for v_0, giving

$$\begin{aligned} \frac{1}{2}mv_0^2 - \frac{GMm}{r_0} &= 0 \\ \frac{1}{2}mv_0^2 &= \frac{GMm}{r_0} \\ v_0 &= \sqrt{\frac{2GM}{r_0}}. \end{aligned}$$

- That's the formula for the escape velocity. If we have enough fuel to go at least that fast, we can escape. Otherwise, we fall in.

- But here's the thing about a black hole: Because there's no surface, there's no minimum value for r_0. There's nothing equivalent to the Earth's surface, within which the gravitational force starts getting weaker. So, the escape velocity grows without bound as r_0 approaches 0.

- At some point, the escape velocity exceeds the speed of light—the fastest speed it's possible for anything to attain—and in that case, no amount of fuel would be enough. To find the value of r_0 where that happens, we set the escape velocity equal to c, the speed of light, and solve for r_0.

$$v_0 = \sqrt{\frac{2GM}{r_0}} = c \longrightarrow r_0 = \frac{2GM}{c^2} = R_S$$

- So, even though the black hole itself has 0 size, $2GM/c^2$ is sort of a radius—the radius of no return. It's called the event horizon or the Schwarzschild radius (R_S), after Karl Schwarzschild, the first person to solve Einstein's equations of general relativity exactly for the case of a point mass.

- Let's find the Schwarzschild radius of a black hole with the mass of the Earth ($M_\oplus$). In this case, $M_\oplus$ is 6×10^{24} kilograms, giving a Schwarzschild radius of 9 millimeters. What that means is that if we could somehow compress the entire Earth down to the size of a marble, we would have a black hole.

$$R_S(\text{Earth}) = \frac{2\overset{6.7\times10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}}{\overset{\downarrow}{G}}\overset{6\times10^{24}\,\mathrm{kg}}{\overset{\downarrow}{M_\oplus}}}{\underset{3\times10^{8}\,\mathrm{m\,s^{-1}}}{\underset{\uparrow}{c^2}}} \approx 0.009\,\mathrm{m}$$

- For the mass of the Sun ($M_\odot$), the Schwarzschild radius comes out to be 3 kilometers, so we can write the formula as a scaling relation.

$$R_S = 3\,\mathrm{km}\left(\frac{M}{M_\odot}\right)$$

Imagine that the Sun is somehow compressed to a radius of 3 kilometers, turning it into a black hole. What would happen to the solar system? How long would we have before Earth fell into the black hole?

Actually, the Earth wouldn't fall in. And the planets' orbits wouldn't change. It's a corollary of Newton's theorem about spherically symmetric mass distributions, which turns out to also hold true in relativity. The force from the Sun is the same as the force from a black hole of equivalent mass.

EVIDENCE FOR BLACK HOLES

- Since the mid-1990s, a group at UCLA led by Andrea Ghez has been watching a cluster of bright stars near the constellation Sagittarius in the center of our galaxy. These stars are moving fast, and not just in random directions. They're being deflected by the gravity of a giant mass at the center. Some stars are orbiting the center in Keplerian ellipses.

- One star in particular, named S0-2, made a complete orbit over 16 years. But in optical and infrared images, there's hardly any light coming from the focus of the ellipse. S0-2 and the neighboring stars are being attracted to an unremarkable spot in the center.

- Let's figure out how massive the attractor must be based on observations of S0-2. The images of S0-2 show that the angular size, $\Delta\theta$, of the long axis of the ellipse is about a quarter of an arc second. And we know from other observations that the distance, d, to the galactic center is 8 kiloparsecs. With that information, we can calculate the semimajor axis, a.

$$\Delta\theta = \frac{2a}{d}$$

$$a = \frac{1}{2}\Delta\theta\, d = \frac{1}{2}\left(\frac{1}{4}\,\text{arcsec}\right)(8{,}000\,\text{pc}) = 1000\,\text{AU}$$

- We also observe that S0-2 takes 16 years to go around, so $P = 16$ years. And whenever we know both P and a, we can apply Kepler's third law to find the mass. First, let's rearrange Kepler's third law to solve for the mass.

$$P = \frac{2\pi}{\sqrt{GM}} a^{3/2} \longrightarrow M = \frac{4\pi^2}{G} \frac{a^3}{P^2}$$

- At this stage, we could plug in numbers, looking up the numerical value of G, converting a to meters and P to seconds, and so forth, but let's make it easier by converting the equation into a scaling relation.

- Because M is proportional to a^3/P^2 and we know the answer is 1 solar mass for the Earth, we can write

$$M = (1\,M_\odot) \left(\frac{a}{1\,\mathrm{AU}}\right)^3 \left(\frac{P}{1\,\mathrm{year}}\right)^{-2}.$$

- This is convenient for our problem because we already know a and P in those units. The mass of the mysterious attractor, in units of solar masses, is

$$\begin{aligned} M &= (1\,M_\odot) \left(\frac{a}{1\,\mathrm{AU}}\right)^3 \left(\frac{P}{1\,\mathrm{year}}\right)^{-2} \\ &= (1\,M_\odot)\,(1000)^3\,(16)^{-2} \\ &\approx 4 \times 10^6\,M_\odot\,. \end{aligned}$$

- In other words, lurking at that nondescript spot in the center—named Sagittarius A*—is something with a mass of 4 million Suns. And it's all crammed into a space that must be smaller than S0-2's distance of closest approach, about 100 AU, which is not much bigger than the solar system.

- Astrophysicists have not been able to think of anything that could have so much mass within such a small space without either glowing very brightly or quickly collapsing under its own gravity. So, it's either a black hole or something even more exotic, beyond our current physical understanding.

- How close are these stars getting to the Schwarzschild radius? Are they in danger of falling in? The Schwarzschild radius is about 12 million kilometers—which is only about 0.1 AU. The orbit of S0-2 has a minimum distance of 100 AU, so it's safe.

$$R_\mathrm{S} = 3\,\mathrm{km}\left(\frac{M}{M_\odot}\right) \approx 12 \times 10^6\,\mathrm{km} \approx 0.1\,\mathrm{AU}$$

OPTICAL AND RADIO TELESCOPES

- Every once in a while, a star or cloud of gas falls into a black hole. The evidence for tidal disruption events takes the form of bright outbursts from the centers of distant galaxies, but those galaxies are all so far away that our telescopes can't resolve the details, so there's no hope of making images that show the star falling in.

- In our own Milky Way, could we watch a star getting torn to shreds? Let's figure out how big a telescope we would need.

- The fundamental limit on the angular resolution of an image, $\Delta\theta_{\mathrm{min}}$, is the diffraction limit, which is of order λ/D, where λ is the wavelength of light and D is the diameter of the telescope.

$$\Delta\theta_{\mathrm{min}} \sim \frac{\lambda}{D}$$

- How big does D need to be to resolve details on the scale of the Schwarzschild radius, 0.1 AU? We can find out by setting λ/D equal to the angular size of the Schwarzschild radius. That's 0.1 AU divided by 8000 parsecs, the distance to the galactic center. For λ, we insert half a micron, which is typical of visible light. After the necessary unit conversions, the diameter comes out to be 8000 meters.

- The world's largest optical telescopes are 10 meters across; in this case, we would need one that is 8 kilometers across. That is crazy. Nevertheless, there's an effort underway to make images sharp enough to resolve the Schwarzschild radius of Sagittarius A*. The trick is to use a radio telescope instead of an optical telescope. The project is called the Event Horizon Telescope.

- At radio wavelengths, it's possible to combine the information from widely separated telescopes to mimic the performance, in terms of angular resolution, of a much larger telescope. So, if we connect telescopes in different parts of the world, we can get the necessary angular resolution—at least in principle. This technique, called very long baseline interferometry, has become routine for observations at wavelengths of meters and centimeters, but getting it to work at a millimeter is still a challenge.

ORBITING A BLACK HOLE

- If you get too close to a black hole, within the Schwarzschild radius, you fall in, and if you're far away, the orbits are the same as they would be around a star, or planet, or any "normal" astronomical body. But what happens in between those extremes—when you're orbiting near a black hole but not right at the Schwarzschild radius?

- To answer this question, we'll use the effective potential energy, a concept we used previously to analyze planetary motion. In Newtonian gravity, the equation for total energy is as follows, where U_{eff} is the sum of the real potential energy and a term that depends on angular momentum but can be written in a form that makes it resemble a type of potential energy.

$$E = \frac{1}{2}mv_r^2 + U_{\text{eff}}(r)$$

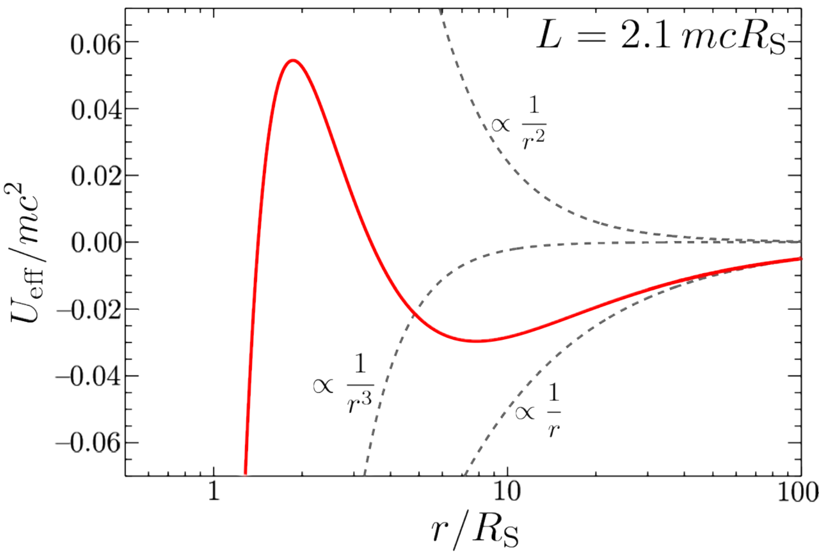
$$U_{\text{eff}}(r) = -\frac{GMm}{r} + \frac{L^2}{2mr^2}$$

- This gives us a graphical way to understand how the radial coordinate changes with time as a body whirls around an attractor.

- In general relativity, there is an extra term in the effective potential energy.

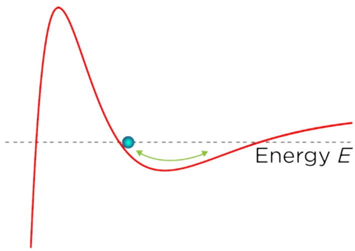
$$U_{\mathrm{eff}}(r) = -\frac{GMm}{r} + \frac{L^2}{2mr^2} - \frac{L^2 R_S}{2mr^3}$$

- When we replot U_{eff} for a black hole, we need to add the contributions of the $-1/r$ term, the $1/r^2$ term, and the $-1/r^3$ term. As r gets smaller, the new $1/r^3$ term grows the fastest of all, and it's negative, so the effective potential dives downward at a small radius.

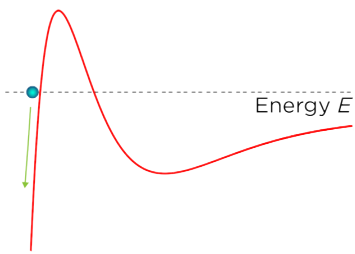


- To make the numbers come out nice, the vertical axis shows U_{eff}/mc^2 and the horizontal axis shows the radius in units of the Schwarzschild radius. For concreteness, a value of angular momentum equal to $2.1mc$ times the Schwarzschild radius was used.

- Below 1 Schwarzschild radius, there's a pit in the potential. We can visualize what's going to happen to the radial coordinate r by imagining we drop a marble onto this curve starting from an initial height E, given by the total energy.

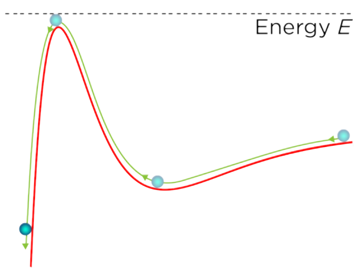


- If the marble is far away from the Schwarzschild radius, then it rolls back and forth. That means r oscillates back and forth, as was the case in Newtonian gravity; the particle follows an orbit with some minimum and maximum distance.



- But if we let go of the marble inside the Schwarzschild radius, it falls into the pit, regardless of energy.

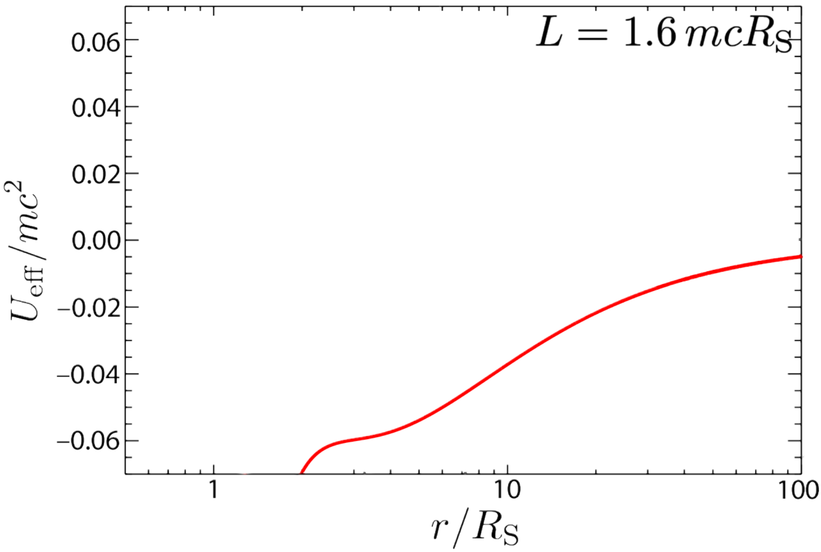
- And even if we start the marble farther out in the bowl, if the energy is high enough, it will go over the inner wall of the bowl and fall into the pit.



- This means that in general relativity, you can hit the origin even if you do have some nonzero angular momentum. There's no more barrier here, sealing off the singularity, as there was in Newtonian gravity. It's as though when you get close enough, the hole reaches out and grabs you. It absorbs both your mass and your angular momentum. In other words, black holes—rotating point masses—can rotate.

- This curve was drawn for a specific value of angular momentum. Different values lead to different shapes for the bowl.

- As we lower the angular momentum, the bowl gets pounded down on the left side until it's not even a bowl any more. There's no more minimum, which means there's no place where the particle can sit still at a constant radius. A circular orbit is impossible.



- This is a famous result in general relativity. It's called the innermost stable circular orbit (ISCO). For a nonrotating black hole, the ISCO turns out to be 3 times the Schwarzschild radius. For a rotating black hole, the ISCO can be larger or smaller, depending on which way it's rotating.

- Let's return to a case with more angular momentum, in which the effective potential energy curve still looks like a bowl. If you give the particle a modest amount of energy, the radial coordinate oscillates back and forth, just as it does for a planet going around the Sun.

- For a planet, the trajectory is an elliptical orbit with a minimum distance of $a(1 - e)$ and a maximum distance of $a(1 + e)$—but we only get ellipses for the very specific case of a force that goes as $1/r^2$. And in general relativity, the force law doesn't go exactly as $1/r^2$, due to that extra term in the potential. So, we don't get perfect ellipses.

- Close to the black hole, the orbits are not even approximately ellipses. They're rosettes—whipping around fast as they approach the black hole, then slowing down as they recede, and coming back again from a different angle.

$$\vec{e} = \frac{\vec{v} \times \vec{L}}{GMm} - \hat{r}$$

- If you're relatively far away, the orbits are very nearly ellipses—but not quite. The orientation of the ellipse gradually wheels around in space. It's an effect called apsidal precession.

- In general relativity, the eccentricity vector—that combination of velocity, momentum, and position that turns out to be conserved in classical gravity—is no longer conserved. It slowly changes with time. It rotates around the direction of the angular momentum vector.

- So, if you looked carefully enough at planets orbiting a black hole, you would see that they're not quite obeying Kepler's first law. And it's not just black holes. Once you're outside a spherical mass distribution, the gravitational effects are the same, no matter whether you're orbiting a planet, star, or black hole.

- That means Kepler's first law must be at least a little bit wrong in the solar system, too. The planetary orbits should be precessing. And they are. The effect is strongest for Mercury, the planet with the smallest orbital distance.

READINGS

Blundell, *Black Holes*.

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 17.

Thorne, *Black Holes and Time Warps*.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 8

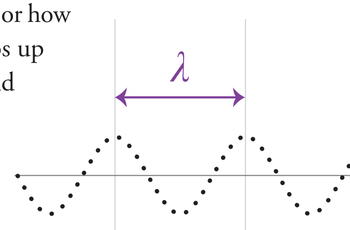
PHOTONS AND PARTICLES

When it comes to the task of solving the mysteries of the cosmos, the only clues that astrophysicists get—with few exceptions—are from the trickle of radiation that arrives at the Earth from distant sources. In this lecture, you will become familiar with electromagnetic radiation.

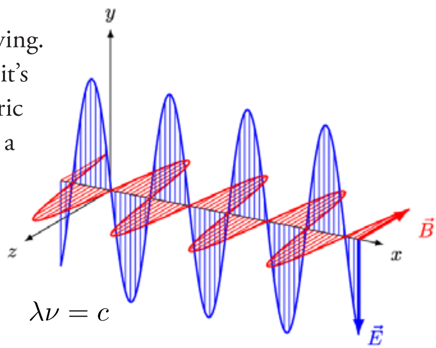
ELECTROMAGNETIC WAVES

- In classical physics—that is, not quantum physics—electromagnetic radiation takes the form of waves. Waves result whenever some physical quantity varies smoothly throughout space and time—such as the height of the surface of a pond or the pressure of the air in a room—that, when disturbed, produces an oscillating pattern.

- No matter what kind of wave it is, the wavelength, λ, can be defined as the distance between maximum values of whatever is waving—for example, between the crests of a water wave. And the frequency, v, can be defined as the rate at which the pattern oscillates, or how many times per second the height of the water bobs up and down, completing a cycle. When wavelength and frequency are multiplied, the result is the wave's phase velocity—the speed with which the pattern moves—which, for electromagnetic radiation, is the speed of light, c.

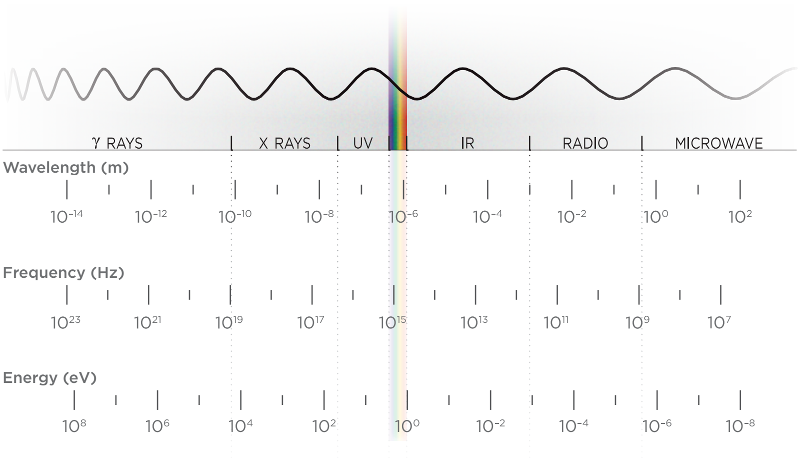


- But with light, it's less obvious what's waving. It's not the air, or anything material; it's the more abstract oscillations of electric and magnetic fields. The electric field is a vector whose magnitude wobbles back and forth, while the magnetic field does the same thing, but tilted by 90°—and the whole pattern moves in the direction perpendicular to both vectors.



- The wavelengths of visible light range from 0.4 to 0.7 microns. The corresponding frequency, v, is just under 10^{15} cycles per second, or Hertz. But that's just a tiny slice of the whole spectrum. Toward longer wavelengths, there's infrared radiation, and when the wavelength exceeds 1 millimeter, they are called microwaves, and then radio waves. Likewise, on the short-wavelength end is ultraviolet radiation, and when the wavelength is shorter than 10 nanometers, they are called x-rays, and then gamma rays (see image on the following page).

- All of these names and boundaries are arbitrary; they're based on the different technologies humans use to study electromagnetic radiation. Mother Nature makes no sharp distinctions.



- Waves are produced by a disturbance, such as dropping a stone in water or clapping your hands. For electromagnetic waves, the disturbance is the acceleration of an electrically charged particle.

- When a charge accelerates, it radiates. And it goes backward, too: Charges can absorb electromagnetic energy, causing them to accelerate.

- These facts come from solving James Clerk Maxwell's equations of electromagnetism.

- The pattern and energy of the waves depend on the details of the acceleration. An electron slowing down in a block of lead produces a certain pattern while an electron whirling in a magnetic field produces a different one. And an electron falling from one orbit to another inside an atom produces yet another kind of radiation.

- But if the charges are moving randomly—if there are zillions of charges rattling around, colliding with each other, and they've been at it long enough to reach a steady state—there's an enormous simplification. The radiation takes on a universal character, depending only on the temperature.

INDIVIDUAL PROPERTIES OF PARTICLES AND PHOTONS

- Quantum theory teaches us that if we look closely enough at electromagnetic energy, we'll see it comes in tiny lumps, called photons. In principle, we could count the photons arriving from the Sun—like counting raindrops falling from the sky or grains of sand in an hourglass—but photons have some weird properties that normal particles do not have.

- Think of a box full of tiny particles whizzing around and knocking into each other. This is an idealized model of a gas. Each particle has a mass, m, and a speed, v—from which the particle's kinetic energy and momentum can be computed. (The symbol ε is used for the energy of a single particle to distinguish it from the energy of the entire gas, E.)

	PARTICLES	PHOTONS
MASS	m	zero
SPEED	v	c
ENERGY	$\epsilon = \frac{1}{2}mv^2$	$\epsilon = h\nu = \frac{hc}{\lambda}$
MOMENTUM	$p = mv$	$p = \frac{\epsilon}{c} = \frac{h\nu}{c}$

- Photons have 0 mass. In addition, they always travel in empty space at the same speed, c, which is equal to 3×10^8 meters per second. And even though they have 0 mass, photons have energy and momentum. (The constant of proportionality, h, is Planck's constant, which is equal to 6.6×10^{-34} joule-seconds.)

COLLECTIVE PROPERTIES OF PARTICLES

- Let's say the box of particles has a volume V and is filled with a certain number of particles, N. The number density, n, is defined as the number per unit volume.

NUMBER DENSITY

$$n = \frac{N}{V}$$

- Into the box a total energy of E is injected. That's the sum of all the kinetic energies of all the particles—which is constant in time, because the box is sealed tight. If the particles are allowed to knock around for a long time, their positions and velocities become randomized; a particle could turn up anywhere in the box with equal probability. And they'll come to share the total energy more or less equally.

- That's the conceptual basis of temperature. The temperature, T, is defined to be proportional to the average energy per particle. For an ideal gas, the proportionality constant turns out to be $3/2k$, where k is Boltzmann's constant, which is equal to 1.4×10^{-23} joules per Kelvin.

- The general rule is that the average energy is $1/2kT$ times the number of independent ways a particle can store or exhibit energy—the number of degrees of freedom. Our particles can move in 3 dimensions, so the kinetic energy has 3 terms, each of which counts as a degree of freedom.

AVERAGE ENERGY

$$\langle \epsilon \rangle = \frac{3}{2}kT$$

$$\epsilon = \frac{1}{2}mv^2 = \frac{1}{2}m\left(v_x^2 + v_y^2 + v_z^2\right)$$

- So, the average energy per particle is $3/2kT$.

- This means that the energy density, u, which is the total energy per unit volume, is equal to 3/2 times the number density, n, times kT.

ENERGY DENSITY

$$u = \frac{3}{2}nkT$$

- So, the temperature of a gas is a scale for the energy associated with the random motion of the particles.

- In addition to energy, the particles have momentum. And the scale for that is pressure. The particles are constantly knocking into the walls of the box, or any surface that might be inserted in the gas. Those collisions exert a force on the surface—that's pressure.

PRESSURE

$$P = nkT$$

- To simplify the math, let's imagine a universe in which particles can move in only one direction (back and forth).

- When a particle with speed v hits the wall, it reflects back with speed v in the opposite direction. Its momentum changes from $+mv$ to $-mv$, a change of $-2mv$. Because momentum is conserved, the wall must have absorbed a momentum of $+2mv$. It feels a push. And that keeps happening as more particles hit the wall. In a time Δt, how much momentum does the wall absorb?

- Let's say all the particles have the same speed, v, but half are moving right and half are moving left. The particles that hit the wall are the ones moving to the right that start within a distance of $v\Delta t$ from the wall.

- If we focus on an area ΔA of the wall, that singles out a box of volume $v\Delta t\Delta A$. So, the total momentum absorbed by the wall will be the momentum from each collision ($2mv$) times the number of collisions ($n/2$, the number density of particles moving to the right) times the volume of the box, $v\Delta t\Delta A$.

$$\Delta p = (2mv)\left(\frac{n}{2}\right)(v\Delta t\Delta A) = nmv^2\Delta t\Delta A$$

- Force is momentum per unit time and pressure is the force per unit area, so to get the pressure, we divide our equation by Δt and ΔA, giving nmv^2. And because mv^2 is twice the kinetic energy, ε, we can also write it as $2n\varepsilon$.

$$P_v = \frac{\Delta p/\Delta t}{\Delta A} = nmv^2 = 2n\,\frac{1}{2}mv^2 = 2n\epsilon$$

- We just assumed that all the particles have the same speed, v, but that's not true. There is a range of speeds. So, we should replace ε by its average value, which would be $1/2kT$ in a 1-dimensional universe.

$$P = 2n\langle\epsilon\rangle = 2n\,\frac{1}{2}kT$$
$$P = nkT$$

If you have a background in chemistry, you might be used to seeing $P = nkT$ as $P = nRT$. These equations are the same. The R is just k in chemists' units; they like to count particles in units of moles. Physicists tend to just count the particles individually.

- We just derived the ideal gas law: Pressure is proportional to number density and temperature. We did it for a 1-dimensional gas, but in 3 dimensions, you end up getting the same equation.

- Suppose we pop a tiny hole of area A in the wall. Gas particles will start leaking out and the gas will lose energy. What's the rate of energy loss?

- In our 1-dimensional universe, the number of particles that leak out in time Δt is equal to the number density of right-moving particles, $n/2$, times the volume of that same box, $v\Delta t\Delta A$. Each one has energy ε, giving

$$\Delta E = \epsilon\left(\frac{n}{2}\right)(v\Delta t\Delta A).$$

- Let's divide by ΔA and Δt to give the power per unit area—the flux—of escaping energy.

FLUX

$$F \propto n\langle v\epsilon\rangle \propto T^{3/2}$$

- Again, because there's a range of speed and energies, we need to take the average. The important thing is that it's proportional to $nv\varepsilon$, which also turns out to be proportional to $T(3/2)$.

$$F = \frac{\text{power}}{\text{area}} = \frac{\Delta E/\Delta t}{\Delta A} = \frac{1}{2}n\langle v\epsilon\rangle$$

COLLECTIVE PROPERTIES OF PHOTONS

- Photons, unlike particles, do not collide with each other; they sail right through each other. The only things photons interact with are charged particles. In order to randomize the positions and energies of photons, you need to have charged particles—which are themselves in thermodynamic equilibrium—around. So, we'll fill our box with charged particles, colliding all the time, producing momentary accelerations and thereby producing and absorbing photons.

- For an ideal gas, the average energy per particle is $3/2kT$. For the photons, it turns out to be $2.7kT$. That's not so weird. They're both proportional to temperature; it's just a different numerical constant in front.

AVERAGE ENERGY

$$\langle\epsilon\rangle \approx 2.7\,kT$$

- Things get weird, though, with the number density. For an ideal gas in a closed box, n is constant; the particles don't spontaneously appear out of nowhere or vanish. Even if we add energy, speeding up the particles, their number stays the same.

NUMBER DENSITY

$$n \approx \left(\frac{3.9\,kT}{hc}\right)^3$$

- But photons do appear out of nowhere—whenever a charged particle accelerates. The particle flings away some of its own energy in the form of photons. Likewise, a photon vanishes when its energy is absorbed by a charged particle.

- So, for photons, we shouldn't expect n to be a constant. If we inject energy into the gas, speeding up the particles, the magnitude of their accelerations will rise and they will produce more photons. It turns out that the number density of photons rises as temperature to the third power.

ENERGY DENSITY

$$u = \frac{4\sigma}{c}T^4$$

- For the ideal gas, energy density, u, is proportional to nT. For photons, n varies as T^3, and if energy density varies as nT, then we might expect the energy density of photons to go as T^4. And it does. The constant of proportionality is traditionally written $(4\sigma)/c$, where σ is the Stefan-Boltzmann constant.

PRESSURE

$$P = \frac{4\sigma}{3c}T^4$$

- For the gas, pressure equals nkT, the ideal gas law. For photons, again, n goes as T^3, so we might expect pressure to go as T^4, and it does. In this case, the proportionality constant is $(4\sigma)/3c$.

- For the gas, flux—the power per unit area that would emerge from a tiny hole in the box—is n times the average of $v\varepsilon$, which is proportional to $T^{3/2}$. For photons, n scales with T^3, v is always c, and ε is proportional to kT, so we might guess that flux is proportional to T^4, and it is.

We never notice radiation pressure because at ordinary temperatures, it's negligible. In principle, though, when you're at the beach, the sunlight hitting your body pushes you down into the sand—a little. Inside really massive stars, though, the radiation pressure can be so large that it blows the star apart.

- This result—the flux of electromagnetic radiation from a body at temperature T is proportional to T^4—is important enough to deserve its own name: the Stefan-Boltzmann law.

FLUX

$$F = \sigma T^4$$

- The constant of proportionality is σ, the one that also appeared in the equations for energy density and pressure. It's not a new fundamental constant; it's a certain combination of h, c, and k. But it occurs so frequently that the abbreviation is helpful.

$$\sigma = \frac{2\pi^5 k^4}{15h^3c^2} = 5.7 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$$

COMPARING DISTRIBUTIONS OF ENERGIES

- For the case of the gas, the average energy is $3/2kT$. If we pick a particle at random, we expect its energy to be about $3/2kT$—but not exactly—depending on its recent history of collisions. Likewise, the speed of a given particle is always fluctuating.

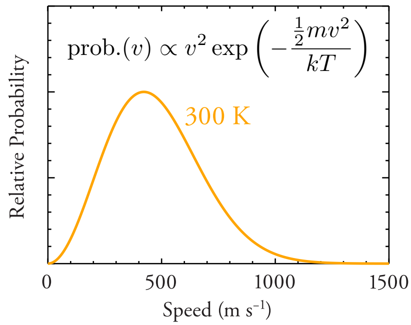
$$\langle \epsilon \rangle = \frac{3}{2}kT$$

- A fundamental rule that emerges from classical statistical physics is that the probability to find a particle in a state with energy ε is proportional to $e^{-\varepsilon/kT}$, an exponential function called the Boltzmann factor.

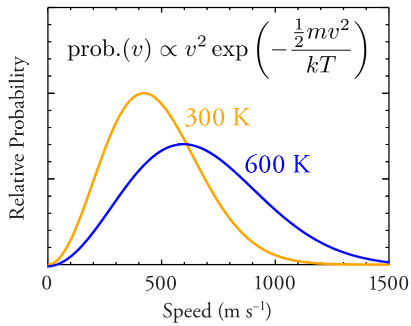
$$\text{Boltzmann factor: prob. (state)} \propto e^{-E/kT}$$

- This means that the energy will almost always be of order kT. Much larger energies are vanishingly rare, because of the exponential falloff. The particles tend to share the energy equally; there's little chance that one particle is going to end up with a disproportionate share of the total energy.

- From that basic rule, it's possible to derive the probability distribution for the energy, or the speed, of a particle in an ideal gas. It's called the Maxwell-Boltzmann distribution, and it's the product of the Boltzmann factor and a factor of v^2 that comes from counting the number of possible states for the particle to have energy E.

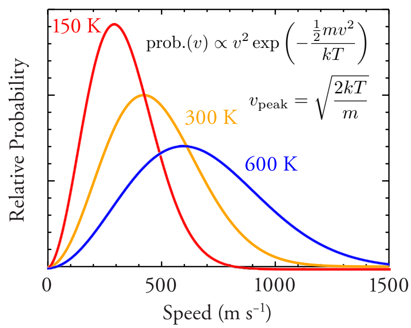


- The horizontal axis is speed (in meters per second), and the vertical axis is the relative probability that you'll find a particle to have that speed; in other words, it's the fraction of particles that have that speed at any given time.



- The function depends on the particle mass and the temperature. For this chart, the case of nitrogen molecules at 300 Kelvin—basically, air at room temperature—was chosen. The distribution rises from 0 to a peak at around 400 meters per second and then falls off again.

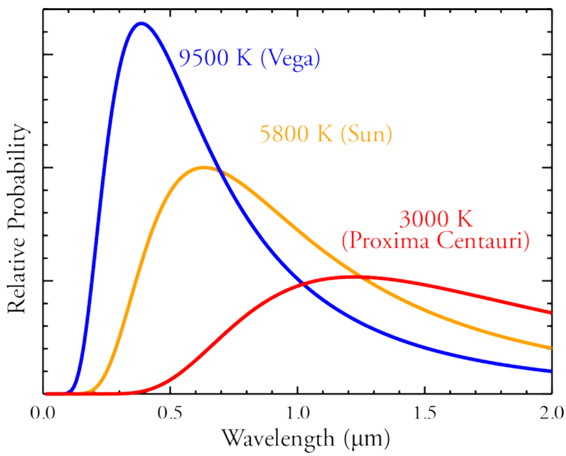
- If the temperature is increased, the peak spreads out and moves to higher velocities.



- If the temperature is decreased, the peak moves to the left. In general, the most probable speed is $\sqrt{2kT/m}$.

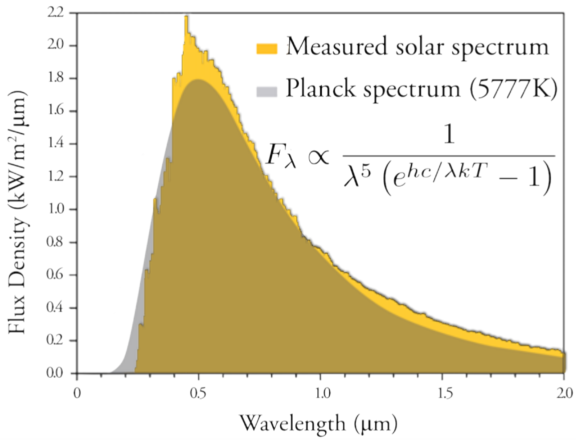
- Let's switch to photons. Imagine trapping some photons from the Sun in a reflecting box and measuring the wavelength of each one. Once we collect enough photons, we find that there's a peak at around 0.6 microns. That's the most popular wavelength to have. The shape of the function looks like the Maxwell-Boltzmann distribution, but it's different in detail, because photons are not your everyday particles. It is called a Planck spectrum.

- One of the curves is for 5800 Kelvin, approximately the temperature of the Sun's outer layers. The bright star Vega is hotter—closer to 9500 Kelvin—so its spectrum is shifted toward higher energies, which means shorter wavelengths. And the faint, nearby star Proxima Centauri is only about 3000 Kelvin, so its photons generally have lower energies and longer wavelengths.



- If we pop a tiny hole in our box, the total flux that leaks out is σT^4. But the contributions to that flux from photons of various wavelengths is the flux density—power per unit area per unit wavelength.

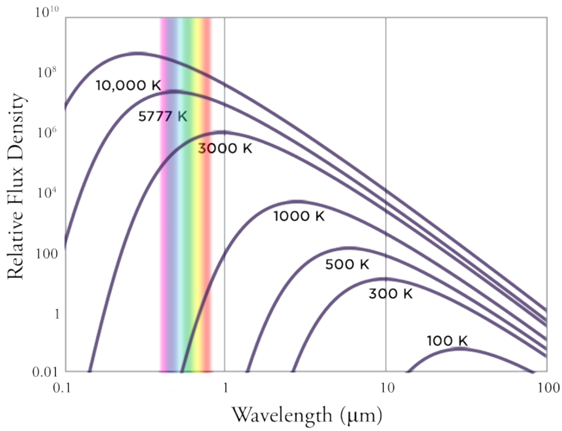
- When we measure the flux density of the Sun as a function of wavelength, we find it's a pretty good fit to a theoretical Planck spectrum. It peaks at around 2 kilowatts per square meter per micron at a wavelength of half a micron.



- It doesn't fit exactly because the Planck spectrum describes the radiation you get from particles that have been knocking around long enough to reach a constant temperature. They're in thermodynamic equilibrium. It's often called a blackbody spectrum, because the derivation relies on the material being a perfect absorber of photons and therefore "black."

- The Sun, or any other real object, does not meet those criteria exactly. The Sun is not all at one temperature; it gets hotter as you go deeper. And the Sun's material is not perfectly absorbing. But the spectrum of the Sun and other stars are nevertheless reasonably well described by the Planck function, which is what allows us to say that the Sun is approximately a blackbody.

- Let's look at the Planck spectrum on logarithmic axes. That way, we can let the wavelength scale range over a factor of 1000, from ultraviolet to infrared, and we can let the flux density scale over a factor of a trillion.



- As the temperature increases, the curve lifts up vertically: Hotter sources produce more radiation at all wavelengths. The area under each curve—the integral of flux density over wavelength—is the total flux, which is equal to σT^4. That's the Stefan-Boltzmann law. Double the temperature and the flux rises a factor of 16.

$$F = \int F_\lambda d\lambda = \sigma T^4$$

- As the temperature increases, the peak of the spectrum shifts to shorter wavelengths. This makes sense because we expect the typical photon energy, hc/λ, to be of order kT, implying λ should be of order hc/kT—inversely proportional to temperature.

$$\frac{hc}{\lambda_{\text{peak}}} \sim kT$$

- When we do the math exactly, we find that the peak of the spectrum occurs when λ is about 1/5 of hc/kT. That's called Wien's law. We can also write it as a scaling relation.

$$\lambda_{\text{peak}} \approx \frac{hc}{5kT} \approx 10\,\mu\text{m} \left(\frac{T}{300\,\text{K}}\right)^{-1}$$

We are constantly bathed by photons whose spectrum follows the Planck function with an accuracy better than 1 part in 10,000 and a temperature of 2.7 Kelvin. According to Wien's law, that corresponds to a wavelength of 1 millimeter, which is in the microwave band of the spectrum.

Why is the universe permeated with this microwave blackbody radiation? It's a clue that at some point in the past, the universe itself was a "gas" of particles at a single temperature—in thermodynamic equilibrium—long before it became the place we know today, with tiny pockets of extreme heat and vast expanses of frigid cold.

This so-called cosmic microwave background radiation is some of the best evidence we have for the big bang.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 3.

Choudhuri, *Astrophysics for Physicists*, chap. 2.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 3.

Maoz, *Astrophysics in a Nutshell*, chap. 2.

PhET, "Radiating Charge," https://phet.colorado.edu/en/simulation/radiating-charge.

Ryden and Peterson, *Foundations of Astrophysics*, chap. 5.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 9

COMPARATIVE PLANETOLOGY

This lecture will analyze the planets from an astrophysicist's perspective. Mercury, Venus, Earth, and Jupiter—a quartet of planets spanning a wide range in orbital distance, size, mass, temperature, and type of atmosphere—make for an interesting comparative study. The lecture will address the following questions: Why are these planets so different? Why doesn't Mercury have any atmosphere? Why is Venus so much hotter than Earth? Why is Jupiter so huge?

COMPARING ORBITAL DISTANCE, RADIUS, MASS, AND TEMPERATURE

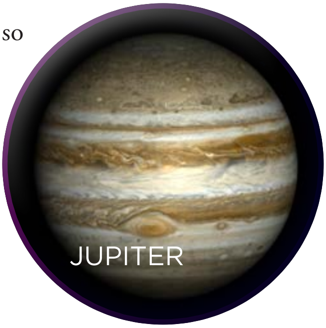
	a [AU]	R [$R_\oplus$]	M [$M_\oplus$]	T [K]
EARTH	1	1	1	288
VENUS	0.72	0.95	0.81	737
MERCURY	0.39	0.38	0.055	100–725
JUPITER	5.2	11.2	317	124

- Earth orbits the Sun at 1 AU; that's its orbital distance (a). Its radius (R) is 1 Earth radius, and its mass (M) is 1 Earth mass. The surface temperature averaged over the whole globe (T) is about 59° Fahrenheit, or 288 Kelvin. And it has an atmosphere composed mainly of nitrogen and oxygen.

- In many ways, Venus is like Earth's twin sister, except its atmosphere is mainly carbon dioxide and it's much thicker than Earth—almost 100 times as massive. It's also scorching hot, with an average surface temperature of 737 Kelvin, 2.6 times hotter than Earth.

- Mercury, the closest planet to the Sun, has a semimajor axis of 0.39 AU. Mercury is relatively small, 38% the size of the Earth and 5.5% of the mass. It looks like the Moon: barren and pocked with craters. Also like the Moon, Mercury has no atmosphere to speak of. Its surface temperature varies from a freezing 100 Kelvin at night to a broiling 725 Kelvin at noon.

- Going the other direction—beyond Earth and past Mars—Jupiter is out at 5.2 AU. Jupiter is a different beast from Mercury, Venus, and Earth. It's a gas giant planet. It's basically all atmosphere; it's a ball of hydrogen and helium gas with no solid surface. As you dive deeper into the planet, it keeps getting denser, until eventually the pressure becomes so high that hydrogen liquefies and even turns into a metal. For Jupiter's surface temperature, we can use the Stefan-Boltzmann law to define an effective temperature—the temperature of the outermost layers of gas—which comes out to be 124 Kelvin. That's colder than it ever gets on Earth.



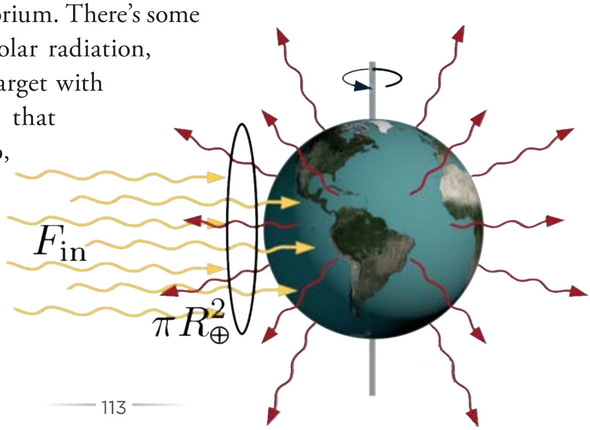
WHAT DETERMINES THE TEMPERATURE OF A PLANET?

- Why is the Earth's average surface temperature 288 Kelvin and not much higher or lower? The Earth absorbs sunlight. But it doesn't retain all that solar energy. It radiates. As the Earth's temperature rises, it radiates more power, rising as T^4.

- At some point, the radiated power equals the incoming solar power—at which point the Earth stops heating up. It reaches radiative equilibrium, with no net gain or loss of energy.

- Let's calculate the temperature of a planet that has achieved radiative equilibrium. There's some flux, F_{in}, of incoming solar radiation, and the planet is a big target with cross-sectional area πR^2 that intercepts that flux. So, the incident power is

$$P_{\text{in}} = F_{\text{in}} \cdot \pi R_{\oplus}^2.$$



- In equilibrium, this must equal the outgoing power. For simplicity, let's assume the entire surface of the planet is radiating like a blackbody at a single temperature T, so, according to the Stefan-Boltzmann law, the flux is σT^4, and the total radiated power is σT^4 times the total surface area, $4\pi R^2$.

$$P_{\text{out}} = \sigma T^4 \cdot 4\pi R_{\oplus}^2$$

- In radiative equilibrium, we set power in equal to power out and solve for T, giving

$$P_{\text{in}} = P_{\text{out}} \longrightarrow F_{\text{in}} = 4\sigma T^4$$
$$T = \left(\frac{F_{\text{in}}}{4\sigma}\right)^{1/4}.$$

- F_{in}, the solar flux at the orbital distance of the planet, is equal to the luminosity of the Sun spread out over a giant sphere with a radius equal to the orbital distance, a.

$$F_{\text{in}} = \frac{L}{4\pi a^2}$$

- We can simplify this if we also approximate the Sun as a blackbody and replace L with the following.

$$F_{\text{in}} = \frac{L}{4\pi a^2} = \frac{4\pi R_{\odot}^2 \sigma T_{\odot}^4}{4\pi a^2} = \frac{R_{\odot}^2 \sigma T_{\odot}^4}{a^2}$$

- That way, we get some nice cancellations, leading to our final answer.

$$T = \left(\frac{F_{\text{in}}}{4\sigma}\right)^{1/4} = \left(\frac{R_{\odot}^2 \sigma T_{\odot}^4}{4\sigma a^2}\right)^{1/4} = T_{\odot}\sqrt{\frac{R_{\odot}}{2a}}$$

- If we evaluate the numbers for the case of Earth, we get 280 Kelvin. That's not far from the true average temperature (T) of 288 Kelvin—an encouraging sign that our calculations captured the essential physics. We can write the result as a scaling relation.

$$\frac{280\,\mathrm{K}}{\sqrt{a/(1\,\mathrm{AU})}}$$

- This makes it easy to apply to the other planets in the quartet. We can add a column to the chart, called equilibrium temperature (T_{eq}), with the result of this calculation.

	a [AU]	R [$R_\oplus$]	M [$M_\oplus$]	T [K]	T_{eq} [K]
EARTH	1	1	1	288	278
VENUS	0.72	0.95	0.81	737	328
MERCURY	0.39	0.38	0.055	100–725	446
JUPITER	5.2	11.2	317	124	122

- This formula works well for Jupiter, too. But the calculated temperature for Venus is more than 400° too cold. And what about Mercury, where there's no single temperature? Clearly, we're missing something important.

- What we're missing are the effects of the atmosphere. Venus has a very thick atmosphere, and Mercury has none.

- When we derived the equilibrium temperature, we equated the power from the Sun to the power radiated by the entire surface of a spherical planet. We assumed that, somehow, the surface maintains a constant temperature.

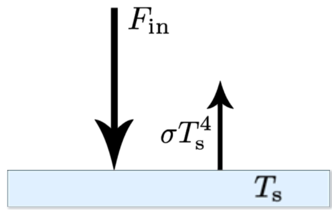
	Atm.
EARTH	N_2, O_2
VENUS	CO_2
MERCURY	none
JUPITER	H_2, He

- For the Earth and Jupiter, that's not a bad assumption, because the global circulation of the atmosphere tends to smooth out temperature differences. But on Mercury, there's no way for heat to flow quickly around the surface, so the dayside gets cooked and the nightside freezes. Therefore, our approximation of a constant temperature is inappropriate.

- Let's recalculate the temperature, using the opposite approximation: no heat transfer. Every square meter of the surface has its own temperature. For simplicity, let's target the hottest-possible temperature—which is at local noon, when the Sun is directly overhead.

- The incoming solar flux, F_{in}, is the same as before, but the outgoing flux, F_{out}, is the blackbody radiation at the local surface temperature, σT^4, or

$$T_{\text{s}} = \left(\frac{F_{\text{in}}}{\sigma}\right)^{1/4}.$$



- It's the same as before, except we're missing a 4 that used to be in the denominator—which means the result will be without the 2.

$$T_{\odot}\sqrt{\frac{R_{\odot}}{a}}$$

- So, the temperature we calculate under the assumption of local reradiation is higher by a factor of $\sqrt{2}$.

- In addition, for Mercury, the semimajor axis, a, is 0.39 AU. But if we want the absolute hottest temperature, we should plug in the distance of closest approach to the Sun, $a(1 - e)$, where e is the orbital eccentricity.

$$T_{\text{s, max}} = T_{\odot}\sqrt{\frac{R_{\odot}}{a(1-e)}}$$

- With that correction, we get 708 Kelvin, which agrees pretty well with the data.

- Venus has plenty of atmosphere, so a constant surface temperature is a reasonable approximation. This time, our calculation didn't work because we ignored the possibility that the atmosphere—not just the surface—absorbs and emits radiation.

- Venus's atmosphere is full of carbon dioxide molecules. What kinds of photons do those molecules absorb and emit?

- For something to absorb a photon, there has to be somewhere the photon's energy can go. There must be something ready to accept the few electron volts that a solar photon has to offer. Usually, the things that can absorb a few electron volts are electrons; they use the energy to jump between different orbits in atoms and molecules.

- But the problem is that carbon dioxide is electronically stable. The electrons are happy where they are, so it takes tens of electron volts before they can be persuaded to jump into higher orbits. That makes carbon dioxide transparent to visible light. The same is true of nitrogen and oxygen.

- But molecules do have other ways to absorb energy, besides shifting around electrons. The atoms are joined by chemical bonds that act sort of like springs, which can absorb energy and start vibrating and rotating. And the energies of those motions are on the order of a few hundredths of an electron volt.

- Objects that radiate photons with those kinds of energies are at room temperature. For a blackbody, at 300 Kelvin, the characteristic energy, kT, is 1/40 of an electron volt, and the typical wavelength is 10 microns—infrared radiation.

- So, the Sun's photons, with energies of a few electron volts, sail through the atmosphere as if it weren't there and get absorbed by the surface. The surface temperature rises to a few hundred Kelvin and starts radiating infrared photons, which can't just escape into space. They get absorbed by the atmosphere.

- All this means we need to modify our calculation of the equilibrium temperature.

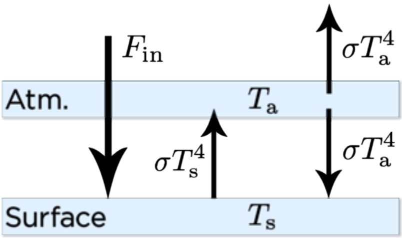
- Rays of sunlight are striking the planet's surface with a flux of F_{in}. The surface is at temperature T_s, and it radiates infrared photons with a flux of σT_s^4.

- When we were analyzing Mercury, we set the surface flux equal to F_{in}, but now let's include an atmosphere. We'll use the simplest-possible model: a layer of gas completely transparent to visible light and completely opaque to infrared. All the upward flux from the surface gets absorbed.

- Next, the atmosphere heats up and starts radiating. It reaches some equilibrium temperature, T_a, and thereby produces a flux of σT_a^4, both upward (into space) and downward (back to the surface).

- In equilibrium, the flux radiated away from the surface must equal the flux hitting the surface—from both the Sun and the atmosphere. Therefore,

$$\sigma T_{\text{s}}^4 = F_{\text{in}} + \sigma T_{\text{a}}^4.$$



- The atmosphere also reaches equilibrium between outgoing and incoming flux:

$$2\sigma T_{\text{a}}^4 = \sigma T_{\text{s}}^4.$$

- The 2 is there because the atmosphere radiates from both the top and bottom surfaces.

- This time, we have 2 equations instead of one. To solve for T_s, the surface temperature, we double the first equation and then use the second one to eliminate the $\sigma T_a^{\,4}$.

$$\begin{aligned} 2\sigma T_\mathrm{s}^4 &= 2F_\mathrm{in} + 2\sigma T_\mathrm{a}^4 \\ 2\sigma T_\mathrm{s}^4 &= 2F_\mathrm{in} + \sigma T_\mathrm{s}^4 \\ \sigma T_\mathrm{s}^4 &= 2\,F_\mathrm{in} \end{aligned}$$

- There's an extra factor of 2 on the right side of the equation, compared to the case of no atmosphere, so in this model, the atmosphere increases the surface temperature by a factor of $\sqrt[4]{2}$, or about 19%.

- This is the famous greenhouse effect—the same one that has everyone worried on Earth. If we make the atmosphere more opaque to the Earth's own thermal radiation, by pumping out megatons of CO_2, then the surface will heat up.

- For Venus, the greenhouse effect is much larger than 20%. It's more like 220%!

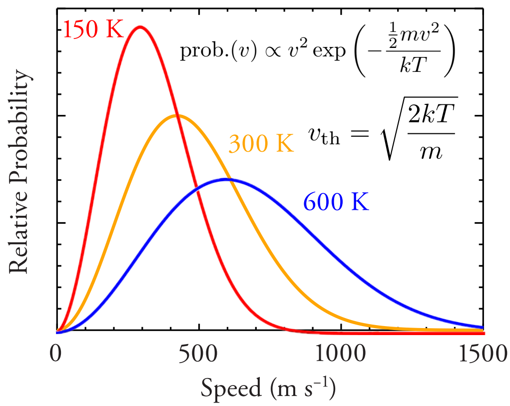
WHY ARE THE PLANETS' ATMOSPHERES SO DIFFERENT?

- The answer to the question of why planets' atmospheres are so different has 3 essential physical ingredients: the equilibrium temperature, the Maxwell-Boltzmann distribution, and the escape velocity.

- The Maxwell-Boltzmann distribution tells us the distribution of speeds of the gas molecules and how it depends on temperature. The peak is at a speed of

$$\sqrt{\frac{2kT}{m}}.$$

- That's the most probable speed—v_{th}, the speed due to thermal fluctuations.



- The thermal speed is the most probable one, but there's a distribution of velocities, a spread of an order of magnitude or so.

- This means that for hot gases and lightweight molecules—high T and low m—there's a danger that some molecules will be moving faster than the planet's escape velocity. Deep in the atmosphere, that doesn't matter, because after a few nanoseconds, the molecule will bash into another one, randomizing its speed again. But at the top of the atmosphere, where the density is low, the fastest-moving particles might escape into space and never return.

- Let's consider hydrogen gas, H_2, with a mass of 2 proton masses. At the Earth's average surface temperature of 288 Kelvin, the thermal velocity is 1.5 kilometers per second, which means a small fraction of hydrogen molecules have speeds as high as this speed.

$$v_{\rm th} = \sqrt{\frac{2kT}{m}} = 1.5 \text{ km s}^{-1} \text{ for } H_2 \text{ at } 300 \text{ K}$$

- Meanwhile, the escape velocity is $\sqrt{\frac{2GM}{R}}$, which for Earth is 11 kilometers per second.

$$v_{\rm esc} = \sqrt{\frac{2GM}{R}} = 11 \text{ km s}^{-1} \text{ for Earth}$$

- So, if there were hydrogen gas in Earth's atmosphere, the random jostling of those molecules would cause some of them to leave—and by the time a billion years goes by, almost all the hydrogen would be gone.

- To be sure a planet can hold on to a molecule, let's require the escape velocity to be at least 10 times higher than the thermal velocity. That leads to an inequality that we can solve for m, the mass of the molecule. It will be useful to express that mass in units of proton masses.

$$v_{\mathrm{esc}} > 10\, v_{\mathrm{th}} \longrightarrow \sqrt{\frac{2GM}{R}} > 10\sqrt{\frac{2kT}{m}}$$

$$\frac{GM}{R} > 100\,\frac{kT}{m}$$

$$\frac{m}{m_{\mathrm{p}}} > 100\,\frac{kTR}{GMm_{\mathrm{p}}}$$

- Running the numbers for the case of the Earth at room temperature, 300 K, we get a minimum molecular mass of 4.

$$> 4\left(\frac{T}{300\,\mathrm{K}}\right)\left(\frac{R/M}{R_{\oplus}/M_{\oplus}}\right)$$

- That's the mass of helium—2 protons and 2 neutrons. So, this calculation suggests the Earth can't retain hydrogen and helium is a marginal case. But nitrogen and oxygen are plenty heavy. N_2 has a molecular mass of 28, and for O_2, it's 32.

- Let's evaluate the minimum mass for the other planets. For Venus, it's 11; it's higher than Earth's because Venus is hotter. Venus can't retain hydrogen and helium either, but holding on to CO_2, with a molecular mass of 44, is no problem.

	m_{min} [m_{p}]
EARTH	4
VENUS	11
MERCURY	66
JUPITER	0.06

- Mercury is hot and has a low escape velocity, so it can't hold anything lighter than 66. That rules out all the common molecules.

- For Jupiter, we get a number that is less than 1, meaning that it's massive and cold enough to retain even the lightest gases.

- By comparing thermal velocity to escape velocity, we can understand some of the patterns in planetary atmospheres. But keep in mind that real atmospheres are complicated.

Venus and Earth are the same size and have the same mass, so why is Venus's atmosphere 100 times more massive than Earth's?

The current thinking is that Venus and Earth originally did have similar atmospheres, but Venus underwent a runaway greenhouse effect—a positive feedback loop that ultimately led to complete evaporation of all the surface water. Venus was left as a blistering hot, dry planet smothered in carbon dioxide.

WHY IS JUPITER SO MASSIVE?

- Jupiter is massive enough to retain hydrogen and helium. But why is it so massive?

- Hydrogen and helium are the most common elements in the universe—by far. Carbon, nitrogen, and oxygen are only a percent or 2 by mass of the total inventory of atoms. The Sun and all the stars are basically giant spheres of hydrogen and helium. So, then, why aren't the planets like that, too? Why aren't all planets like Jupiter?

- It's because planets form in a very different way from stars. Stars are what happens when a cloud of gas collapses under its own gravity into a ball of gas dense and hot enough to ignite nuclear fusion reactions.

- Planets form out of the stuff left over after star formation. A new star is surrounded by a rotating disk of hydrogen and helium gas. Planets are thought to start from the microscopic flecks of heavier elements that are mixed in with this gas—dust grains. Over time, gravity causes the dust to settle down into a layer that is thin and dense enough that the dust grains start colliding and stick to each other. Over millions of years, they grow to the size of planets.

- The details of this process, called planetesimal formation, are still hazy, but what is clear is that the hydrogen and helium don't participate. They're too lightweight; their high thermal speeds prevent them from condensing.

- Why aren't all planets rocky? Where did Jupiter come from?

- If a rocky planet gets big enough, its escape velocity will start to exceed 10 times the thermal velocity of the gas. That allows the planet to start attracting and retaining gas from the huge reservoir of hydrogen and helium all around it.

- In the solar system, this process had no trouble making objects as large as Venus and the Earth. Those planets can't hold on to hydrogen. They're too hot, and their escape velocities are too low. How much farther from the Sun would we need to move them so they could retain hydrogen?

- Let's take our equation for the minimum mass of molecules that can be retained—which is proportional to temperature—and combine that with our equation for the equilibrium temperature in terms of orbital distance.

$$\left.\begin{array}{c} \dfrac{m}{m_{\mathrm{p}}} > 100\,\dfrac{kTR}{GMm_{\mathrm{p}}} \\[2ex] T = T_{\odot}\sqrt{\dfrac{R_{\odot}}{2a}} \end{array}\right. \longrightarrow \quad \frac{m}{m_{\mathrm{p}}} > 100\,\frac{kT_{\odot}R}{GMm_{\mathrm{p}}}\sqrt{\frac{R_{\odot}}{2a}}$$

- Then, we can plug in the mass of hydrogen and solve for a, the orbital distance. This comes out to be 3.4 AU. That's right in between the orbits of Mars and Jupiter—that is, at the dividing line in the solar system between the inner rocky planets and the outer gas giant planets. So, a rocky planet needs to be far away from the Sun to be cold enough to start attracting hydrogen and have a chance of becoming a gas giant.

- Of course, in reality, planet formation is more complicated than we've accounted for in our calculations.

Why are hydrogen and helium by far the most common elements in the universe? Why is everything else so much rarer?

This observation, like the cosmic microwave background radiation, is another pillar of evidence supporting the theory of the big bang.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 19.

de Pater and Lissauer, *Planetary Sciences*, chap. 1.

Planetary Fact Sheets, http://nssdc.gsfc.nasa.gov/planetary/planetfact.html.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 10

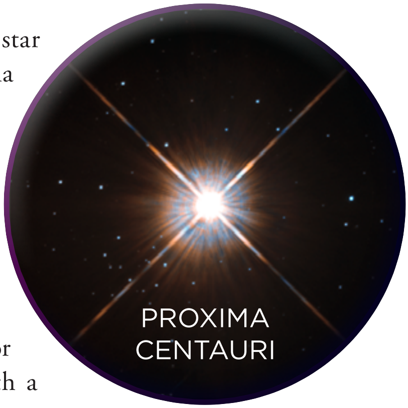
OPTICAL TELESCOPES

Most people think that the main purpose of a telescope is to magnify, or to make distant objects seem closer. But for professional astronomers, magnification is one of the least important reasons to build a big telescope. As you will learn in this lecture, the most important reasons are to collect more light and to improve angular resolution, spectral resolution, and temporal resolution.

COLLECTING MORE LIGHT

- You can think of the light from distant sources as a gentle rain of photons falling onto the Earth's surface. When we look up, our eyes collect the rain that happens to fall through our pupils. When we use a telescope, we're using a bigger bucket to collect the rain.

- Consider Proxima Centauri, the closest star to the Sun. The rain coming from Proxima Centauri amounts to about 100 photons per square centimeter per second. The pupils of our eyes are only a fifth of a square centimeter, and our retinas don't respond to 90% of the photons that get in. Even worse, we only have about a tenth of a second to detect a signal. Our visual system can't accumulate signals for longer than that; it's like a camera with a hardwired shutter speed.



- Putting all that together, the average number of photons from Proxima Centauri that enter our eyes and get detected during a tenth of a second is 0.2. That's less than 1, which is why Proxima Centauri is invisible to the naked eye.

$$\text{Photon flux} = 100/\text{cm}^2\ \text{s}$$

$$\#\ \text{detected} = 100/\text{cm}^2\ \text{s} \times 0.2\ \text{cm}^2 \times 0.1 \times 0.1\ \text{s} = 0.2$$

- With a telescope and a digital camera, though, we can boost all the factors in this calculation. We can increase the collecting area, detect nearly all the photons, and use whatever shutter speed we want.

- The reason this is so important is that the rain of photons is not completely steady. There are fluctuations.

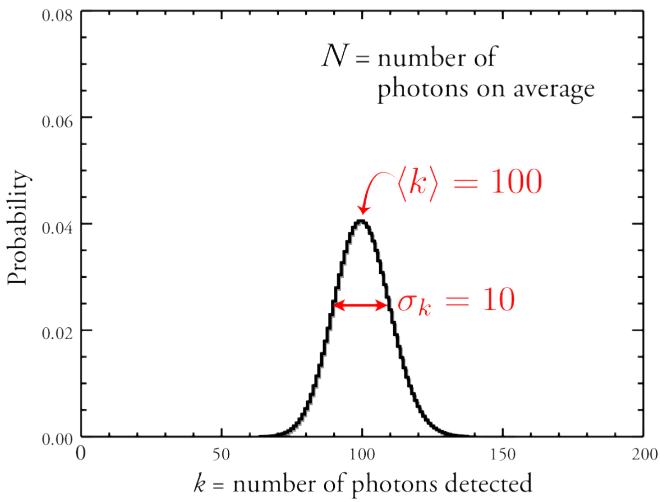
- Let's say we're looking at a star that's spraying our telescope with 100 photons per second, on average. The key here is the phrase "on average." In reality, the number of photons arriving each second is a random number. Whenever we're dealing with events that occur at random times but with a well-defined average rate—such as radioactive decays, earthquakes, or the arrival of photons from a distant star—the relevant piece of mathematics is the Poisson distribution.

- Let's say N is the average number of events we expect to occur in some time interval. Then, the probability we will actually observe k events is

$$\text{prob.}(k) = \frac{N^k}{k!}e^{-N}.$$

- That's the Poisson distribution.

- Let's plot it for the case of $N = 100$ photons. The horizontal axis is k, the number of photons detected, and the vertical axis is the corresponding probability.



$$\frac{\text{signal}}{\text{noise}} = \frac{100}{10} = 10$$

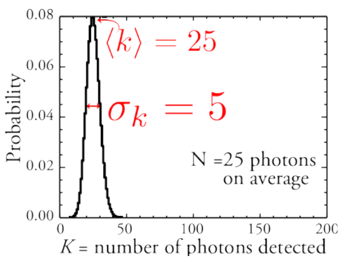
- The most likely outcome is 100. But any value between 90 and 110 is also likely. Mathematically, the mean value of k, which is written as $\langle k \rangle$, is equal to N. And the standard deviation of k, or σ_k—a measure of the spread in the likely values—is equal to $\sqrt{N}$.

$$\text{mean} \quad \langle k \rangle = N$$
$$\text{standard deviation} \quad \sigma_k = \sqrt{\langle (k - \langle k \rangle^2 \rangle} = \sqrt{N}$$

- In this case, the mean is 100 and the standard deviation is 10, which means there's about a 68% chance that the number of photons in our image will be between 90 and 110.

- This implies that our flux measurement could deviate from the average by about 10 units out of 100, or 10%. In other words, the signal-to-noise ratio of our measurement is 10. The signal is the average flux of the star, and the noise refers to the random deviations from the average.

- What if we're looking at a fainter source that only delivers an average of 25 photons to our detector? Let's replot the Poisson distribution for $N = 25$. Now the mean is down to 25, and the standard deviation is 5, so the signal-to-noise ratio is 25/5, or 5.



- In general, the maximum-possible signal-to-noise ratio is

$$\frac{\text{signal}}{\text{noise}} = \frac{N}{\sqrt{N}} = \sqrt{N}.$$

- In fact, it's usually worse than that, because the light from your favorite star might be blended together with other sources of light—from other nearby stars in the sky that you can't resolve with your telescope or from the faint glow of the Earth's atmosphere. The photons from those other sources, called sky noise, also contribute to the fluctuations.

- When sky noise is significant, we need to revise our signal-to-noise equation to be

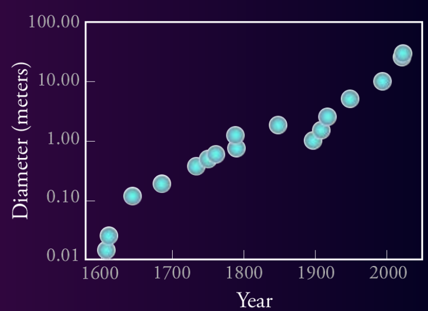
$$\frac{\text{signal}}{\text{noise}} = \frac{N}{\sqrt{N + N_{\text{sky}}}},$$

where N_{sky} is the number of photons from other sources besides the star.

- The inevitable fluctuations in the photon count, called Poisson noise, is an unforgiving fact of life in astronomy. There are many ways a measurement can be wrong, but even if you have perfect equipment and make no mistakes, you can't eliminate the Poisson noise.

- But there are ways to decrease it. We can increase the exposure time, thereby increasing both N and N_{sky}. We can use a camera that responds to a wider range of photon energies, or wavelengths, boosting the photon count. But often, these measures are not enough, and we have no choice but to build a bigger telescope.

This chart shows the diameter of the largest optical telescope in the world over the centuries, starting from Galileo's telescopes, which were only 10 centimeters across, to today's largest telescopes, which use mirrors that are 10 meters across.



The vertical axis is logarithmic, which means that the upward linear trend represents exponential growth. Telescopes have been doubling in size every 40 years or so.

IMPROVING ANGULAR RESOLUTION

- Another advantage of big telescopes is they improve our angular resolution: our ability to measure the direction from which photons are coming and make sharper images.

- The diffraction limit is the tightest-possible focus we can achieve given that light is a wave. The smallest angle we can resolve is about 1.22 λ/D, where λ is the wavelength and D is the diameter of the primary mirror or lens.

- Visible light has a wavelength of about half a micron, and our eyes have a D of about 5 millimeters. The diffraction limit works out to be 1.2×10^{-4} radians, or 25 arc seconds.

$$\Delta\theta_{\min} = 1.22\,\frac{0.5 \times 10^{-6}\,\mathrm{m}}{5 \times 10^{-3}\,\mathrm{m}} = 1.2 \times 10^{-4}\,\mathrm{rad} = 0.4\,\mathrm{arcmin}$$

- But astronomers can do better by increasing D to 10 meters, an improvement by a factor of 2000. In that case, the equation gives $\Delta\theta_{\min}$ of only 0.013 arcseconds.

$$\Delta\theta_{\min} = 1.22\,\frac{0.5 \times 10^{-6}\,\mathrm{m}}{10\,\mathrm{m}} = 6.1 \times 10^{-8}\,\mathrm{rad} = 0.013\,\mathrm{arcsec}$$

- There's a catch, though. Fluctuations in the temperature and density of air produce a blurring effect of order 1 arc second, even on a high mountaintop, so we can't take full advantage of the large D to improve our angular resolution—at least not easily.

- There are 2 ways around the problem: Put your telescope in space, above the air, like the famous Hubble Space Telescope; or stay on the ground and use a technique called adaptive optics, which is when you put a deformable mirror somewhere in the light path of your telescope. Many mechanical actuators are mounted on the backside of the mirror, which can apply tiny localized forces under computer control, pulling and pushing by a fraction of a micron.

- The goal is to distort the mirror in just the right way to reverse the distorting effects of the air. You have a separate camera stare at an extremely bright star, which you know should appear as a sharp point in the image. But it doesn't, because of the turbulent air; it looks like a big blotch.

- The computer measures the shape of that blotch and uses an algorithm that tells it how to distort the mirror to turn the blotch into a point. And it all has to happen within a few milliseconds, because the atmosphere is constantly changing.

- Adaptive optics allows us to get close to the diffraction limit even with a 10-meter telescope.

- Better angular resolution also gives us another way to reduce the Poisson noise. The reason that helps is it improves our ability to separate the star's light from the other light sources; it reduces N_{sky} in our signal-to-noise equation. So, in situations where the sky noise is the main problem, improving the angular resolution leads to a higher signal-to-noise ratio.

IMPROVING SPECTRAL RESOLUTION

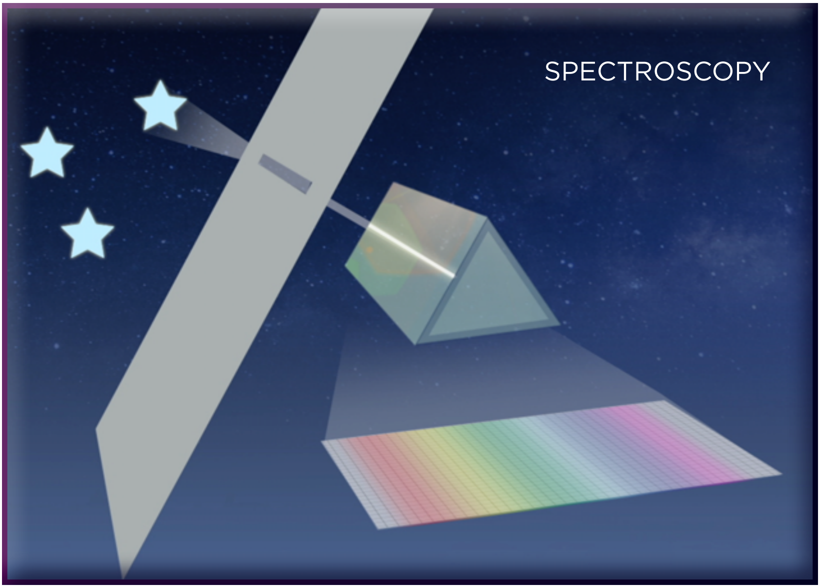
- A third advantage of telescopes is they improve our spectral resolution: our ability to measure the energies, or wavelengths, of the incoming photons.

- Our eyes can sense millions of different colors, making very fine distinctions between photons with wavelengths ranging from about 0.4 to 0.7 microns. But as miraculous as color vision is, our eyes can't analyze those different hues in any quantitative way.

- With a telescope, we can distinguish colors in several different ways. We can put colored filters in the light path, admitting only the light within a narrow range of wavelengths. Then, we can take exposures through different filters and combine them.

- If we want to make finer distinctions, we can put a prism in the light path, which deflects light by an amount depending on wavelength. That way, the light from a star, instead of appearing as a point in our image, shows up as a little stripe of light, a minirainbow.

- If there are other stars nearby, we don't want all those rainbows to overlap. That would be confusing. So, before the beam of light reaches the prism, we interrupt it with an opaque sheet with a slit in it and position the slit so that the light from our favorite star, or galaxy, goes through the slit, and all the other sources of light are blocked.



- That's spectroscopy. The camera takes a picture of the spectrum of light so that each pixel in the image records how many of the arriving photons have a wavelength in a particular range. Sorting the photons by wavelength is a powerful technique for understanding the nature of a light source.

- Modern spectrographs differ in detail from the simple prism-and-camera design. Instead of a prism, they're more likely to use a diffraction grating, a series of regularly spaced grooves in an otherwise flat surface. The grooves have a different reflectivity from the surrounding material, and when a light beam bounces off, the reflections from the stripes interfere with each other in such a way that the direction of the reflected beam depends on wavelength.

IMPROVING TEMPORAL RESOLUTION

- The fourth advantage of telescopes is they improve our temporal resolution: our ability to measure the time of arrival of a burst of photons or any more gradual variations in their rate of arrival.

- Our eyes have a sort of built-in exposure time of about 20 milliseconds. For astronomy, a fixed exposure time would be a severe limitation. Cameras allow us to choose whichever exposure time is most appropriate for the purpose.

- With short exposures, we can witness events that happen quickly. Today, there are astronomical cameras capable of capturing scenes on the scale of milliseconds, or even microseconds. This has led to the discovery of all sorts of fascinating, rapidly varying sources.

- Cameras also allow us to accumulate light for much longer than our eyes—to build up a high signal-to-noise ratio.

- The best way to think of a research telescope is a machine for measuring the properties of photons from celestial bodies: the direction it came from, its energy or wavelength, and the time of its arrival. There is also one more: polarization, one of a photon's wavelike properties. The polarization is the direction in which the electric field of the wave is oscillating.

- The traditional system of units that optical astronomers use to measure the flux of a source—the power per unit area arriving at the Earth in the form of light—is the magnitude scale. (It would be logical to express flux in standard metric units, or watts per square meter, but astronomers hardly ever do that). The apparent magnitude, m, of a source is defined as

$$-2.5\log\left(\frac{F}{F_0}\right),$$

where F is the flux and F_0 is a reference flux.

- The choice was made to set F_0 equal to the flux of Vega, a bright star in the northern sky. That way, we don't have to worry about calibrating our camera to measure starlight in watts per square meter. We can just compare the flux of our star to the flux of Vega. Then, if we ever want to convert to standard units, we just look up the flux of Vega in watts per square meter, which has already been measured.

- Because of the minus sign, brighter objects have lower magnitudes. It makes the magnitude scale like a ranking system for flux: a first-magnitude star is brighter than a second-magnitude star, which is brighter than a third-magnitude star, and so on.

- When we invert the magnitude equation, we find

$$F = F_0 \cdot 10^{-0.4m}.$$

- So, a star with 0 magnitude has $F = F_0$—that is, the star has the same flux as Vega. And a star with magnitude 1 is fainter by a factor of $10^{-0.4}$, or about 40%.

$$\begin{aligned} m = 0 &\longrightarrow F = F_0 \\ m = 1 &\longrightarrow F = F_0 \cdot 10^{-0.4} \approx 0.40\,F_0 \end{aligned}$$

- If we go down to a fifth-magnitude star, we find that increasing the magnitude by 5 units corresponds to lowering the flux by a factor of 100.

$$m = 5 \longrightarrow F = F_0 \cdot 10^{-0.4\times5} = F_0 \cdot 10^{-2} = 0.01\, F_0$$

- In addition to flux, we can use magnitudes to represent color. One way to quantify the color of a light source is to measure its flux through a colored filter—for example, blue—that only admits short-wavelength photons. Then, measure its flux through a red filter and take the ratio, F_B/F_R. Objects that are intrinsically blue will have a higher ratio than red objects.

- The color index is defined as the difference between the 2 corresponding apparent magnitudes.

$$\text{color index} = m_B - m_R$$

- When we write that difference in terms of fluxes, using the definition of apparent magnitudes and then use some standard log properties to rearrange things, we see that the color index is a logarithmic scale for the flux ratio, F_B/F_R. For this pair of filters, a high color index means a high blue magnitude relative to red, and because high magnitude means faint, a source with a high blue-red color index is relatively red.

$$= -2.5\log\frac{F_B}{F_{B,0}} - 2.5\log\frac{F_R}{F_{R,0}}$$

$$= -2.5\log\frac{F_B}{F_R} - 2.5\log\frac{F_{B,0}}{F_{R,0}}$$

- We can also use magnitudes to represent luminosity—the intrinsic power of a source independent of its distance from Earth. For a source at distance d that emits equally in all directions, the relation between luminosity and flux is $F = L/4\pi d^2$.

- Instead of expressing the luminosity of a source in watts or as some multiple of the Sun's luminosity, we sometimes express it as an absolute magnitude, defined as the apparent magnitude the source would have if we could magically position it 10 parsecs away. (This number is an arbitrary choice; the point is that if we put everything at a common distance, differences in apparent magnitude signify differences in luminosity.)

- Usually, we use *m* for apparent magnitude and *M* for absolute magnitude. To find the relationship between them, first we need to compute the flux the source would deliver to Earth if it were parked 10 parsecs away. Let's call that F_{10}. Because flux is proportional to $1/d^2$,

$$F_{10} = F\left(\frac{d}{10\,\mathrm{pc}}\right)^2.$$

- Then, we form an apparent magnitude based on F_{10}. We divide by the reference flux, take the log, and multiply by –2.5. That's the absolute magnitude.

$$\begin{aligned} M &= -2.5\log\left(\frac{F_{10}}{F_0}\right) \\ &= -2.5\log\left[\frac{F}{F_0}\left(\frac{d}{10\,\mathrm{pc}}\right)^2\right] \\ &= -2.5\log\left(\frac{F}{F_0}\right) - 2.5\log\left(\frac{d}{10\,\mathrm{pc}}\right)^2 \\ M &= m - 5\log\left(\frac{d}{10\,\mathrm{pc}}\right) \end{aligned}$$

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 6.

Ryden and Peterson, *Foundations of Astrophysics*, chap. 6.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 11

RADIO AND X-RAY TELESCOPES

Different parts of the electromagnetic spectrum go by different names, but for the purposes of astronomy, the spectrum can be divided into 3 parts. The infrared, visible, and ultraviolet are lumped together and called optical. Everything with lower energies and longer wavelengths is called radio, and everything with higher energies and shorter wavelengths is called x-rays. These are the 3 main cultures in observational astronomy. Optical, radio, and x-ray astronomers use different technologies, different units, and different jargon. In this lecture, you will learn about radio and x-ray astronomy.

MAKING AN IMAGE WITH CAMERAS AND MIRRORS

- Among the reasons that we build telescopes is to detect invisible radiation—to go beyond the visible range and explore all the other orders of magnitude.

- The analogy from the previous lecture of a telescope as a bucket for collecting rain fails in one major respect: A bucket collects rain no matter what direction it's coming from, whether it's falling vertically or coming in at a slant. For a telescope, that would be bad; we'd have no idea where the photons were coming from.

- We want to point the telescope at something—a star or a galaxy—and only collect the photons coming from a narrow range of directions. We want to make an image, in which each pixel of the image corresponds to a different point on the sky.

- Mathematically, an image is a mapping between the direction the photons are coming from and the position on some surface. Ideally, we want the number of photons that hits each point on the surface to be proportional to the rate of photons arriving at the Earth from a certain direction in space.

- But how does a telescope perform that mathematical mapping? How does it sort the photons by incoming direction?

- Conceptually, the simplest way to make an image is with a pinhole camera, where no lens or mirror is needed.

- Think of a completely dark room except for a small hole in one wall that is admitting light, and there's a screen on the opposite wall. The light that hits a given spot on the screen must have come from the specific direction defined by the line from the hole to that spot, which means there is a 1-to-1 correspondence between the x and y coordinates on the screen and the direction from which light enters the hole.

- Actually, there *would* be a 1-to-1 correspondence if diffraction didn't exist. Photons coming from angles differing by less than about λ/D radians get blurred together in the image because of diffraction of electromagnetic waves.

- How can we achieve the tightest-possible focus? Let's say the distance from the hole to the screen is F, the so-called focal length, and the hole has diameter D.

- Let's start with a big hole and imagine the light from a distant star goes straight through the hole. Ideally, we want the image of the star to be a pinpoint on the screen, but because the hole is so big, the image is a luminous circle of diameter D.

- To shrink the circle down into a pinpoint, we should reduce the diameter of the hole. But at some point, D becomes so small that the diffraction limit starts to dominate—the image of the hole stops being a crisp circle and starts fuzzing out. If we keep reducing D beyond that point, the image gets worse, because the diffraction limit goes as λ/D.

- The optimal case, the tightest focus, is when the angular diameter of the circular image, D/F, is equal to the diffraction limit, which is roughly λ/D.

$$\frac{D}{F} \sim \frac{\lambda}{D}$$

- Solving for D, $D \sim \sqrt{\lambda F}$.

- We've just learned that the ideal size for the hole is on the order of the geometric mean of the focal length and the wavelength.

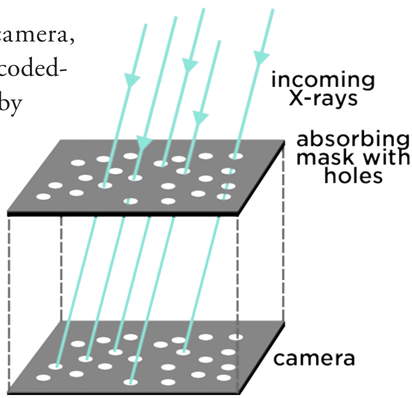
- There's one big drawback to using a pinhole camera. The optimal hole is so small that hardly any light gets through. So, while it can make sharp images over a wide field of view, the images are faint, with a low signal-to-noise ratio. Even a daylight scene might require an exposure lasting minutes or hours.

- This makes pinhole cameras totally useless for optical astronomy, where the light levels are down by many orders of magnitude. But variations on the pinhole camera do find use in astronomy—when there is no other choice.

- Very-high-energy photons—x-rays and gamma rays—punch straight through most materials. We can't easily build mirrors or lenses to redirect their trajectories and focus them. But we can drill a hole in a layer of a dense metal, such as lead or tungsten, that is thick enough to block them. Then, for the screen, we can use germanium crystals or other materials that produce an optical flash when a high-energy photon hits it.

- To mitigate the problem of the small hole, astronomers drill several widely spaced holes. This makes the pattern on the screen confusing, because now it's the overlap from lots of different pinholes, but if you observe the same scene multiple times with the camera in different orientations, computer algorithms can disentangle the information and reconstruct the scene.

- This is called a multiple-pinhole camera, and there's another variation called a coded-aperture telescope that has been used by x-ray astronomers.



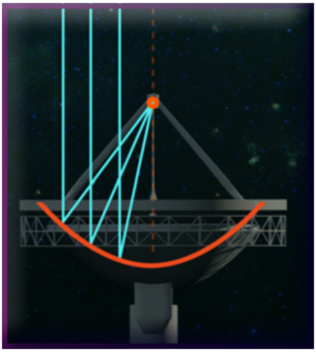
- In the optical and radio domains, photons don't pack as much energy, so we can use mirrors to focus them. This allows us to collect much more light than a pinhole camera.

- At optical wavelengths, we could use a lens instead of a mirror, but lenses have a problem. Because they're made of glass, they act like prisms, even when you don't want them to; the amount they deflect the light depends at least slightly on its wavelength. This introduces chromatic aberration: What should be a white point in the image turns into a multicolored blob.

- That's one reason why astronomers use reflective optics: glass mirrors coated with aluminum or silver.

- Another reason is that a big mirror is easier to support than a lens. Only one side of the mirror, the reflective side, matters, so you can lay it shiny-side up on a stable supporting structure—as opposed to a lens, which you can only grip on the rim or else you'll block the light.

- The mirror has to be curved to focus light. One possibility is to use a mirror whose surface has the shape of a parabola, in which all the photons coming in along the symmetry axis get bounced to the same point: the focal point. So, a parabolic bucket not only collects the photons but also redirects and concentrates them into a small area, where we can put a detector.



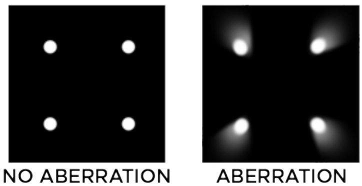
- Parabolic dishes are used in radio telescopes. The dish is pointed straight at a source of radio waves and the radio static coming from that direction is measured. To make an image, we can slew the dish through a range of directions, recording the intensity of the static as we go.

- At shorter wavelengths, it's more common to use the imaging property of the parabola, in addition to the focusing property. Photons arriving head-on, straight down the symmetry axis of the parabola, all end up at the focal point. The ones coming from a slightly different angle, off axis, don't get directed to the focal point, but they do get concentrated near a different point that is displaced to the side from the focal point.

- For small angles, the displacement is proportional to the incoming angle—which is just what we want to make an image: a mapping between incident direction and location in a surface.

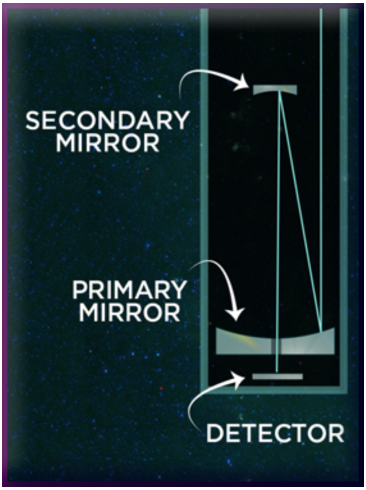
- To capture the image, we insert a 2-dimensional detector with many separate pixels, each of which can record the intensity at a point in the focal plane. For optical astronomers, the detector of choice is the charge-coupled device, which uses a thin layer of pure silicon. Photons hit the silicon and knock loose some electrons, which can be trapped and counted by electronics mounted on the silicon surface.

- Optical astronomers rarely use parabolas. It would be wonderful if all the photons from a given direction get sent to a single point in the focal plane but that's not the case. Only the on-axis light is focused perfectly. Off the axis, the rays hitting different parts of the parabola land in slightly different locations in the focal plane.



- This causes image distortions, or aberrations. Stars near the middle of the image appear pretty sharp, but away from the center, they look like little cones, or comets—which is why this type of distortion is called a coma.

- Perhaps there is a better shape than the parabola, one that will focus light to a single point no matter which way it's coming from. Unfortunately, in 1856, James Clerk Maxwell and Ernst Abbe proved that there's no shape for a mirror, even in principle, that lacks aberrations.



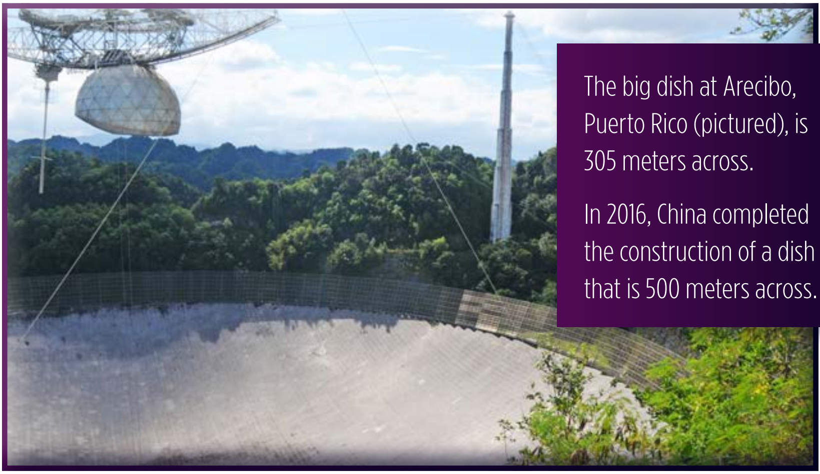
- Instead, astronomers reduce aberrations by using multiple mirrors. The light hits a primary mirror and bounces to a secondary mirror, which is also curved, before going to the detector.

- Having more than one surface that can be adjusted gives the optical designer the freedom to reduce whatever kind of aberration is most worrisome. Mirrors with elliptical or hyperbolic cross sections, instead of parabolic ones, can be used. Even 3 mirrors can be used.

RADIO ASTRONOMY

- What distinguishes radio astronomy from optical and x-ray astronomy is that the wavelengths are long. This has important implications.

- First, it's easier to build a focusing mirror. As a rule of thumb, the mirror needs to be polished with an accuracy of at least a tenth of a wavelength—that is, it needs to conform to the shape of a parabola, or whatever surface has been chosen, to within $\lambda/10$. That's much easier when λ is a meter than when it's a millionth of a meter. This is why the world's largest telescopes are radio telescopes.



The big dish at Arecibo, Puerto Rico (pictured), is 305 meters across.

In 2016, China completed the construction of a dish that is 500 meters across.

Radio astronomy grew out of electrical engineering. Karl Jansky, a radio engineer at Bell Telephone Laboratories, was the first person to detect radio static coming from astronomical sources in 1931.

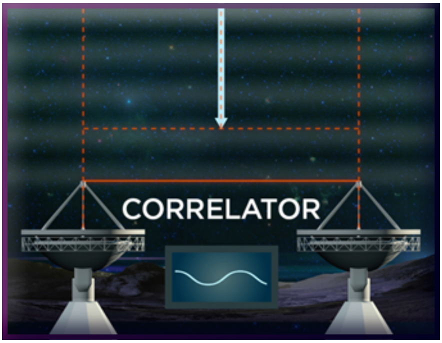
For nearly a decade after that, the world's leading radio astronomer—essentially the *only* radio astronomer—was Grote Reber, an engineer who built his own backyard radio telescope as a hobby and mapped all the brightest radio sources in the sky.

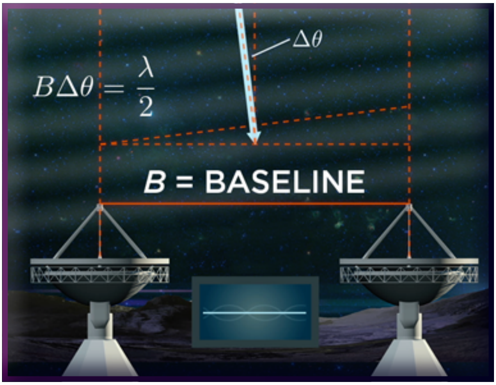
- Another implication of the long wavelengths of radio photons is that their wave nature is much more conspicuous than their particle nature. In fact, nobody ever calls them "radio photons"—always radio waves. We can use purely wave-based methods to detect, amplify, and combine radio waves. It's not like counting photons; it's more like following the rise and fall of ocean waves.

- This means that Poisson noise does not apply in radio astronomy; the statistics are different. It also allows for the possibility of interferometry—an amazing trick to improve our angular resolution.

- You'd think that radio images would have terrible angular resolution because the diffraction limit is λ/D and λ is a million times bigger than in the optical range. But, in fact, radio astronomers enjoy the best angular resolution of all. This is because they can combine the waves from 2 widely separated receivers.

- Let's say we have 2 dishes that are a distance B (for baseline) apart and there's a radio source straight overhead sending down waves, a pattern of crests and troughs. Both dishes detect the waves and amplify them, without losing track of the phase, and send the signals to a central computer called a correlator.





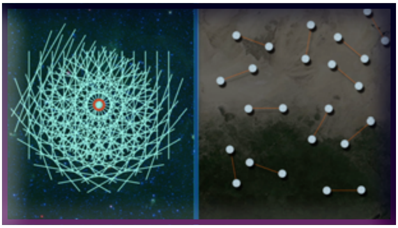
- The correlator merges the 2 signals, which combine constructively because the crests from straight overhead hit the dishes at the same time, resulting in a big signal at the output.

- If the source is not straight overhead, the waves are coming in at some small angle $\Delta\theta$. In this case, the crests will hit one of the dishes first and then the other. In the small-angle approximation, the extra distance the wave must travel to hit the second dish is $B\Delta\theta$. If that is equal to half a wavelength, then the waves will interfere destructively; a crest plus a trough gives 0 signal at the output.

- The output of the correlator can tell us whether the source is straight overhead or displaced by an angle of $\lambda/2B$. We can resolve details in our radio image on that angular scale.

- That's just like the diffraction limit, except B is not the diameter of a single dish; it's the separation between the dishes. We can put them kilometers apart, if we want, and achieve the same angular resolution as an enormously large telescope.

- What if we increase θ some more, so that $B\Delta\theta$ is an entire wavelength? At that point, we return to constructive interference. This means that with only the output of the correlator, we can't tell whether a source is directly overhead or at an angle of λ/B from the vertical—or $2\lambda/B$, or $3\lambda/B$. They all lead to constructive interference.



- We can get around this problem by building more than 2 dishes. Each pair of dishes is sensitive to angles that form a pattern of fringes on the sky, with the fringes perpendicular to the line between the dishes. If the correlator is getting information from many different pairs of telescopes, with baselines in different directions, then the crisscrossing fringe patterns allow us to pinpoint the source to a specific location on the sky so that we can make a proper image.

The Very Large Array Radio Telescope in New Mexico is an interferometer with 27 radio dishes that can be moved along railroad tracks to spread them apart as much as 36 kilometers. It can achieve the same angular resolution as a 36-kilometer radio dish!

The catch is that it doesn't have the sensitivity of such a large dish, so it doesn't collect nearly as much radiation as the bigger dish would, but it does solve the angular resolution problem.

X-RAY ASTRONOMY

- The Earth's atmosphere is totally opaque to x-rays. X-ray photons are energetic enough to knock apart molecules and rip off electrons, so they slam into a molecule and get stopped. They never make it down to the ground.

- That's great news for life on Earth; it keeps us from getting fried. But it's bad news for x-ray astronomy—a field that couldn't get started until the 1960s, when advances in the space program made it possible to put telescopes above the atmosphere. The sky turned out to have a glittering display of x-ray sources, which are now understood to be related to black holes, neutron stars, supernovas, and other fascinating phenomena.

- Focusing x-rays is difficult. We can't just use a regular mirror; the photons would penetrate through it instead of getting reflected. The coded-aperture mask referenced previously is one solution. But there is another.

> The pioneers of x-ray astronomy were from high-energy particle physics and cosmic-ray physics. Two of these pioneers were Bruno Rossi and Riccardo Giacconi, who convinced NASA to launch an x-ray telescope in 1962.

- It turns out that x-rays will reflect from metals as long as they strike the surface at a very grazing angle, like skipping a stone off the surface of a pond. X-ray mirrors look like polished metal cylinders. But they're not exactly cylinders. The sides are slightly sloped so that x-rays can skip off the surface and land on a charge-coupled device for detection.

- It's a serious challenge, though, to polish the surface of an x-ray mirror to an accuracy of $\lambda/10$ when lambda is just a few nanometers—it needs to be smooth at the atomic scale! But if we want sharp x-ray images, we have to figure out how to do it.

- In 1999, NASA launched the Chandra X-ray Observatory, which makes the sharpest x-ray images of any facility. The Chandra mirrors are glass coated with iridium. To boost the collecting area, Chandra uses multiple mirrors, nesting small ones inside the larger ones. And there's a second set of hyperboloid mirrors to reduce the image aberrations. But the time to research, develop, and implement the mirror technology was more than 20 years, and the mission cost billions of dollars.


CHANDRA

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 6.

Ryden and Peterson, *Foundations of Astrophysics*, chap. 6.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

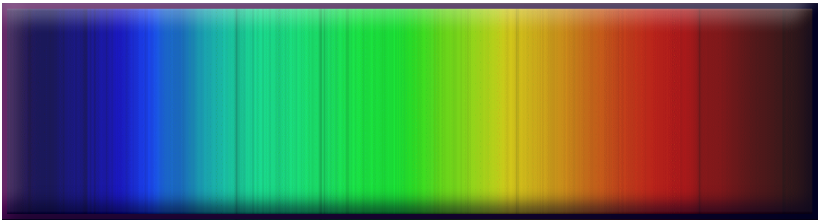
Lecture 12

THE MESSAGE IN A SPECTRUM

This lecture will take a deep dive into spectroscopy, which is one of the most important things we can do with a telescope. We can sort photons by wavelength or, equivalently, by energy. Spectroscopy is our main way to learn about the physical conditions of a star, planet, nebula, or galaxy—its temperature, pressure, and composition.

THE SOLAR SPECTRUM

- In the spectrum of sunlight, along with the continuous ribbon of color, there are dark lines. Sunlight is missing certain colors. For example, there's a conspicuous line in the red part of the spectrum and another one in the yellow.



The dark lines in the Sun's spectrum were first observed in 1802, but the message wasn't decoded until the 1920s because it was written in the language of quantum theory, which wasn't developed until the 1920s.

- The pattern of lines looks sort of random, but not completely random—almost as though it were a secret code. In fact, it *is* a code. It contains a message from the atoms and ions in the Sun's outer layers, which are broadcasting information about their temperature, abundance, and much more.

- Setting aside quantum physics, we can imagine an atom as a tiny solar system in which the attraction is provided by the electrical force instead of gravity. Electrons orbit the nucleus, just as planets orbit the Sun.

- The classical theory of electromagnetism says that any accelerating charge will radiate. And an electron orbiting a nucleus is accelerating. It's a centripetal acceleration; it feels an inward force. That's what keeps it bound to the nucleus. Therefore, an orbiting electron should emit electromagnetic waves, which carry away energy.

- In the case of planets, the total orbital energy—kinetic plus gravitational potential—is

$$E = -\frac{GMm}{2a},$$

where *a* is the orbital distance.

- A similar formula applies to the case of an electron orbiting a proton under the influence of the electrical attraction:

$$E = -\frac{\eta e^2}{2a},$$

where η is the Coulomb constant and *e* signifies the magnitude of the charge on both the proton and the electron.

- If the system is gradually losing energy by radiating electromagnetic waves, *E* is getting lower. It becomes more negative. This means *a* must be shrinking. The electron should spiral inward toward the nucleus.

- This should happen to planets, too. An accelerating planet produces gravitational waves that fly away at the speed of light. In principle, this causes planets to spiral inward and crash into the Sun. But the timescale over which the orbit shrinks is orders of magnitude longer than the age of the universe, making gravitational waves irrelevant to planetary motion.

- But electromagnetism is stronger than gravity, and atoms are smaller than planetary systems. When you calculate how long it should take for an electron to spiral inward and crash into the nucleus, it comes out to be on the order of 10 nanoseconds. In other words, all the atoms in the universe should collapse within 10 nanoseconds. But they don't, so there must be something wrong with the classical model of the atom.

- What's wrong is that it ignores quantum theory. Electrons are not just particles; they have wavelike properties, too. An electron in an atom is like a wave that's trapped near the nucleus.

- Because the electron's wave nature—its wave function—is confined by the electrical attraction to the nucleus, it is a standing wave, and the possibilities for its energy are restricted to a discrete set. Those energies are found by solving an equation: the time-independent Schrödinger equation.

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2}-\frac{\eta e^2}{r}\Psi=E\Psi$$

- We get a discrete set of energies, instead of a continuum, which means there's some lowest-energy state—a minimum orbital distance—called the ground state. In the context of gravity, that would be like a certain distance from the Sun inside which planets can't exist.

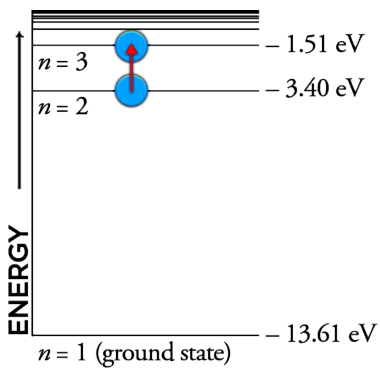
- For hydrogen, the energy levels of the electron obey a simple equation:

$$E_n=-\frac{13.6\,\mathrm{eV}}{n^2},$$

where n is a whole number. The ground state is –13.6 eV.

- These levels can be represented on an energy diagram in which height is proportional to energy. Because the energies vary as $1/n^2$, they bunch up near 0 as n increases.

- For bigger atoms with more electrons and for molecules with more than one nucleus, the energy diagrams are more complicated, but the point remains that the energy levels are discrete and there's a ground state.



- This explains why atoms and molecules are stable: the electrons in the ground state can't radiate. To do so, they would need to lose energy, and they're already in the minimum energy state.

- It also explains the dark lines in the Sun's spectrum. An electron can jump from one level to a higher one, but only if it absorbs just the right amount of energy, such as from a passing photon whose energy is equal to the difference between 2 energy levels.

- Suppose there's an electron in the $n = 2$ level of hydrogen. To jump to $n = 3$, the electron needs to absorb an energy—based on our equation—of 1.89 eV.

$$E_3 - E_2 = -13.6\,\mathrm{eV}\left(\frac{1}{3^2} - \frac{1}{2^2}\right) = 1.89\,\mathrm{eV}$$

- A photon with 1 eV won't work; there's not enough energy. A photon with 2 eV also won't work; the electron either absorbs a whole photon or does not absorb a photon at all. It can't carve off a fraction of the photon's energy.

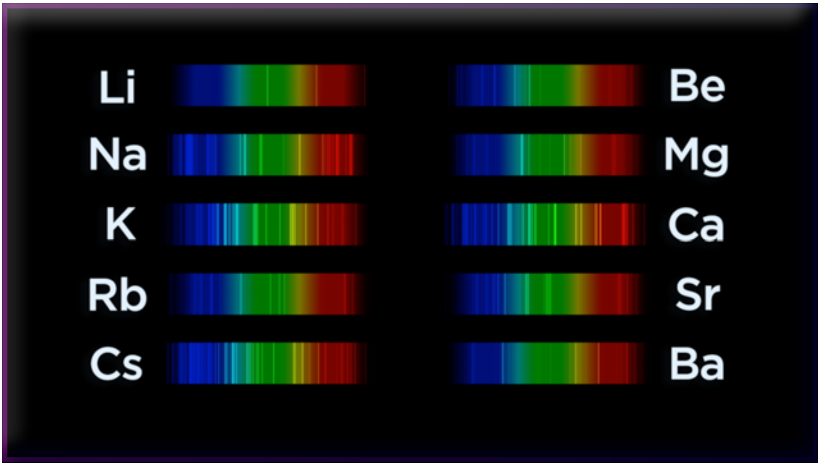
- Because photon energy, E, equals hc/λ and λ equals hc/E, an energy of 1.89 eV corresponds to a wavelength of 0.656 microns.

$$\lambda = \frac{hc}{E_3 - E_2} = 0.656\,\mu\mathrm{m}$$

- This is in the red part of the spectrum. In fact, it's exactly the shade of red that was missing from the solar spectrum!

- The dark lines are called absorption lines. It's a bit of a misnomer; the atoms are emitting those special photons just as much as they're absorbing. The electron can fall back down from $n = 3$ to 2, releasing a photon with 1.89 eV.

- Because we know the energy levels of atoms from laboratory experiments and computer calculations, whenever we measure the wavelengths of absorption lines, we can learn the identities of the atoms responsible for those lines. Each atom produces a bunch of lines with a known pattern.

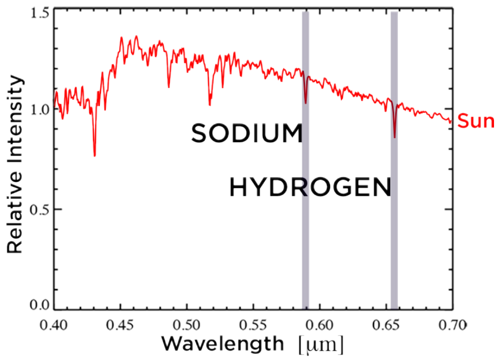


STELLAR SPECTRA

- Let's look at some stellar spectra in a different way. Instead of rainbows, we'll plot intensity versus wavelength across the visible range of the spectrum.

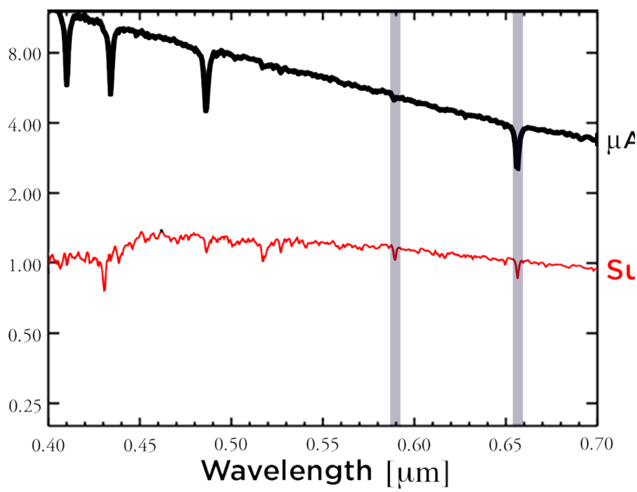
- If the Sun were a perfect blackbody, the spectrum would be a smooth curve peaking at around half a micron, but a real spectrum has lots of divots—the absorption lines. We see now that the star is not completely black at those wavelengths; there's some intensity, but it's lower by as much as 30% than the surrounding spectrum.

- For example, there is a dip at 0.656 microns from hydrogen and one at 0.589 microns from sodium.



- Let's compare the solar spectrum with the spectra of other stars. For that, we'll need to use a logarithmic intensity axis to capture the wide range of intensities on a single chart.

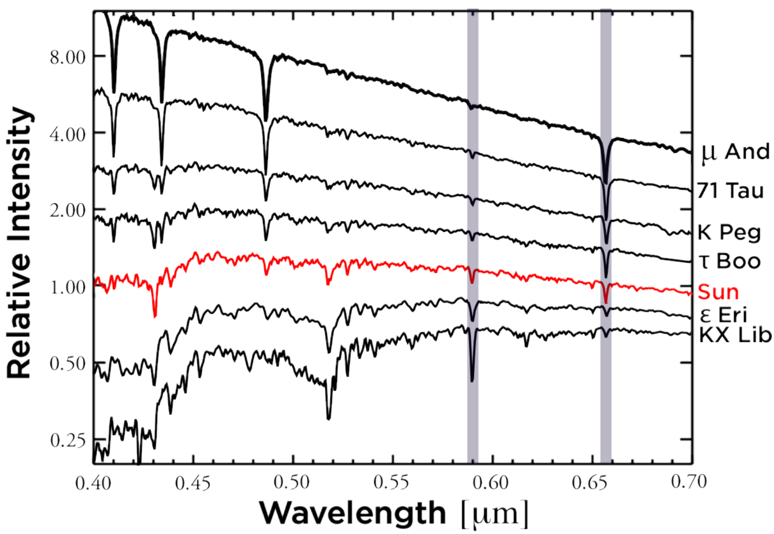
- First, let's pick a bluish-white star, such as μ Andromedae.



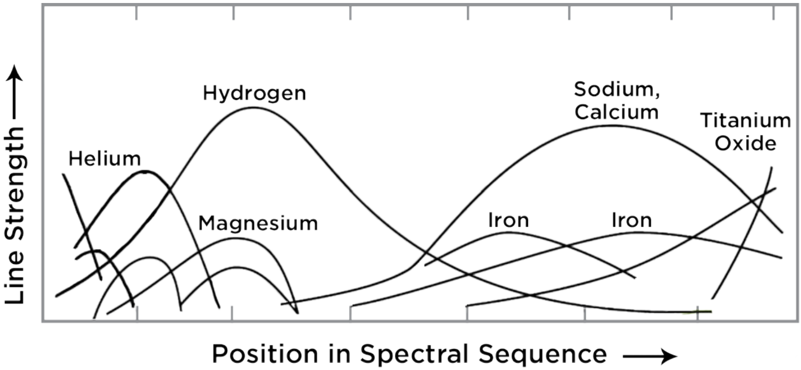
- We can tell it's bluer than the Sun because the intensity rises toward shorter wavelengths, the blue end of the spectrum. And the absorption lines are different, too. We hardly see the sodium line, and the hydrogen line is deeper. The deep lines at shorter wavelengths are also from hydrogen; they represent jumps from $n = 2$ to 4, 5, and 6.

- Does that mean μ Andromedae has more hydrogen than the Sun and the Sun has more sodium? That's what some astronomers used to think. The truth is subtler.

- A big clue came when the pioneers of astronomical spectroscopy discovered that stellar spectra can be sorted into a single sequence. The following is a graph of the spectra of some representative stars in that sequence. From one to the next, the shape of the spectrum and the pattern of lines change smoothly. As we work our way downward on this plot, the hydrogen lines go away and the sodium line builds up.



- This pattern can be shown as a chart in which the *x*-axis is the position of a star in this sequence and the *y*-axis is the strength of various absorption lines—how dark they are in the spectrum. At one end, some helium lines are strongest; then, as you move down, the hydrogen lines take over. Later, the lines of sodium and calcium rise and fall in strength.



- At the very end of the sequence, there are lines from titanium oxide, which is closely related to the active ingredient in some sunscreen lotions. Does this mean that the stars at this end of the sequence are made out of sunscreen lotion? No.

- It turns out that all the stars are made of basically the same ingredients in the same proportions. What's changing as we move along the sequence is the temperature of the star's outer layers.

The pattern of absorption lines in spectra encodes temperature. The first person to crack this code was Cecilia Payne (later Payne-Gaposchkin) in 1925. By interpreting stellar spectra with the newly discovered laws of quantum theory, she could deduce the temperature and composition of a star based on its spectrum. She was the first person to figure out what stars are made of.

- To produce an absorption line, not only do we need photons with the right energy, but we also need a supply of atoms with electrons sitting in the lower energy level ready to absorb those photons. And the supply of such atoms will depend on temperature.

- In general, stars are about 75% hydrogen by mass and 24% helium, with the remaining percent or so from heavier elements, especially carbon, nitrogen, and oxygen. Sodium, calcium, and titanium oxide are only present in minute quantities, but they happen to have electron energy levels in the right places to produce strong lines when the temperature is right.

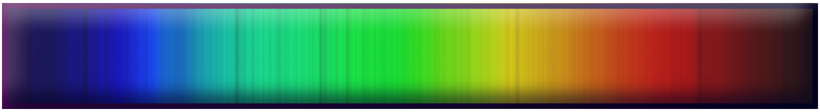
NEBULAR SPECTRA

- A nebula is an interstellar cloud of gas and dust. There are many different kinds, and they are found all over the galaxy.

- The Orion Nebula is a glowing, colorful, complex cloud that appears in the sword of Orion. At 400 parsecs away and 7 or 8 parsecs across, it's one of the nearest big star-forming regions, where clumps of gas are contracting under their own gravity, eventually becoming stars.

- The spectrum of the Orion Nebula looks different from that of a star. It's totally dark, except for a few bright spikes at particular wavelengths.

ORION NEBULA SPECTRUM



SUN SPECTRUM


ORION NEBULA

- The key difference between the star and the nebula is density, not temperature. The density of gas in the Orion Nebula is about 10 billion times less than in a star.

- The low density means our usual assumption of thermal equilibrium between atoms and photons breaks down. The atoms in the nebula are in thermal equilibrium; they collide and share energy, achieving a uniform temperature of about 10,000 Kelvin for the case of Orion.

- But photons don't interact as frequently. If an atom inside the nebula emits a photon, it won't always crash into another atom and get absorbed. There's a good chance it will just sail off into space.

- In the nebula, collisions between atoms push the electrons up to higher energy levels. They eventually fall back down and emit photons, which escape the nebula and land in our telescope, having never interacted with anything during its journey across 400 parsecs.

- The photons we detect, then, only have the energies that are emitted by electron transitions within the atoms of the nebula. This is called an emission-line spectrum.

- The situation is different in a dense medium, like a star, where a photon emitted by one atom almost always gets absorbed by a neighboring atom. That atom will then reemit another photon, but not necessarily with the exact same energy; the high density and high frequency of collisions cause the energy levels to be blurred out. Photons can also get scattered from charged particles—electrons or ions—which alter the photon's energy in a continuous way, producing more blurring of the spectrum.

- We can imagine the photons as pinballs getting knocked around in a pinball machine. All the bumpers, flippers, and kickers are the atoms and ions. The directions and energies of the photons are constantly changing.

- Given enough time, the atoms and photons come into thermal equilibrium, and the photons attain a Planck spectrum. The little bit of light that leaks out will have the spectrum of a continuous rainbow.

- Quantitatively, the key concept separating the nebula and the star is the mean free path: the average distance a particle travels between interactions. The mean free path of a visible photon in air might be 10 kilometers on a clear, dry day. That means photons from 10 kilometers away can propagate unimpeded all the way to the eye.

- The mean free path of the molecules in air is much shorter, on the order of a tenth of a micron. They're constantly colliding with each other. The photon mean free path is longer because visible photons don't have the right energy to be absorbed by nitrogen or oxygen.

- In general, a medium is transparent if its spatial extent is substantially smaller than the mean free path of photons, such as a piece of glass or a diffuse interstellar gas cloud like the Orion Nebula. Instead of using the word "transparent," astronomers use the term "optically thin." The idea is that from the point of view of a photon, the medium is thin; there's little chance a photon will hit something when it goes through.

- Conversely, when the spatial extent is larger than the mean free path, astronomers say it is optically thick, or opaque.

- A related concept is optical depth, which is defined as the distance through some medium divided by the mean free path. It's a dimensionless number that is often written as τ.

- Whenever you're peering into an optically thick medium, the photons reaching your eye originate from an optical depth of order 1.

- That's what we mean when we refer to the "surface" of Jupiter or of the Sun. Both objects are gaseous spheres; there's no solid surface. The gas gradually gets denser as you go from the outside to the inside. But we can define the "surface" in such cases as the level at which the optical depth is of order 1. The more technical term is the photosphere.

- The mean free path is not the same for all photons. It depends strongly on wavelength. Photons with a wavelength just right to be absorbed by the surrounding atoms have a much shorter mean free path than photons with some random wavelength.

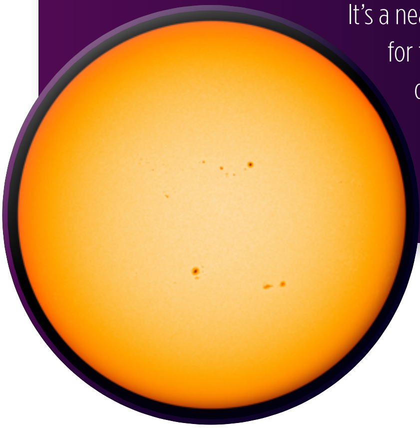
The Basic Rules of Spectroscopy (Kirchhoff's Laws)

1. Optically thick sources emit a continuous Planck spectrum. That's because the density is high enough for the photons to reach thermal equilibrium with the atoms.
2. Optically thin sources, such as the Orion Nebula, produce emission lines. Photons don't interact with atoms on their way out, so there's no way for them to reach equilibrium.
3. If a hot, optically thick source is surrounded by a cooler, optically thin layer, that's when you get an absorption spectrum.

- Now you can understand why it's a little too simplistic to say that the dark lines in a spectrum come from absorption. It's true that those photons are getting absorbed, but they're getting emitted at the same rate. It's more accurate to say that the dark lines are there because the star is opaquer at those wavelengths—we can't see as far down into the stellar fog.

An image of the Sun with an appropriate telescope shows that it gets fainter near the rim of the circle—the so-called limb of the solar disk. It's not an optical illusion; the radiation from the limb really is less intense than the radiation from the center. This phenomenon is called limb darkening.

At first, you might surmise that this phenomenon is a result of the fact that the Sun is a sphere, not a flat circle, and that there is a shadowing effect. But what causes limb darkening is that the Sun, like all stars, is a gaseous sphere that's hotter and denser on the inside than the outside.



It's a neat effect, and a useful way to test models for the Sun's photosphere. By observing limb darkening in different colors, we can learn about how temperature increases with depth and how the overall opacity of the Sun varies with wavelength.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 9.

Ricker, et al., "Transiting Exoplanet Survey Satellite."

Tyson, Strauss, and Gott, *Welcome to the Universe*.

QUIZ
LECTURES 7–12

1 Does the observation of the motion of stars around Sagittarius A* prove the existence of a black hole? If not, what further evidence would be required? [LECTURE 7]

2 A star falls directly into a black hole and is tidally disrupted. But if the black hole is too massive, this happens inside the Schwarzschild radius and the tidal disruption event cannot be observed. What is the maximum black hole mass for which we could observe the tidal destruction of a Sunlike star? [LECTURE 7]

3 The star Sirius has a radius 1.7 times larger than the Sun and an effective temperature of 9940 Kelvin. What is the total luminosity of Sirius relative to the Sun? [LECTURE 8]

4 What is the peak in the spectrum of thermal radiation from your own body? At what rate (in watts) does your body radiate energy? [LECTURE 8]

5 Which properties of the solar system can be understood from basic physical principles? Which properties were contingent on events happening in the remote past with no fundamental explanation? [LECTURE 9]

6 Life on Earth seems to require liquid water. This motivates the definition of the habitable zone of a star as the range of orbital distances where a planet would have a surface temperature from 0°C to 100°C. What are the boundaries of the Sun's habitable zone? How might your answer change depending on the planet's atmosphere? [LECTURE 9]

7 What are the advantages and disadvantages of putting an optical telescope in space? [LECTURE 10]

8 A diffraction-limited telescope of diameter 0.5 m is used to make a 10-second exposure of a star at a wavelength of 2 μm. During the exposure, the detector records 300 photons from the star. In the resulting image, the sky brightness is 1000 photons per square arc second. Calculate the signal-to-noise ratio of the detection of the star. [LECTURE 10]

9 What is the main obstacle to building a telescope optimized for ultraviolet observations? How about a radio telescope operating at 100 MHz? [LECTURE 11]

10 The Atacama Large Millimeter/submillimeter Array (ALMA) is a radio interferometer in northern Chile. It consists of 66 antennas that observe at wavelengths ranging from 0.3 to 9.6 mm. The maximum separation between antennas is 16 km. What is ALMA's best-possible angular resolution? [LECTURE 11]

11 Imagine a hypothetical star for which temperature decreases with depth (i.e., it is hotter on the outside than the inside). How would the appearance of the star differ from that of the Sun? What would its spectrum look like? [LECTURE 12]

12 Which of the following objects do you expect would have a spectrum resembling a blackbody? Venus, neon sign, human body, incandescent lamp, LED lamp, interior of an oven, interior of a freezer, radio antenna, lava erupting from a volcano, lightsaber. [LECTURE 12]

Go to page 337 for solutions.

Lecture 13

THE PROPERTIES OF STARS

For hundreds of years, we've known that the stars in the sky are distant Suns, producing vast quantities of light and heat. But are there different kinds of stars? Why are they so bright and hot? Will they keep shining forever? The answers to these questions came only after astronomers learned to measure the basic properties of stars: luminosity, temperature, radius, and mass.

MEASURING LUMINOSITY, TEMPERATURE, AND RADIUS

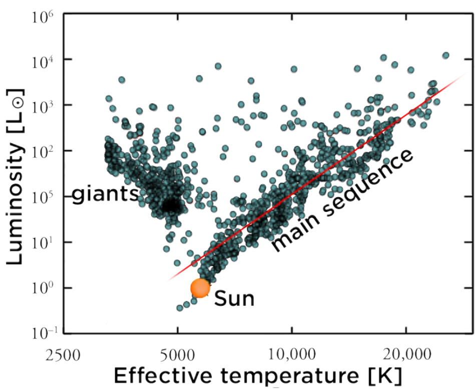
- To measure a star's luminosity—its total power output—we can use the flux-luminosity relationship. First, we measure distance to the star, d, maybe with parallax. Then, we measure the flux, F, the power per unit area we detect with our telescope. Finally, we calculate the luminosity as

$$L = 4\pi d^2 F.$$

- We get the effective temperature (T_{eff}), the temperature of its photosphere, from spectroscopy. We can measure the strength of the absorption lines and place it on the spectral sequence, which is a temperature scale.

- Let's plot L versus T_{eff} using logarithmic axes. To get oriented, we'll mark the location of the Sun, which has an effective temperature of about 5800 Kelvin and a luminosity of 1—that is, 1 solar luminosity.

- Next, let's plot the data for the 1000 brightest stars in the sky, which range in effective temperature from around 3000 to 30,000 Kelvin and range in luminosity from around 1/3 to 100,000 times the Sun's. The Sun is way below average; it's one of the least luminous stars on the chart.



- Many of the data points follow a diagonal stripe from the lower left to the upper right called the main sequence. The upward slope means stars with hotter photospheres are more luminous. That makes intuitive sense. But there's also a bunch of points higher up on the left—very luminous stars but with relatively cool photospheres. These are called giants.

- Because stars are approximately blackbodies, we can use the Stefan-Boltzmann law: The luminosity equals the total surface area, $4\pi R^2$, times the blackbody flux, $\sigma T_{\text{eff}}{}^4$.

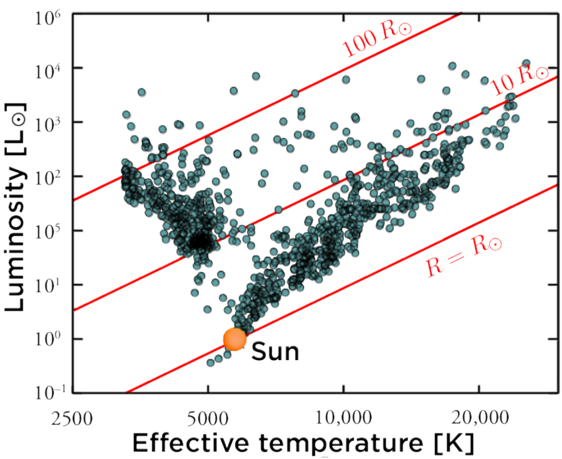
- And because we're using a logarithmic chart, let's take the log. The result is an equation that connects the location of a point in the chart with the radius of the star.

$$\log L = \log(4\pi\sigma) + 2\log R + 4\log T_{\text{eff}}$$

- The stars in the upper left of the chart are called giants because they're 10 to 100 times bigger than the Sun.

- The stars on the main sequence don't all have the same size. Stars on the faint, cool end are the size of the Sun or smaller, and stars on the luminous, hot end are up to 10 times bigger than the Sun.

- This type of chart, luminosity versus temperature, is called a Hertzsprung-Russell (HR) diagram after the astronomers who first drew these diagrams in the early 1900s.



- But their charts were a little different. Following astronomical tradition, they plotted absolute magnitude instead of luminosity, and instead of effective temperatures (which they didn't know at that time), they plotted a color index.

- The apparent magnitude is a logarithmic measure of flux: –2.5 times the log of flux relative to Vega. The minus sign makes bright stars have low magnitudes, and the 2.5 means that a difference of 5 magnitudes corresponds to a flux ratio of 100.

$$m = -2.5 \log \left(\frac{F}{F_0} \right)$$

bright stars have low magnitudes

Δmag = 5 corresponds to 100×

flux of vega

- The absolute magnitude is a log scale for luminosity. It's equal to the apparent magnitude minus 5 times the log of the distance divided by 10 parsecs.

$$M = m - 5 \log \left(\frac{d}{10\,\mathrm{pc}} \right)$$

- A color index is the difference between 2 apparent magnitudes measured through different filters. It's a log scale for a flux ratio. For example, the B – V color index is the apparent magnitude measured through a standard blue filter minus the apparent magnitude measured through a so-called visual filter, which is centered in the middle of the visible range.

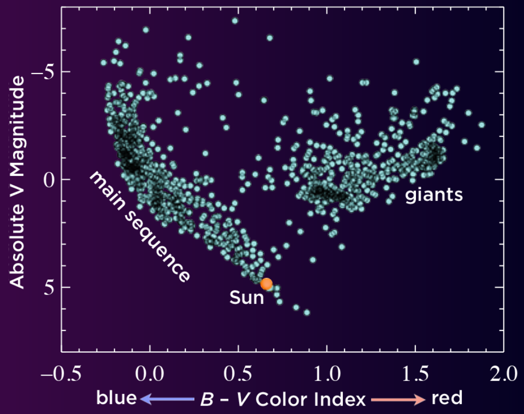
$$B - V = m(\text{blue}) - m(\text{visual})$$

- And because it's all referenced to Vega, the color index of Vega is 0. For the Sun, B – V happens to be 0.66.

- The color index is a proxy for effective temperature—blue stars are hot and red stars are cool—and the color index is easier to measure. You don't need spectroscopy; you just need a pair of colored filters for your camera.
- A traditional HR diagram shows the absolute *V* magnitude against *B* – *V*. It has the same features as the previous chart, but everything is flipped left to right. That's because a low value of *B* – *V* means the star is relatively bright in *B* compared to *V*—it's blue. And a star with high *B* – *V* is red. So, the horizontal axis goes from blue to red, hot to cool. It's a backward temperature scale.

There's a big difference between the *nearest* stars and the *brightest* stars. If we pick a certain volume of space and count all the stars inside, we find that small, faint stars are most common.

Bigger and more luminous stars are rare. To find them, we need to look far away. But because they're so bright, they dominate our naked-eye view of the night sky. That's why Betelgeuse, Arcturus, Aldebaran, and many other famous stars are giants.



In contrast, most of the nearest stars are invisible without a telescope.

MEASURING MASS AND SIZE

- We now know how to find the luminosity, temperature, and radius of a star, but if we want to understand stars, we need to know mass, which determines a star's gravitational pull and internal pressure as well as the amount of fuel available to keep it shining.

- How can we learn a star's mass? To answer this question, we look toward the constellation of Perseus, in which there is a red star called Algol. Every 3 days, Algol drops in flux by 30% and stays that way for 10 hours before going back to normal. We can tell the difference by eye.

- It turns out that Algol is actually a pair of stars that orbit each other and periodically eclipse one another. Pairs of stars—called binaries—are common. About half of the points of light in the night sky represent the combined light of more than one star, but our eyes lack the angular resolution to see them separately.

- Most binary stars do not show eclipses, though. Only a small percentage happen to have orbits oriented nearly parallel to our line of sight so that the stars cross directly in front of each other, blocking each other's light and making the whole system appear fainter.

- Eclipsing binaries like Algol allow us to measure stellar masses. We use Kepler's third law, which can be written as

$$a^3 = \frac{GM}{4\pi^2}P^2.$$

- In the planetary context, a was the semimajor axis of the planet's orbit, P was the period, and M was the mass of the Sun.

- When discussing Kepler's laws in lecture 4, we said that the Sun sits still at one focus of the elliptical orbit. But the truth is that the Sun moves a little bit. This follows from Newton's third law: Every reaction is accompanied by an equal and opposite reaction. If the Sun pulls on a planet, the planet must pull back on the Sun with equal force.

- But because acceleration equals force over mass and the Sun is so much more massive than the planets, the Sun's acceleration is much smaller than that of the planets. That's why we often neglect the Sun's motion.

- We can't do that when 2 stars of comparable mass are pulling on each other. We need to solve the 2-body problem. We'll find that both stars travel in elliptical orbits with a focal point at the center of mass of the system.

- The center of mass is the average location of all the components of a system weighted by mass. For 2 stars with vector positions $\vec{r}_1$ and $\vec{r}_2$, the center of mass is located at

$$\vec{R}_{\text{com}} = \frac{m_1\vec{r}_1 + m_2\vec{r}_2}{m_1 + m_2},$$

where the $m_1 + m_2$ is the total mass.

- Let's solve the 2-body problem. We have 2 stars, each of which obeys Newton's second law, $F = ma$. For star 2,

$$m_2\frac{d^2\vec{r}_2}{dt^2} = \vec{F},$$

where $\vec{F}$ is the gravitational force exerted by star 1. We can write a similar equation for star 1, except this time it's $-\vec{F}$ because of Newton's third law: The forces must be equal and opposite.

$$m_1\frac{d^2\vec{r}_1}{dt^2} = -\vec{F}$$

- When we add these equations, the left side can be rewritten as the second derivative of $m_1\vec{r}_1 + m_2\vec{r}_2$, which is proportional to the center of mass vector.

$$m_1\frac{d^2\vec{r}_1}{dt^2} + m_2\frac{d^2\vec{r}_2}{dt^2} = 0 \longrightarrow \frac{d^2}{dt^2}\left(m_1\vec{r}_1 + m_2\vec{r}_2\right) = 0$$

$$\frac{d^2\vec{R}_{\text{com}}}{dt^2} = 0$$

- We've just proven that the location of the center of mass does not accelerate. The 2 stars pull each other around, but the center of mass moves in a straight line at a constant speed.

- This invites us to work in a reference frame that moves along with the center of mass. In other words, we'll choose the origin of our coordinate system to be at the center of mass so that

$$m_1\vec{r}_1 + m_2\vec{r}_2 = 0.$$

- Next, let's go back to our equations of motion and strategically subtract them. We'll multiply the first equation by m_1 and the second by m_2 and then subtract. The product of the masses times the second derivative of the relative separation, $(\vec{r}_2 - \vec{r}_1)$, is equal to the total mass times $\vec{F}$.

$$\begin{aligned} m_1m_2\frac{d^2\vec{r}_2}{dt^2} &= m_1\vec{F} \\ -\quad m_1m_2\frac{d^2\vec{r}_1}{dt^2} &= -m_2\vec{F} \\ \hline m_1m_2\frac{d^2}{dt^2}\underbrace{(\vec{r}_2 - \vec{r}_1)}_{\vec{r}} &= (m_1 + m_2)\vec{F} \end{aligned}$$

- Let's use the shorthand $\vec{r}$ for the relative separation, and for $\vec{F}$, let's plug in Newton's law of gravity:

$$m_1 m_2 \frac{d^2\vec{r}}{dt^2} = -(m_1 + m_2)\frac{Gm_1 m_2}{r^2}\,\hat{r},$$

where $\hat{r}$ is the unit vector in the r direction.

- The product of masses cancels out, and we're left with a simple equation:

$$\frac{d^2\vec{r}}{dt^2} = -\frac{G(m_1 + m_2)}{r^2}\,\hat{r}.$$

- This looks just like the equation of motion for a single mass—which is great because we already solved the 1-body problem in lectures 4 and 5. The only difference is that what used to be M, the mass of the Sun, is now replaced by the total mass of both bodies.

- We previously proved that this equation leads to Kepler's 3 laws, so Kepler's laws also apply to binary stars.

    - The first law says that the vector r traces out an ellipse, although in this case the origin is the center of mass, not the Sun.

    - The second law applies, so the relative speed of the stars will rise as they approach each other.

    - The third law also applies. The only differences are that the M in Kepler's third law is the total mass of both stars and the a is the semimajor axis of the relative orbit.

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

- That means if we can measure a and P for a binary star system, we can calculate the total mass.

- We'd also like to know m_1 and m_2 individually, not just their total. For that, we use the fact that the center of mass is at the origin:

$$m_1\vec{r}_1 + m_2\vec{r}_2 = 0,$$

which implies

$$\vec{r}_2 = -\frac{m_1}{m_2}\vec{r}_1.$$

- And because $\vec{r} = \vec{r}_2 - \vec{r}_1$, we can replace $\vec{r}_1$ with $\vec{r}_2 - \vec{r}$ and then solve for $\vec{r}_2$ again, which leads to

$$\vec{r}_2 = -\frac{m_1}{m_2}(\vec{r}_2 - \vec{r})$$
$$\vec{r}_2\left(1 + \frac{m_1}{m_2}\right) = \frac{m_1}{m_2}\vec{r}$$
$$\vec{r}_2 = \frac{m_1}{m_1 + m_2}\vec{r}.$$

- We already knew that the relative separation $\vec{r}$ traces out an ellipse, but now we know that star 2 itself moves in an ellipse but scaled down—by an amount that depends on the mass of star 1.

- When you do the same thing for star 1, you get

$$\vec{r}_1 = -\frac{m_2}{m_1 + m_2}\vec{r}.$$

- The minus sign means that star 1 is moving in an ellipse with the opposite orientation.

- Each star moves in an ellipse, and the 2 ellipses have a common focus: the center of mass. The size of each ellipse is proportional to the other star's mass. The ratio of semimajor axes, a_2/a_1, equals m_1/m_2. This makes sense; the heavier body doesn't move as much.

- As they move, the stars are always on opposite sides of the center of mass. They also have equal and opposite momenta. We can see that by taking the derivative of our center-of-mass equation, giving

$$m_1\vec{r}_1 + m_2\vec{r}_2 = 0$$
$$m_1\vec{v}_1 + m_2\vec{v}_2 = 0$$
$$m_1\vec{v}_1 = -m_2\vec{v}_2.$$

- Their velocities are always in opposite directions, and the ratio of their speeds tells us the mass ratio: $v_1 / v_2 = m_2 / m_1$. The heavier star moves more slowly.

- If we can measure the relative sizes of the orbits, or orbital speeds, we learn the mass ratio, and if we measure the orbital period, we can use Kepler's third law to learn the total mass—giving us enough information to solve for m_1 and m_2.

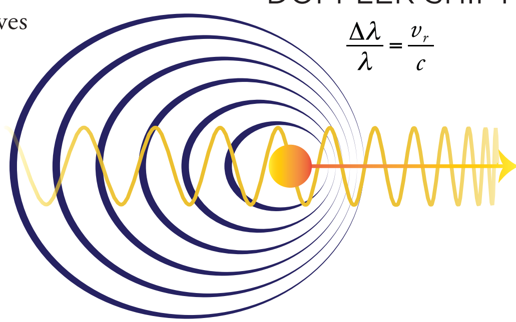
- But how do we measure the sizes of the orbits and orbital period? If our telescope has good enough angular resolution to see both stars as distinct points of light, we can track them over the course of at least 1 full orbit.

- Some binaries are close enough to resolve, but not many. In general, we need to rely on a more indirect method, based on spectroscopy: Doppler spectroscopy.

- The Doppler effect is the shift in wavelength that you observe whenever the source of the waves is moving. For example, when a car is speeding past you, the pitch is higher when the car is coming at you and lower when it goes away. The sound waves from the approaching car are compressed, and shorter sound waves mean higher pitch. But when the car is speeding away, the waves get stretched out, and longer waves mean lower pitch.

DOPPLER SHIFT

$$\frac{\Delta\lambda}{\lambda} = \frac{v_r}{c}$$



- The same thing happens when you have a moving source of light. The fractional shift in wavelength is equal to v_r/c, where c is the speed of light and v_r is the radial velocity: the component of the velocity along the line between the star and us. That's the only component of the velocity that produces a Doppler shift.

- When a star is moving toward us, the wavelengths of all its absorption lines get shifted toward the blue end of the spectrum. When it's moving away, the shift is toward the red end. If we measure that shift, we can calculate the radial velocity.

- Even though the 2 stars of Algol are blended together in our images, the spectrum of that single point of light reveals 2 different sets of absorption lines. They're shifted in wavelength with respect to each other because the stars are moving at different speeds, and as the stars go around, we can watch those lines shift back and forth.

- By using eclipsing binaries, we can measure stellar masses and make precise measurements of the sizes of the stars by tracking the changes in brightness during eclipses. Even though all we see is a point of light, we can nevertheless measure the masses and sizes of both stars, sometimes to within a few percent.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 7.

Doyle, et al., "Kepler-16."

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 5.

Maoz, *Astrophysics in a Nutshell*, chap. 2.

Ryden and Peterson, *Foundations of Astrophysics*, chap. 13.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 14

PLANETS AROUND OTHER STARS

The discovery of planets around other stars—which we call extrasolar planets, or simply exoplanets—didn't get underway until the mid-1990s. That's when technology advanced to the point that exoplanets could be detected.

For exoplanets, the approach of direct imaging—which would involve making a sharp image of a nearby star and then searching for any points of light going around the star—turns out to be nearly impossible. This is because planets are so small and faint compared to stars.

In fact, an Earth twin is more than a billion times fainter than its star. To order of magnitude, the planet is like a firefly buzzing a few meters away from a giant searchlight in Las Vegas and we're trying to detect it using a telescope in New York City.

DOPPLER SPECTROSCOPY

- As with binary stars, the Doppler and eclipse techniques are where most of our knowledge about exoplanets comes from. The physics is the same as with binary stars, but there are some practical differences because of the extreme contrast between planets and stars.

- For example, a spectrum only shows the absorption lines from the star, not the planet. The planet is way too faint. That means we can track the star's radial velocity with the Doppler method, but not the planet's. But let's find out if we can still learn something interesting.

- We'll use the equations from the previous lecture, but instead of m_1, we'll use $m_\star$ for the star, and instead of m_2, we'll use m_p for the planet.

- For a circular orbit, the maximum amplitude of the star's radial velocity is the circumference of the star's orbit around the center of mass, divided by the period, times the geometrical factor of sin(I) that picks out the line-of-sight, or radial, component of the star's velocity.

$$\frac{2\pi a_\star}{P}\sin I$$

- Through Doppler spectroscopy, we can measure that amplitude, and we can measure the orbital period. That means we can calculate $a_\star \sin(I)$—which is not very interesting in itself but does tell us something about the planet's mass.

- Recall that both the star and the planet are orbiting the center of mass, and from the definition of the center of mass, we know that $m_\star a_\star = m_\mathrm{p} a_\mathrm{p}$. We can use that fact to eliminate $a_\star$ from the equation.

$$\frac{2\pi a_\mathrm{p}}{P}\frac{m_\mathrm{p}}{m_\star}\sin I$$

- Next, let's bring in the information about the orbital period. We'll use Kepler's third law, which connects the orbital period with the total mass and the total orbital separation, $a_\star + a_\mathrm{p}$. In this case, though, the total mass is very nearly equal to the mass of the star. And because the star doesn't move as much as the planet, the total separation is nearly equal to a_p. That gives an expression for a_p that we can substitute into the radial velocity equation.

$$a^3 = \frac{G(m_\star + m_\mathrm{p})}{4\pi^2} P^2$$

$$a_\mathrm{p}^3 \approx \frac{Gm_\star}{4\pi^2} P^2$$

$$\approx \frac{2\pi}{P} \left(\frac{Gm_\star}{4\pi^2} P^2 \right)^{1/3} \frac{m_\mathrm{p}}{m_\star} \sin I$$

- After tidying up, we notice that the radial velocity amplitude is proportional to the planet mass.

$$\approx \left(\frac{2\pi G}{P} \right)^{1/3} \frac{m_\mathrm{p}}{m_\star^{2/3}} \sin I$$

- We don't have enough information to solve for both masses, but if we already have a reliable estimate for the star's mass—based on its similarity to other stars for which we can measure the mass—then we can calculate the planet's mass from this equation.

- Actually, because sin(I) is on the right side, we can calculate $m_p$sin(I), but not m_p by itself. That's not ideal, but at least it gives us a lower bound on the planet mass. If $m_p$sin(I) is 1 Earth mass, then m_p must be 1 or larger, because the sine of an angle is always 1 or smaller.

- So, Doppler spectroscopy reveals the orbital period and the minimum mass of the planet. In addition, the Doppler method reveals the orbital eccentricity from the way the radial velocity varies over the course of a full period. A circular orbit will show a sinusoidal variation, but an eccentric orbit will be faster in some parts and slower in others, leading to a skewed, nonsinusoidal signal.

PLANETARY ECLIPSES

- When the planet eclipses the star, we get even more information. Whenever we detect eclipses, we know we must be viewing the orbit nearly edge on, which implies that $\sin(I)$ is very close to 1. In that case, the Doppler signal tells us the planet mass without any ambiguity.

- For an Earth-mass planet orbiting a solar-mass star with a period of 1 year, the star's orbital velocity is 9 centimeters per second.

$$
\begin{aligned}
v_{\star,\mathrm{r}} &\approx \left(\frac{2\pi G}{P}\right)^{1/3} \frac{m_\mathrm{p}}{m_\star^{2/3}} \\
&\approx 9\,\mathrm{cm\,s^{-1}} \left(\frac{P}{1\,\mathrm{year}}\right)^{-1/3} \left(\frac{m_\star}{M_\odot}\right)^{-2/3} \frac{m_\mathrm{p}}{M_\oplus}
\end{aligned}
$$

- To get the corresponding wavelength shift, we need to divide by the speed of light, resulting in 3 parts in 10 billion.

$$\frac{\Delta\lambda}{\lambda} = \frac{v_{\star,\mathrm{r}}}{c} = 3 \times 10^{-10}$$

- Nobody has ever achieved this level of precision; we have not yet detected an Earth twin with this method.

- But if we make the planet 10 times more massive than the Earth or shrink the orbital period by a factor of 1000, then the star's velocity increases to about 1 meter per second, which can be achieved with present technology.

- The good news is that there are many planets more massive than the Earth, with periods much shorter than a year.

- While we can calculate the radii of stars based on eclipse durations, this doesn't work well for planets. Instead, we take advantage of the fact that the planet is dark—essentially black compared to the star. So, when the planet goes in front of the star, it's as though the stellar disk has a circular black spot. The fraction of starlight being blocked is the area of the planet's silhouette divided by the area of the stellar disk, which is equal to the square of the radius ratio between the planet and the star.

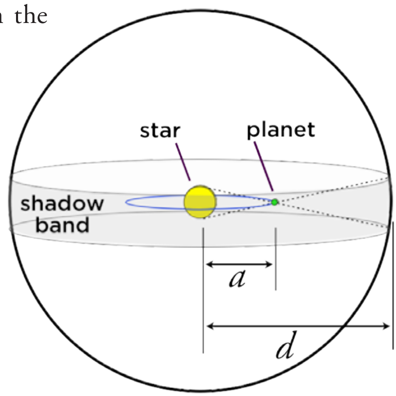
$$\frac{\Delta F}{F} = \frac{\pi R_{\mathrm{p}}^2}{\pi R_\star^2} = \left(\frac{R_{\mathrm{p}}}{R_\star}\right)^2 = 8.4 \times 10^{-5} \left(\frac{R_{\mathrm{p}}/R_\star}{R_\oplus/R_\odot}\right)^2$$

- If we were on an alien planet monitoring the Sun as the Earth crossed in front, the Sun would appear to get fainter by 84 parts per million. That's not easy to detect, but at least now we're working with parts per million, not billion.

- Measuring fluxes with a precision of parts per million is very difficult from beneath the Earth's atmosphere, but we can do it using space telescopes. This is why almost all of our knowledge about Earth-sized planets comes from space telescopes.

- Eclipsing planets are astrophysical treasures, just like eclipsing binaries. Unfortunately, they're rare treasures. It takes a special coincidence to be viewing the planetary system from just the right angle.

- To find the odds of this happening, let's draw the imaginary celestial sphere surrounding a star with a planet. The star illuminates the sphere, but the planet casts a shadow, and as it orbits, the planet's shadow traces out a band on the celestial sphere.



- If there are observers all over the galaxy monitoring the star from all possible directions, the only ones who see eclipses are the ones inside the shadow band. So, the probability of seeing eclipses is equal to the area of the shadow band divided by the total area of the sphere.

- To calculate that probability, let's assume that $R_\star$, the radius of the star, is much smaller than a, the radius of the planet's orbit, which in turn is much smaller than d, the distance to the celestial sphere.

- From the planet's point of view, the angular width of the shadow band is equal to the angular diameter of the star, $2R_\star/a$. The shadow band makes a thin ribbon around the celestial sphere with a width equal to the angular width, $2R_\star/a$, times the distance, d. The area of the shadow band is equal to its circumference, $2\pi d$, times its width, and if we divide that area by the total area of $4\pi d^2$, we find the eclipse probability to be $R_\star/a$.

$$\text{prob.} = \frac{2\pi d \times 2R_\star d/a}{4\pi d^2} = \frac{R_\star}{a} = 0.005\left(\frac{R_\star/a}{R_\odot/1\,\text{AU}}\right)$$

- If we plug in 1 solar radius for $R_\star$ and 1 AU for a, the probability is 1/215, which is approximately 0.005, or about half a percent. This means we need to monitor hundreds of Sunlike stars with Earthlike planets before we're likely to find even 1 that eclipses.

- Even worse, the eclipses are brief and easy to miss. The maximum duration is the diameter of the star divided by the planet's orbital speed, which we can also calculate in terms of the orbital period and the star's average density, using Kepler's third law.

$$\text{max. dur.} = \frac{2R_\star}{v_\text{p}} = 13\,\text{hours}\left(\frac{P}{1\,\text{year}}\right)^{1/3}\left(\frac{\rho_\star}{\rho_\odot}\right)^{-1/3}$$

- For the Earth crossing the Sun, the maximum duration is 13 hours.

- The goal, then, is to monitor the brightness of hundreds of stars, waiting for the one day each year when one of the stars might dip in brightness by 84 parts per million.

- That was precisely the goal of a NASA mission called Kepler, a 1-meter space telescope that spent 4 years monitoring several hundred thousand stars, a few thousand of which were bright enough that Kepler could have detected eclipses by planets just like the Earth.

- And Kepler did find some—a few dozen—but it was a struggle because the eclipse signals are so tiny. The primary mission ended in 2013, and since then, astronomers have been arguing about which of those signals represent real planets and which are other types of astronomical phenomena or simply noise fluctuations that happen to mimic a planetary signal.

- In 2018, NASA launched a new spacecraft called TESS, which stands for Transiting Exoplanet Survey Satellite, that will also detect eclipsing exoplanets but, this time, around brighter stars that are closer to the Earth. That will make any newly discovered planets easier to confirm and study.

A transit is the passage of a small body in front of a larger body.

- TESS isn't specifically designed to find Earth twins. That quest is only part of what makes exoplanetary science so interesting. Even though Kepler only found a few dozen potentially Earthlike planets, it found thousands of planets very different from Earth—different from any of the planets in the solar system.

THEORIES AND RULES OF PLANETARY SCIENCE

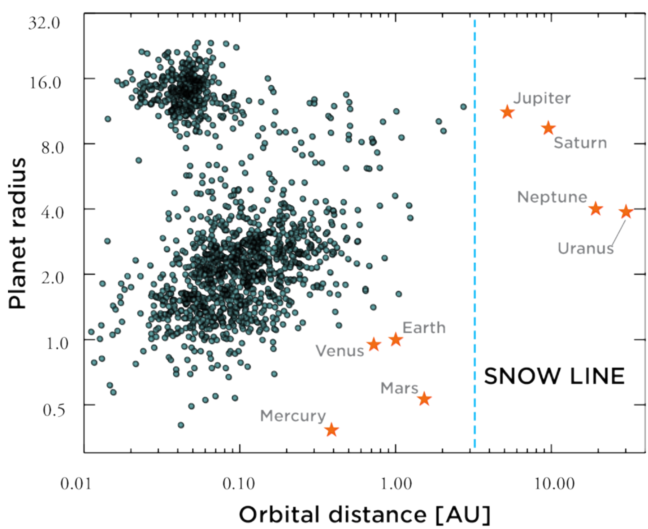
- Let's review the expectations we had prior to the discovery of exoplanets based on the patterns we see in the solar system.

- In the solar system, all the planets have nearly circular orbits, and all the orbits are aligned with each other: They lie flat, in a single plane, to within a few degrees. Those 2 patterns are evidence for planet formation within a flat, circular disk of material swirling around a newborn star.

- A third pattern is that all the small, rocky planets are close to the Sun and the gas giants are farther away. This is the observation that led to the theory of core accretion, in which all planets start out rocky. They can only transform into gas giants far away from the star—beyond the so-called snow line, where it's cold enough for water, methane, and ammonia to freeze, providing more solid material to build a planet massive enough to be able to accrete hydrogen gas and puff up to become a gas giant.

- According to core accretion theory, the most massive planets, which produce the biggest signals, should have very wide orbits with periods of decades, making them hard to detect with the Doppler or eclipse techniques. But this is not what we actually found.

- A logarithmic chart of orbital distance against planet size for all the known eclipsing planets shows that for solar system planets, the snow line is at around 3 or 4 AU. The data for exoplanets shows that there are thousands of them, many of which are very close to the star, at distances of a tenth or even a hundredth of an AU.



- The chart shows many planets that are as large as Jupiter, but instead of having orbits at 5 AU, like the real Jupiter, theirs are smaller than a tenth of an AU. These are called hot Jupiters, and their existence seems to contradict the story about core accretion and the necessity for giant planets to form beyond the snow line.

- However, this chart is not a representative sample of exoplanets. It's a sample of detected planets, and as such, it exaggerates the abundance of the planets that are easy to detect—namely, large planets in tiny orbits. The chart makes it seem like hot Jupiters are very common, but in fact, they only occur around half a percent of Sunlike stars.

- Likewise, the absence of exoplanets with orbits wider than 1 AU on this chart does not imply that such planets are intrinsically rare; all it means is that we have a hard time detecting them. The eclipse probability is too low.

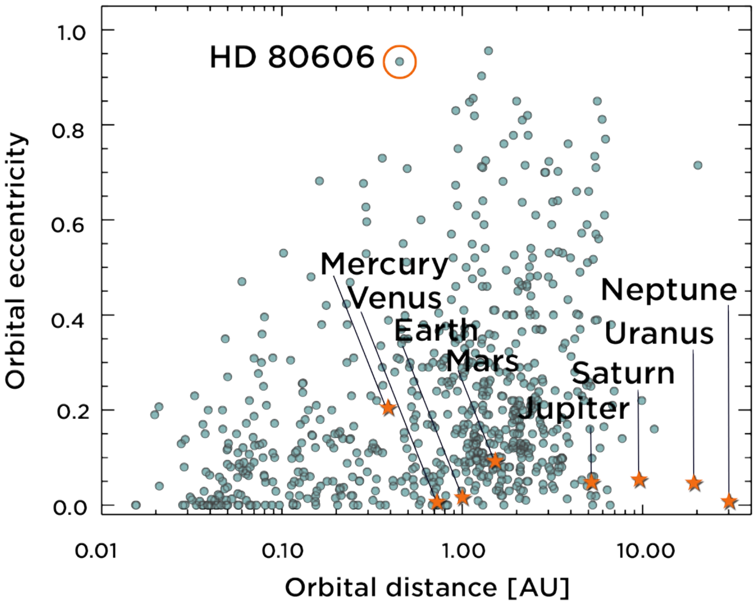
- The most commonly detected planets have sizes in between those of Earth and Neptune and periods of less than a year. We find them in tiny multiplanet systems, such as Kepler-11, which has 6 planets scrunched inside what would be Venus's orbit around the Sun. These compact multiplanet systems occur around 30% of Sunlike stars, but they were unanticipated by theorists.

- How did that happen? What was so wrong with the theory of planet formation?

- Some theorists say that the theory of planet formation was correct but incomplete; nobody had thought hard enough about what might happen after planets form. Their orbits can get rearranged and shrink through interactions between planets and the gaseous disk, or close encounters between planets, or tidal effects from a passing star.

- Other theorists are ready to abandon the idea that the snow line plays a crucial role in giant planet formation. To make progress, we need to keep looking for patterns among the exoplanets.

- We can change the chart so that it shows orbital eccentricity against orbital distance (semimajor axis) and plot the data for planets found through the Doppler method, rather than eclipses, because it's the Doppler data that reveal the eccentricity.



- The solar system planets have low eccentricities—nearly circular orbits. But the exoplanets have eccentricities that range all the way up to 0.9! Another supposed rule of planetary science is broken.

- Another pattern in the chart is that the planets with the smallest orbits tend to have small eccentricities—more circular orbits. That's one pattern about exoplanets that we actually do understand. It's a consequence of tidal forces.

- When a planet is close to a star, it gets slightly stretched by tidal forces. If the orbit is eccentric, the stretching varies with time: It's stronger when the planet approaches the star and weaker when it recedes. The friction associated with that constant stretching converts orbital energy into heat, and when an orbit loses energy without changing its angular momentum, the orbit becomes circular.

For more information about exoplanets, check out Joshua Winn's Great Course *The Search for Exoplanets: What Astronomers Know.*

READINGS

Carter, et al., "Kepler-36."

de Pater and Lissauer, *Planetary Sciences*, chap. 12.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 5.

NASA Exoplanet Archive, https://exoplanetarchive.ipac.caltech.edu/.

Perryman, *The Exoplanet Handbook*.

Ricker, et al., "Transiting Exoplanet Survey Satellite."

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Winn and Fabrycky, "The Occurrence and Architecture of Exoplanetary Systems."

Lecture 15

WHY STARS SHINE

The Earth intercepts less than a billionth of the Sun's total luminosity—which is 3.8×10^{26} watts—yet even that tiny fraction amounts to 10,000 times more power than all the world's power plants put together. In addition, the Sun has been shining for 4.6 billion years, based on evidence from primitive meteorites. And geologists have found evidence that liquid-water oceans have also existed for billions of years, so the Sun's luminosity couldn't have been much lower in the past or else the oceans would have frozen. With these numbers, the Sun's total energy output since it formed can be calculated by multiplying the Sun's luminosity by its age, giving 5.5×10^{43} joules of energy. Where did all that energy come from? The answer is nuclear fusion.

Chemical reactions involve electrons changing orbits or getting traded between nuclei, and the energy scale for those transitions is always on the order of 1 electron volt per proton mass, or 100 megajoules per kilogram. A chemical fuel with orders of magnitude better than that cannot be found.

On the other hand, nuclear reactions involve rearrangements of nucleons—protons and neutrons. The typical energies involved with those rearrangements are millions of electron volts. That's what gives the Sun its longevity.

A CRASH COURSE IN NUCLEAR PHYSICS

- An atomic nucleus is the tiny bundle of protons and neutrons at the center of a cloud of electrons. In a neutral atom, the number of protons equals the number of electrons because they have equal and opposite electric charges. The number of neutrons is about the same as, or somewhat higher than, the number of protons.

A carbon atom has 6 electrons around a nucleus of 6 protons and 6 neutrons. Some carbon nuclei, though, have 7 neutrons. We say that carbon has 2 main isotopes: nuclei with the same number of protons but different numbers of neutrons.

Why do we call them both carbon if their nuclei are different? It's because elements are named by chemists, not nuclear physicists. The chemical properties of an atom depend almost entirely on the number of electrons, which is equal to the number of protons. The neutron count doesn't matter much.

But it matters a lot for nuclear reactions. When carbon is involved in a nuclear reaction, we need to know which isotope we're talking about. The convention is to specify the total number of nucleons—protons plus neutrons—which is often called the atomic mass (*A*). For example, carbon-12 has 6 protons and 6 neutrons, so $A = 12$, and it's written like this: ^{12}C.

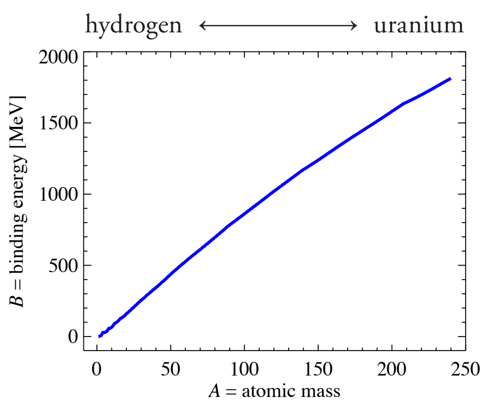
- A big difference between chemical and nuclear reactions is the spatial scale over which the changes occur. The smaller scale of nuclei and the millionfold more energy released are consequences of the strength of the strong nuclear force, and the short range over which it acts, compared to electromagnetism.

- In chemistry, a key concept is the heat of formation: the amount of energy released if you make a molecule from its elemental constituents. For example, if you mix hydrogen and oxygen to make water, you release 2.5 electron volts per water molecule. That's the heat of formation of water.

- The analogous concept in nuclear physics is the nuclear binding energy, defined as the energy released if you make a nucleus out of free protons and neutrons. Or, equivalently, it's the energy required to completely disintegrate a nucleus. Carbon-12, for example, has a binding energy of 92 million electron volts (MeV).

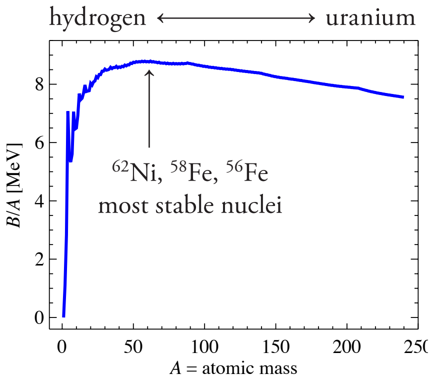
- The binding energies of all the most stable isotopes have been measured in laboratory experiments. On a chart of the measured binding energy, B, versus atomic mass, A, the x-axis ranges from hydrogen at $A = 1$ to uranium at 238.

- The data form an upward sloping, nearly straight line. That makes sense: The bigger the nucleus, the harder we need to work to tear it apart. The slope, B/A, tells us *how* hard—based on this chart, it's about 2000 MeV/250, or 8 MeV per nucleon.



- But despite appearances, the line is not exactly straight. To get a clearer view, let's plot B/A (instead of B) versus A.

- Now we see that it's not as simple as 8 MeV per nucleon. Small nuclei have a relatively low binding energy per nucleon. B/A rises to a maximum value at an atomic mass of around 60.



- Why does the binding energy curve have this shape? Think back to our mental model of a nucleus as a cluster of marbles. Each one is coated with glue—that's the strong nuclear force. If a marble is surrounded by other marbles, it takes about 8 MeV of energy to overcome all the glue and pull it out. That's the average value of B/A.

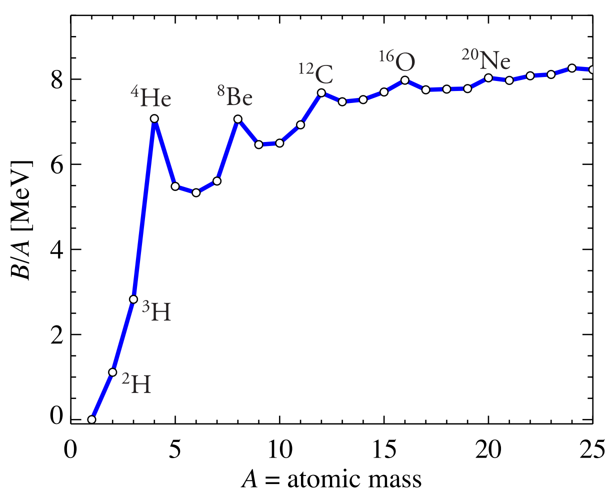
- But that's not all there is to it. Some marbles are near the surface of the cluster. They're not surrounded by marbles in all directions, so they're not glued as tightly in place. That lowers the binding energy relative to the average.

- This "surface penalty" is especially bad for small nuclei, because they have a higher surface-to-volume ratio. The surface area of a sphere goes as radius squared, and volume goes as radius cubed, so the surface-to-volume ratio varies as 1 divided by the radius, meaning it gets worse for small nuclei.

$$\begin{array}{r} \text{Surface area: } S \propto R^2 \\ \text{Volume: } V \propto R^3 \end{array} \quad \longrightarrow \quad \frac{S}{V} \propto \frac{1}{R}$$

- Another problem is that some of the marbles—the protons—repel each other. The electric repulsion works against the glue, lowering the binding energy. So, there's a penalty for having protons.

- Then why would a nucleus want to have any protons? Why don't we see nuclei made entirely from neutrons?
- It's because of another effect called symmetry energy. The "glue" sticking nucleons together is a little stronger when it's between a proton and a neutron than it is between 2 protons or 2 neutrons. That increases the binding energy of nuclei with lots of proton-neutron pairs.
- Now we can understand the shape of the binding energy curve. Light nuclei suffer from high surface area; that's why their binding energy is lower than average. Very heavy nuclei also have low binding energies, because they suffer from an internal conflict: As you keep adding nucleons to a nucleus, you'd like to avoid the proton penalty and stick with neutrons, but to satisfy the desire for symmetry, you need to add protons. The net effect is to destabilize the nucleus.
- Notice that there are some spikes in the binding energy curve. Let's get a better look at them by zooming in on the lightest nuclei.



- These enhancements in binding energy come from aspects of the strong force that favor even numbers of protons and even numbers of neutrons. Helium-4, for example, has 2 of each, so it has especially high binding energy. Also in this category are beryllium-8, carbon-12, oxygen-16, and neon-20.

EXPLAINING THE SUN'S ENERGY: NUCLEAR FUSION

- We know from its spectrum that the Sun is mostly hydrogen. Could it be using hydrogen as a fuel, fusing it into helium?

- The stable nucleus of hydrogen, H-1, is just a proton, with a binding energy of 0. Helium-4 has a binding energy of 7 MeV per nucleon. So, hydrogen fusion releases 7 MeV per nucleon, which is equivalent to 670 million megajoules per kilogram.

- But is that enough to explain the Sun's total energy output? Let's calculate the required mass of hydrogen: the Sun's total energy output divided by the energy per kilogram released by fusion.

$$\begin{aligned}\frac{5.5 \times 10^{43}\ \mathrm{J}}{670{,}000{,}000\ \mathrm{MJ\,kg^{-1}}} &= 8.2 \times 10^{28}\ \mathrm{kg} \\ &= 0.04\, M_{\odot}\end{aligned}$$

- By converting only 4% of its hydrogen into helium, the Sun can release enough energy to shine for billions of years.

- There is a close relationship between mass and energy, which Einstein taught us about with his most famous equation, $E = mc^2$. This equation means that neither energy nor mass is conserved. Interactions can change one into the other, and the exchange rate is $E = mc^2$. Mass and energy turn out to be the same kind of stuff at a fundamental level; it's their combination that's conserved.

- An important implication is that when an interaction releases energy—such as when fusing hydrogen—the total mass of all the particles involved must decrease.

$$\Delta m = \frac{\Delta E}{c^2}$$

- This happens during chemical reactions, but we never notice it. The changes in mass are only at the level of parts per billion. But for nuclear reactions, the changes are more substantial.

- Let's calculate the fractional change in mass, $\Delta m/m$, for hydrogen fusion. We take the 7 MeV per nucleon that is released during fusion and divide by c^2 to get the corresponding mass loss and then divide by the mass of a proton to make it a fraction.

$$\frac{\Delta m}{m} = \frac{7\,\mathrm{MeV}/c^2}{m_\mathrm{p}} = 0.0075$$

- This means that a helium nucleus is less massive, by 0.75%, than the sum of 2 isolated protons and 2 isolated neutrons. This difference in mass can be measured in laboratory experiments.

- All of this implies that the Sun, like anything that radiates energy, is losing mass. And we can calculate how much. In time dt, the Sun emits energy dE and loses mass dE/c^2. So, the rate of mass loss is dE/dt divided by c^2, and dE/dt is the Sun's luminosity. Dividing the Sun's luminosity by c^2, we get

$$\frac{dM_\odot}{dt} = \frac{dE/c^2}{dt} = \frac{L_\odot}{c^2}$$
$$= 4 \times 10^9\,\mathrm{kg\,s^{-1}}.$$

- This means that every second, 4 billion kilograms of mass are vanishing from the Sun, having been converted into energy and flung out into space. But don't be too concerned. The Sun won't disappear any time soon. Its total mass is 2×10^{30} kilograms, so even after 10 billion years, the Sun will lose only about 0.07% of its mass in this way.

PROCESS OF CONVERTING PROTONS INTO HELIUM

- Even after it became clear that the Sun's luminosity has a nuclear origin, the sequence of nuclear reactions was the subject of confusion and debate for decades. We can't expect 4 protons to crash into one another all at the same time. Such 4-body collisions are always vanishingly rare. The only plausible reactions are 2-body collisions. So, you'd think the first step would be to smash 2 protons together to make helium-2. But there are a few problems.

- The first one is that protons repel each other. How do we get them to approach each other closely enough for the strong force to take over?

- We can represent the problem schematically by plotting the potential energy of a proton as a function of its distance from another proton. As we bring the protons together, the electric potential energy (E_e) rises; it's $\eta e^2/r$, where η is the Coulomb constant and e is the proton's charge.

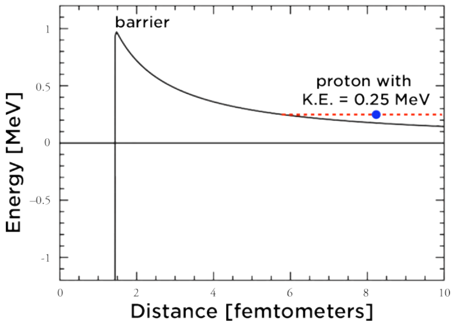
$$E_{\mathrm{e}} = \frac{\eta\, e^2}{r}$$

- But if they manage to get within 1.5 femtometers, or 1.5×10^{-15} meters, the strong force takes over and the energy drastically decreases. The marbles come close enough to touch—and stick.

- To find the maximum height of the potential energy barrier, we evaluate the electrical potential energy of 2 protons separated by 1.5 femtometers, which comes out to be 1 MeV.

$$\frac{\eta e^2}{1.5\,\mathrm{fm}} = 1\,\mathrm{MeV}$$

- So, if we want the proton to stick, we need to throw it at the other proton with at least 1 MeV of kinetic energy. If it only has a quarter of an MeV, the repulsion will cause it slow down, halt at 5.5 femtometers, and then fly away.



- Do the protons at the center of the Sun have at least 1 MeV of kinetic energy?

- We have good reason to believe that the temperature at the center of the Sun is 15 million Kelvin, and we know that the average kinetic energy per particle in a gas at temperature T is $3/2kT$.

$$T = 15 \times 10^6\,\mathrm{K}$$

$$\langle E_\mathrm{k} \rangle = \frac{3}{2}kT = 0.002\,\mathrm{MeV}$$

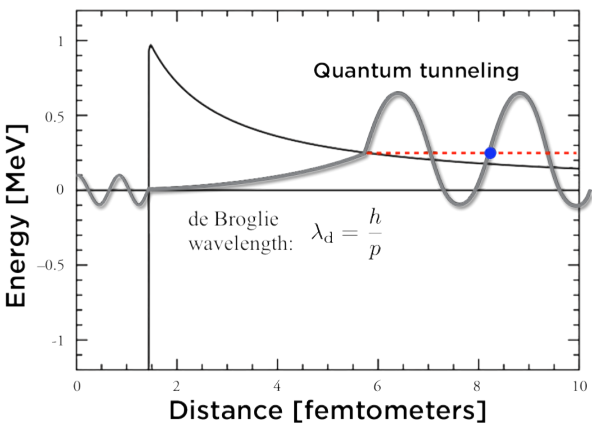
- According to this calculation, the Sun is way too cold! The protons should have no chance of coming close enough to stick. So, how do they even get started making helium?

- The answer is that quantum theory helps the protons.

- The time has come to set aside our mental model of protons as marbles. Like any fundamental particle, protons also have a wavelike nature and are described by a wave function—a cloud of probability extending over a region of space.

- Unlike classical particles, wave functions do not simply halt when they reach a potential energy barrier. They leak through. The amplitude of the wave function decreases exponentially as it leaks through, so the portion that gets through the barrier is tiny. Nevertheless, there's a chance for a proton to tunnel through the barrier—even though, officially, it doesn't have enough energy. This phenomenon is called quantum tunneling, and it's governed by the Schrödinger equation.

- Let's perform an order-of-magnitude calculation to establish how much energy a proton needs to have a decent chance of tunneling through the barrier.



- The key concept is the de Broglie wavelength, λ_d, defined as h/p, which is Planck's constant over momentum. You can think of λ_d as the spatial extent of a particle's cloud of probability. When you observe the particle, it won't necessarily be located where you would expect based on classical physics, but you'll find it within about 1 de Broglie wavelength of the expected location.

- This means that we don't need the protons to get within 1.5 femtometers of each other. It's enough that their wave functions overlap to within about 1 de Broglie wavelength. That turns out to be much easier; it doesn't require as much energy.

- The de Broglie wavelength is defined in terms of momentum, but we're not interested in momentum. Instead, we're interested in kinetic energy, $1/2mv^2$.

$$E_{\mathrm{k}} = \frac{1}{2}mv^2 = \frac{p^2}{2m} \quad \longrightarrow \quad p = \sqrt{2mE_{\mathrm{k}}}$$

- So, we can rewrite the de Broglie wavelength as

$$\lambda_{\mathrm{d}} = \frac{h}{p} = \frac{h}{\sqrt{2mE_{\mathrm{k}}}}.$$

- For the protons to approach within a de Broglie wavelength, the required energy is

$$E_{\mathrm{k}} > \frac{\eta\, e^2}{\lambda_{\mathrm{d}}}.$$

- This leads to an equation with energy on both sides, which we can solve for energy and then plug in the numbers.

$$\begin{aligned} E_{\mathrm{k}} &> \frac{\eta\, e^2 \sqrt{2m_{\mathrm{p}}E_{\mathrm{k}}}}{h} \\ \sqrt{E_{\mathrm{k}}} &> \frac{\eta\, e^2}{h}\sqrt{2m_{\mathrm{p}}} \\ E_{\mathrm{k}} &> \left(\frac{\eta\, e^2}{h}\right)^2 2m_{\mathrm{p}} \\ E_{\mathrm{k}} &> 0.0025\,\mathrm{MeV} \end{aligned}$$

- The answer is 3 orders of magnitude lower than 1 MeV, the required energy for classical particles. And it's the same order of magnitude as the average kinetic energy of protons at the center of the Sun.

- So, by relying on quantum tunneling, protons can fuse even at the "low" temperature of 15 million Kelvin.

- An important aspect of quantum tunneling is the exponential falloff in the wave function as it penetrates a barrier. Because of that, the probability to tunnel through a barrier depends very sensitively on energy. Small changes in energy lead to large changes in the fusion rate.

- Rates of thermonuclear reaction—where the nuclei rely on random kinetic energy to come together and fuse—depend very strongly on temperature. The rate of proton fusion in the Sun varies as T^4. Other important reactions vary as T^8, or even higher powers. In practice, this means that each reaction has a sharply defined ignition temperature below which the power output is negligible and above which it's substantial. This has important consequences for stellar structure.

- But there's another problem: Helium-2, the nucleus made of 2 protons, is extremely unstable! So, after all that work bringing the protons together, the nucleus falls apart and we're back to square one.

- What saves the day is the weak nuclear interaction, which can convert protons into neutrons and vice versa.

- Two protons approach each other, and right when their wave functions merge, there's a slight chance that one of them will change into a neutron, spitting out a positron and a neutrino in the process. The proton and the newborn neutron make a nucleus of hydrogen-2, also known as deuterium, which is stable.

- However, the interaction that allows this transformation is weak. The probability is low. If you follow a single proton bouncing around in the core of the Sun, you'll have to wait several billion years, on average, before you see it fuse into deuterium.

- But after that, things start happening faster. Within a few seconds, the deuterium smacks into another proton and sticks, forming helium-3 and emitting a gamma ray photon.

- From there, there are 3 different paths that lead to helium. In the simplest one, given enough time—about 20,000 years—a helium-3 nucleus will find another one and collide, knocking away 2 protons. What's left is 2 protons and 2 neutrons: that's helium-4, the end point, which is stable and interacts no further.

What's the average power output per cubic meter in the core of the Sun?

$$L_\odot = 3.8 \times 10^{26}\,\mathrm{W}$$

$$R_\mathrm{c} = 70{,}000\,\mathrm{km}$$

$$\text{power density of Sun:}\quad \frac{L_\odot}{\frac{4\pi}{3}R_\mathrm{c}^3} = 300\,\mathrm{W\,m^{-3}}$$

The answer, 300 watts per cubic meter, is not very impressive. A kitchen toaster emits 1000 watts of heat.

This is why we should not envision the Sun as a blast furnace with nuclear bombs going off all the time. Instead, it's more like a collection of toasters spaced apart every few meters.

What makes the Sun so luminous is that it's so big—there are a lot of toasters. And, just as important, the overlying material, the outer 90% of the Sun, is optically thick; it traps the energy for a long time, keeping the interior toasty hot.

READINGS

Choudhuri, *Astrophysics for Physicists*, chaps. 3 and 4.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 5.

Maoz, *Astrophysics in a Nutshell*, chap. 3.

Phillips, *The Physics of Stars*.

Ryden and Peterson, *Foundations of Astrophysics*, chaps. 14–18.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 16

SIMPLE STELLAR MODELS

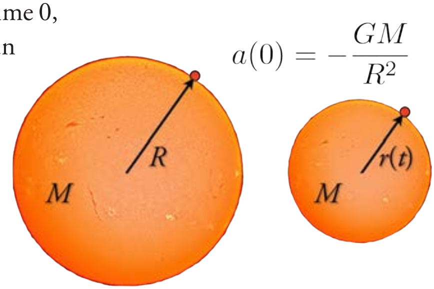
All stars are powered by nuclear fusion. But despite this common power source, stars show a wide range of properties. For stars on the main sequence of the Hertzsprung-Russell diagram, there are patterns, and this lecture will consider these relationships.

The biggest star you can see with your naked eye is Betelgeuse, the star at Orion's upper-left shoulder. It's about 1000 times larger than the Sun! If the Sun were swapped for Betelgeuse, it would engulf all the planets out to Jupiter.

CENTRAL PRESSURE

- How long would it take for a sphere of mass M and radius R to collapse under its own gravity?

- Consider an atom at the surface. At time 0, gravity pulls inward, imparting an acceleration of GM/R^2. As time goes on, the atom falls to a radius $r(t)$. The entire mass is also contracting, so the mass M pulling on the atom is always the same. But because $r(t)$ is decreasing, the acceleration is increasing.



- Acceleration is the second time derivative of position, so we need to solve the second-order differential equation for $r(t)$ and find the time it takes for the atom to reach $r(t) = 0$. Instead of doing this, though, we can perform an order-of-magnitude estimate.

- Let's use T to represent the time to collapse, and instead of trying to solve for $r(t)$, let's replace the functions on both sides of the equation with simple expressions that we have reason to believe have the correct orders of magnitude.

- On the left side, we have acceleration. What's the typical acceleration of a falling atom? We know that the average speed must be R/T because it travels a distance R in a time T. But acceleration is change in speed.

- The atom starts from rest and reaches the average speed somewhere along the way—we don't know where, exactly, but maybe we won't be so far off if we assume it's at time $T/2$. That implies an inward acceleration of magnitude R/T divided by $T/2$, or $2R/T^2$.

- Meanwhile, on the right side, we have GM/r^2. Let's replace that function with the value it takes at the halfway point: $r = R/2$.

- By making those substitutions, we convert the differential equation into a simple algebraic equation, which we can solve for T.

$$a(t) = \frac{d^2r}{dt^2} = -\frac{GM}{r^2}$$

$$v_{\text{avg}} = \frac{R}{T}$$

$$\frac{v_{\text{avg}}}{T/2} = \frac{R/T}{T/2} = \frac{2R}{T^2}$$

$$\frac{GM}{(R/2)^2} = \frac{4GM}{R^2}$$

$$\frac{2R}{T^2} \sim \frac{4GM}{R^2}$$

$$R^3 \sim 2GMT^2$$

$T =$ time to collapse

$$T \sim \sqrt{\frac{R^3}{2GM}}$$ estimate

- If you solve this problem exactly, you get $T = \frac{\pi}{2}\sqrt{\frac{R^3}{2GM}}$, which is only off by a factor of $\pi/2$.

- When we evaluate this free-fall time for a sphere with the Sun's mass (2×10^{30} kilograms) and radius (7×10^{11} meters), we get about half an hour.

$$\frac{\pi}{2}\sqrt{\frac{R_\odot^3}{2GM_\odot}} \approx 0.5\,\text{hour}$$

- So, if gravity were the only force acting, the Sun would collapse into a black hole in half an hour—which is not very long. So how does the Sun survive for billions of years?

- There's a force that opposes gravity: the force that arises from differences in pressure.

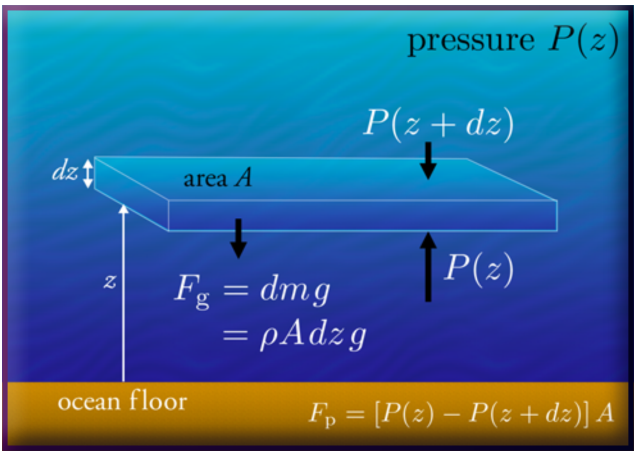
- The pressure of a gas is the force per unit area produced by all the microscopic collisions with the gas particles. When gravity tries to compress a sphere of gas, the contraction brings the particles closer together, increasing the rate of collisions. The contraction also releases gravitational potential energy, which heats the gas, making the collisions more energetic.

- Both effects cause the interior pressure to increase, leading to a net outward force that eventually becomes strong enough to oppose gravity and halt the collapse.

- The equation that expresses that balance between pressure and gravity is called the equation of hydrostatic balance.

- Think about water pressure in the ocean. The pressure at the ocean floor is higher than it is at the surface. So, pressure is not just one number; it's a function of height, $P(z)$.

- Consider a horizontal layer of water at height z with an infinitesimal thickness of dz and a cross-sectional area A. Let's calculate the gravitational force and the pressure forces acting on the water in this layer.



- There are 2 forces: gravity and pressure. Gravity pulls down with a force dmg, where g is the gravitational acceleration near the surface of the Earth and dm is the mass of the infinitesimal layer, which can also be written as density times volume, ρAdz.

- The water in the layer feels inward pressure from every direction. The horizontal forces cancel out, though; the right side gets pushed just as hard as the left side.

- But the vertical forces don't cancel, because pressure decreases with increasing height. The pressure on the top of the water layer is weaker than the pressure on the bottom. So, the net effect of pressure (F_p) is an upward force given by the difference in pressure times the area.

- If the water isn't moving—if it's in balance—then the downward force of gravity (F_g) must equal the upward force from pressure (F_p).

$$F_{\mathrm{p}} = F_{\mathrm{g}}$$

- This leads to an equation in which the *A*'s cancel out.

$$[P(z) - P(z + dz)]\, A = \rho A \, dz \, g$$

$$\frac{P(z + dz) - P(z)}{dz} = -\rho g$$

- This can be rearranged to form the equation of hydrostatic balance.

$$\frac{dP}{dz} = -\rho g$$

- For stars, we need to make a few changes. One is trivial: We replace *z* with *r*, the radial coordinate. More importantly, when we consider a whole star, the density and the gravitational acceleration are not constants—they vary from the center to the surface—so ρ becomes $\rho(r)$ and *g* is replaced by the more general formula.

$$\frac{dP}{dr} = -\rho(r)\,\frac{GM_r}{r^2}$$

- A subtler point is that M is now a function of r, too, so we write it as M_r. It represents all the mass located interior to r. That's from Newton's theorem: We only need to consider the interior mass when calculating the force from a spherically symmetric object. M_r grows from 0 at $r = 0$ to the total mass of the star when r equals the stellar radius, R.

$$M_r = 0, \ M_R = M$$

- Now let's use the equation of hydrostatic balance to estimate the Sun's central pressure, P_c. In the spirit of order-of-magnitude estimates, we can replace the derivative dP/dr with a simple ratio because the pressure decreases from P_c to nearly 0 as we go from the center to the surface.

- And we need to replace the function on the right side with an expression representative of the function's value, so let's replace $\rho(r)$ with the average density. For r, let's use half the total radius, and for M_r, let's use half of the total mass. That turns the differential equation into an algebraic equation, which we can solve for P_c.

- When we plug in the Sun's mass, radius, and average density of 1.4 grams per cubic centimeter, we get 5×10^{14} newtons per square meter—that's 5 billion times higher than the air pressure on Earth.

$$\frac{dP}{dr} = -\rho(r)\,\frac{GM_r}{r^2}$$

$$-\frac{P_\mathrm{c}}{R} \qquad\qquad -\rho_\mathrm{avg}\frac{G(M/2)}{(R/2)^2}$$

$$\Big\downarrow$$

$$\frac{P_\mathrm{c}}{R} \sim \rho_\mathrm{avg}\frac{G(M/2)}{(R/2)^2}$$

$$P_\mathrm{c} \sim \rho_\mathrm{avg}\frac{2GM}{R}$$

$$P_{\mathrm{c},\odot} \sim \rho_\odot \frac{2GM_\odot}{R_\odot} \sim 5 \times 10^{14}\,\mathrm{N\,m^{-2}}$$

CENTRAL TEMPERATURE

- Let's estimate the central temperature, T_c. The temperature of a gas is closely related to pressure, as encoded in the ideal gas law: $P = nkT$, where n is the number of particles per unit volume, k is Boltzmann's constant, and T is the temperature.

- Instead of number density, we will write this in terms of mass density, ρ, because that's what appears in the equation of hydrostatic balance. The relationship is simple: Number density equals mass density divided by the average mass per particle.

$$n = \frac{\rho}{m_{\text{avg}}} \longrightarrow P = \frac{\rho k T}{m_{\text{avg}}}$$

- The ideal gas law tells us that central pressure equals central density, times k, times central temperature, divided by m_{avg}. And we also have our order-of-magnitude expression for central pressure that we obtained from the equation of hydrostatic balance. Let's set them equal.

$$P_{\text{c}} = \frac{\rho_{\text{c}} k T_{\text{c}}}{m_{\text{avg}}} \sim \rho_{\text{avg}} \frac{2GM}{R}$$

- That leaves us with an equation where on one side we have the central density and on the other we have the average density. Those aren't the same—so we make an educated guess. The central density is higher than average, perhaps twice as high. When we replace ρ_c with $2\rho_{\text{avg}}$, the ρ's cancel out and we can solve for T_c.

$$\rho_{\text{c}} \stackrel{?}{=} 2\rho_{\text{avg}} \longrightarrow \frac{2\rho_{\text{avg}} k T_{\text{c}}}{m_{\text{avg}}} \sim \rho_{\text{avg}} \frac{2GM}{R}$$

$$T_{\text{c}} \sim \frac{GM m_{\text{avg}}}{kR}$$

- It's time to plug in some numbers for the Sun. The only thing we don't already know is the average mass per particle. For simplicity, let's say that the Sun is made of pure hydrogen and it's hot enough to be ionized, so there are equal numbers of protons and electrons. Both particles contribute to the number density, but the electrons have essentially 0 mass compared to the protons, so the average particle mass is half a proton mass. Plugging that in, along with the mass and radius of the Sun, we get $T = 1.2 \times 10^7$ Kelvin.

$$m_{\mathrm{avg}} \approx 0.5\, m_{\mathrm{p}} \longrightarrow T_{\mathrm{c},\odot} \sim \frac{GM_\odot(m_{\mathrm{p}}/2)}{kR_\odot} \sim 1.2 \times 10^7\,\mathrm{K}$$

- That's the same order of magnitude that is sufficient to ignite hydrogen fusion.

PROPERTIES OF MAIN-SEQUENCE STARS

- The formula for central temperature is true regardless of the star's energy source.

- For thermonuclear fusion, the reaction rates increase rapidly with temperature. For each reaction, there's a sharply defined ignition temperature. An implication is that we should expect all hydrogen-fusing stars to have the same central temperature, regardless of the mass or size of the star.

- Once a star achieves hydrostatic balance, the central temperature always hovers just above the ignition temperature. That's why, for stars fusing hydrogen in their cores, T_c is a constant, independent of M or R. Our equation now says that M/R is proportional to a constant, or, equivalently, R is proportional to M.

$$T_{\mathrm{c}} \sim \frac{GMm_{\mathrm{avg}}}{kR}$$

$$T_{\mathrm{c}} \sim 10^7\,\mathrm{K} \sim \mathrm{const.} \longrightarrow \mathrm{const.} \sim \frac{GMm_{\mathrm{avg}}}{kR}$$

$$R \propto M$$

- This is exactly what we observed in our chart of radius versus mass for main-sequence stars! Now we understand the reason for that pattern: Main-sequence stars are all fusing hydrogen at their centers.

- What about the observed pattern between luminosity and mass? The data tell us that L goes as M^3. How can we understand that?

$$L \propto M^3$$

- If we make the usual approximation that stars radiate nearly as blackbodies, then $L = 4\pi R^2 \sigma T_{\text{eff}}^4$.

- How does that depend on mass? We know that R is proportional to M. But what about T_{eff}?

- It's important to remember that T_{eff} is the temperature of a star's outer layers—the photosphere, not the center. On the main sequence, the central temperatures are all the same—about 10^7 Kelvin—but the effective temperatures are not. They range from 3000 to 30,000 Kelvin. To understand the relationship between central and effective temperatures, we need to think about how the energy flows from the center to the surface.

- Fusion reactions produce very energetic photons: x-rays and gamma rays. But they don't make it far, because the Sun is opaque. It's optically thick. This is because the interior is dense and because it's ionized.

- Charged particles, especially electrons, interact strongly with photons. So, the photons are nearly trapped; they transfer their energy to the surrounding gas, heating it up. Photons are constantly getting scattered, absorbed, and reradiated with different energies.

- But the energy doesn't stay in the core forever. That's because the temperature can't be constant throughout the whole star. The core is superhot, and above the photosphere is the cold vacuum of space, so temperature must be a function of r. And heat will naturally flow from high to low temperature—from the core to the photosphere.

- The rate of heat flow depends on the temperature gradient. The larger the difference in temperature between the core and photosphere, the more rapidly heat will flow. The temperature gradient rises until the rate of heat flowing upward and escaping is equal to the rate at which energy is being released by fusion. At that point, the star achieves a steady state and $T(r)$ stops changing.

- To be more quantitative, we'll model the heat flow as a diffusion process. We'll assume that heat is being transferred in the form of photons, but the photons can't travel far; they're like pinballs, bumping around from particle to particle.

- Each time the ball hits a "bumper," it collides and flies off with the same speed in a random direction until it hits the next one. After the first collision, it follows a straight path represented by the vector $\vec{r}_1$. After the second collision, it advances by r_2, and so on. After the N^{th} collision, the location of the ball, $\vec{r}$, is the sum of $\vec{r}_1$, $\vec{r}_2$, and so forth, up to $\vec{r}_N$.

$$\vec{r}_1 = \vec{r}_1 + \vec{r}_2 + \cdots + \vec{r}_N$$

- Now let's imagine playing a zillion pinball games. What is the average distance the ball goes after N collisions? To calculate that, let's square the equation for r—that is, let's take the dot product of $\vec{r}$ with itself. Then, we expand that product.

$$= (\vec{r}_1 + \vec{r}_2 + \cdots + \vec{r}_N) \cdot (\vec{r}_1 + \vec{r}_2 + \cdots + \vec{r}_N)$$

$$= \boxed{\begin{array}{c} r_1^2 + r_2^2 + \cdots + r_N^2 \\ \text{each has the} \\ \text{same average, } \ell^2 \end{array}} \begin{array}{c} + \\ + \\ \end{array} \boxed{\begin{array}{l} 2\,(\vec{r}_1 \cdot \vec{r}_2 + \vec{r}_1 \cdot \vec{r}_3 + \cdots) \\ 2\,(\vec{r}_2 \cdot \vec{r}_3 + \vec{r}_2 \cdot \vec{r}_4 + \cdots) + \cdots \\ \qquad\qquad \text{average to zero} \end{array}}$$

- Interestingly, all the dot products average to 0. That's because the vectors have random directions. Any 2 vectors are just as likely to point the same way as opposite ways, so each dot product is just as likely to be positive as negative, and the average is 0.

- The r^2's don't average to 0, but they all have the same average value because the "bumpers" have random locations. The average of each r^2 term will be the square of ℓ, the mean free path. Because there are N terms, the sum is $N\ell^2$.

$$r_{\text{avg}}^2 = \ell^2 + \ell^2 + \cdots + \ell^2 = N\ell^2$$

- Taking the square root, we find that the distance from the origin—the so-called root-mean-square distance—is equal to the square root of N times ℓ. In other words, the distance grows as the square root of the number of collisions.

- Let's rewrite that in terms of elapsed time, t. The number of collisions, N, equals t divided by the mean time between collisions—which is equal to the mean free path divided by the speed of the pinball (c, because our pinball is standing in for a photon).

$$\sqrt{r_{\text{avg}}^2} = r_{\text{rms}} = \sqrt{N}\,\ell = \sqrt{\frac{t}{\ell/c}}\,\ell = \sqrt{\ell c t}$$

$$N = \frac{t}{\ell/c}$$

- The distance grows as the square root of ℓct. Contrast this to the case of a ball rolling in a straight line without any "bumpers"; the distance would be proportional to t, not the square root of t. The dependence on the square root of time is a distinctive feature of diffusion.

- If the Sun were optically thin—a transparent sphere of radius R—the photons could fly free and unimpeded from the core to the surface. The time to reach the photosphere (t_{free})—would equal R/c.

- For the Sun, that comes out to be 2.3 seconds.

$$t_{\text{free}} = \frac{R}{c} \qquad t_{\text{free},\odot} = \frac{R_\odot}{c} = 2.3\,\text{sec}$$

- But real stars are optically thick, and the photons have to play pinball. To see how long that takes, let's take our diffusion formula and set the root-mean-square distance equal to the stellar radius. Solving for t gives a diffusion time of $R^2/\ell c$.

$$r_{\text{rms}} = \sqrt{\ell c t}$$
$$R = \sqrt{\ell c\, t_{\text{diff}}}$$
$$t_{\text{diff}} = \frac{R^2}{\ell c}$$

- To evaluate this for the Sun, we need to know ℓ, the mean free path. That's not something we're equipped to calculate from first principles; it's a job for a quantum mechanic with a good computer. The outcome of those calculations is that the Sun's mean free path averaged over its whole interior is about 1 millimeter.

- If we insert the Sun's radius for R, 1 millimeter for ℓ, and the speed of light for c, we get a diffusion time of 50,000 years.

$$\ell_\odot \sim 1\,\text{mm}$$
$$t_{\text{diff},\odot} \sim \frac{R_\odot^2}{\ell_\odot c} \sim 50{,}000\,\text{years}$$

- A star's opacity dramatically slows the progress of photons and spreads out their arrival times relative to the case of 0 opacity. The ratio of the diffusion time to the free-flying time is R/ℓ. For the Sun, that ratio is of order 1 trillion.

$$\frac{t_{\text{diff}}}{t_{\text{free}}} = \frac{R^2/\ell c}{R/c} = \frac{R}{\ell} \qquad \frac{R_\odot}{\ell_\odot} \sim 10^{12}$$

- Now we can understand the connection between the central and effective temperatures.

$$T_{\text{eff}} = T_{\text{c}} \left(\frac{\ell}{R}\right)^{1/4}$$

- For the Sun, that's the fourth root of 10^{-12}, which is 10^{-3}.

The Sun's core is of order 10^7 degrees because that's when nuclear fusion becomes possible, and the photosphere is 1000 times cooler because the Sun's radius is a trillion times larger than the mean free path.

- We also have all the tools we need to understand the mass-luminosity relation.

$$L \propto M^3$$

- This is what we observed to be true for main-sequence stars.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 10.

Choudhuri, *Astrophysics for Physicists*, chaps. 3 and 4.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 5.

Maoz, *Astrophysics in a Nutshell*, chap. 3.

Phillips, *The Physics of Stars*.

Ryden and Peterson, *Foundations of Astrophysics*, chaps. 14–18.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 17

WHITE DWARFS

Unlike ordinary stars, white dwarfs—planet-sized objects with the mass of the Sun—do not fuse hydrogen. In fact, they don't fuse anything; there's no internal power source. They're held up against the force of gravity by Heisenberg's uncertainty principle.

DEGENERACY PRESSURE

- For ordinary stars, what prevents gravitational collapse is gas pressure. The higher temperature and density toward the center of a star lead to increasing pressure with depth. The resulting outward force opposes the inward pull of gravity, allowing the star to achieve hydrostatic balance. That balance can be sustained because the nuclear reactions keep the center hot; they replenish the energy that diffuses outward and radiates away into space.

- When a star runs out of nuclear fuel, the radiated energy isn't being replenished anymore. The pressure drops. Hydrostatic balance is lost. The star contracts.

- When the core is compressed to a high enough density, quantum theory becomes relevant. The star comes up against the Pauli exclusion principle: No 2 electrons can occupy the same exact quantum state—same position, energy, and angular momentum.

The Pauli exclusion principle is the underlying reason for the most basic fact of chemistry: When you add an electron to an atom, you can't put it in the same state as the other electrons that are already there. Instead, adding electrons leads to wider orbits and more complex atoms, leading to the rules and patterns governing the periodic table of the elements.

- In a white dwarf, the electrons are squeezed so tightly that the Pauli exclusion principle prevents them from occupying the same location in space. That creates a pressure—unrelated to temperature—that pushes them apart. It's this quantum pressure that keeps gravity at bay.

- Let's calculate the order of magnitude of the effect. The Heisenberg uncertainty principle says that for any particle, the minimum-possible product between the uncertainties in momentum and position is of order $\hbar$, which is Planck's constant (h) divided by 2π.

$$(\Delta p\,\Delta x)_{\mathrm{min}} \sim \hbar$$

- Imagine compressing a cold gas to a density of n particles per cubic centimeter. Each particle is confined in a small cube with volume $1/n$. The uncertainty in each coordinate of the particle's location, Δx, has been squeezed down to 1 divided by the cube root of n.

$$n\,\frac{\text{particles}}{\text{cm}^3} \quad\longrightarrow\quad \Delta V = \frac{1}{n}\,\frac{\text{cm}^3}{\text{particle}}$$

$$\Delta x = \frac{1}{n^{1/3}}$$

- Heisenberg taught us that the momentum of each particle will not remain 0; it will inevitably range up to values of $\hbar/\Delta x$, or $\hbar$ times the cube root of n.

$$\Delta p \sim \frac{\hbar}{\Delta x} \sim \hbar n^{1/3}$$

- And because momentum is mass times velocity, the velocity will range up to

$$\Delta v \sim \frac{\hbar}{m} n^{1/3},$$

where m is the mass of each particle. So, at high density, even if the gas is cold and there's no energy source, the particles will be moving. That creates a pressure that can resist further compression.

- To find out how much pressure, we'll use a result from lecture 8, when we derived the ideal gas law (which is $P = nkT$): The pressure of a gas is nm times the average value of $v2$. The same logic applies here. To order of magnitude, we can replace the v in the pressure equation with the Δv that arises from Heisenberg's principle, because the particles have speeds that range from 0 to several times Δv.

$$\Delta v \sim \frac{\hbar}{m} n^{1/3}$$

$$P = nm\langle v^2 \rangle$$

$$P_{\text{deg}} \sim nm \left(\frac{\hbar}{m} n^{1/3} \right)^2$$

$$P_{\text{deg}} \sim \frac{\hbar^2}{m} n^{5/3}$$

- This is called degeneracy pressure.

There are 3 things to notice about the formula for degeneracy pressure.

1. It's independent of temperature. That's why degeneracy pressure can support a white dwarf even when there is no internal energy source to keep the gas hot.
2. Degeneracy pressure varies inversely with particle mass. That means in a neutral gas of electrons and ions, it's the electrons—the low-mass particles—that produce almost all the degeneracy pressure.
3. Degeneracy pressure varies as the 5/3 power of density. That's a stronger dependence than ideal gas pressure, which varies as density to the first power. So, if we crank up the density, eventually degeneracy pressure will dominate over gas pressure.

BUILDING A STAR USING DEGENERACY PRESSURE

- In the previous lecture, we constructed models for main-sequence stars by setting the outward force from the pressure gradient equal to the inward force of gravity. Let's do the same thing now, but using degeneracy pressure instead of ideal gas pressure.

- We need to satisfy the equation of hydrostatic balance:

$$\frac{dP}{dr} = -\rho(r)\,\frac{GM_r}{r^2}.$$

- Let's replace this differential equation with the algebraic approximation we derived in the previous lecture: the central pressure is of order the average density times $2GM/R$.

$$P_c \sim \rho_{\text{avg}} \frac{2GM}{R}$$

- But this time, instead of using the ideal gas law, let's set P_c equal to the degeneracy pressure.

$$\frac{\hbar^2}{m} n^{5/3} \sim \rho_{\mathrm{avg}} \frac{2GM}{R}$$

- Because degeneracy pressure varies inversely with mass and electrons are 1800 times less massive than protons, the electrons provide almost all the degeneracy pressure. So, for m, we'll insert the electron mass, and for n, we'll insert the electron density.

$$\frac{\hbar^2}{m_{\mathrm{e}}} n_{\mathrm{e}}^{5/3} \sim \rho_{\mathrm{avg}} \frac{2GM}{R}$$

- In addition, the mass density, ρ, is dominated by the ions. While the electrons provide most of the pressure, the ions provide most of the mass. That makes the relationship between n and ρ subtler than before.

- Let's suppose our white dwarf is made of carbon and oxygen. Carbon has 6 electrons, 6 protons, and 6 neutrons, for an atomic mass of 12. Oxygen has 8 electrons and an atomic mass of 16. In both cases, there are twice as many nucleons as electrons.

- That means for every 1 electron, there are 2 nucleons. If the electron number density is n, the mass density is $2m_{\mathrm{p}}n_{\mathrm{e}}$, where m_{p} is the proton mass. We can use that relation to rewrite n_{e} in terms of ρ.

$$\rho = 2m_{\mathrm{p}}n_{\mathrm{e}}$$

- Another subtlety is that the left side of the equation refers to the central density, but the right side refers to the average density. If we make the crude assumption that the central density is twice the average density, that will cancel the 2 on the right side of the equation.

$$\frac{\hbar^2}{m_{\mathrm{e}}} \left(\frac{\rho}{2m_{\mathrm{p}}}\right)^{5/3} \sim \rho \frac{GM}{R}$$

- Next, let's bring all the constants over to the left side of the equation and the stellar properties to the right side. The right side becomes

$$\frac{\hbar^2}{Gm_\mathrm{e}(2m_\mathrm{p})^{5/3}} \sim \frac{M}{R\rho^{2/3}}.$$

- M, R, and ρ are not all independent; ρ is M divided by $4/3\pi R^3$. For simplicity, let's just say it's M/R^3.

$$\frac{M}{R}\left(\frac{R^3}{M}\right)^{2/3}$$

- Then, the right side simplifies to

$$R\,M^{1/3}.$$

- We can solve for R, leading to an expression for radius as a function of mass. It looks a little neater if we express the stellar mass in units of the proton mass.

$$R \sim \frac{\hbar^2}{Gm_\mathrm{e}(2m_\mathrm{p})^{5/3}M^{1/3}}$$

$$R \sim \frac{1}{2^{5/3}}\frac{\hbar^2}{Gm_\mathrm{e}m_\mathrm{p}^2}\left(\frac{M}{m_\mathrm{p}}\right)^{-1/3} \quad \text{(approximate)}$$

- There's something strange about this result: R is inversely proportional to the cube root of M. In other words, increase the mass and the radius decreases.

- Although counterintuitive, this is the case because when you add mass, you increase the gravitational compression. For degeneracy pressure to compensate, the particle velocities need to increase. And from Heisenberg's principle, the way to force Δv to increase is to make Δx decrease—shrink the spacing between particles. So, you end up with a smaller, denser star.

- Let's run the numbers for an object with the mass of the Sun. But because we derived this formula with blatant disregard for factors of 2 and π, the exact answer turns out to be bigger by about a factor of 3. When you make that correction and plug in all the numerical constants, you get an R of about 5800 kilometers. And studies of large samples of white dwarfs confirm the amazing truth that the bigger the mass, the smaller the size.

$$R = 5800\,\mathrm{km}\left(\frac{M}{M_\odot}\right)^{-1/3} \quad \text{(more accurate)}$$

THE ORIGIN OF WHITE DWARFS

- A star like the Sun has enough hydrogen to last about 10 billion years. At that point, the core has been converted entirely to helium. Fusion stops, the gas pressure starts to decrease, and the core contracts.

- This causes the density to rise and, less obviously, causes the core to heat up. Gravitational contraction releases gravitational potential energy, which is converted into heat. Eventually, the core gets dense and hot enough to ignite helium fusion.

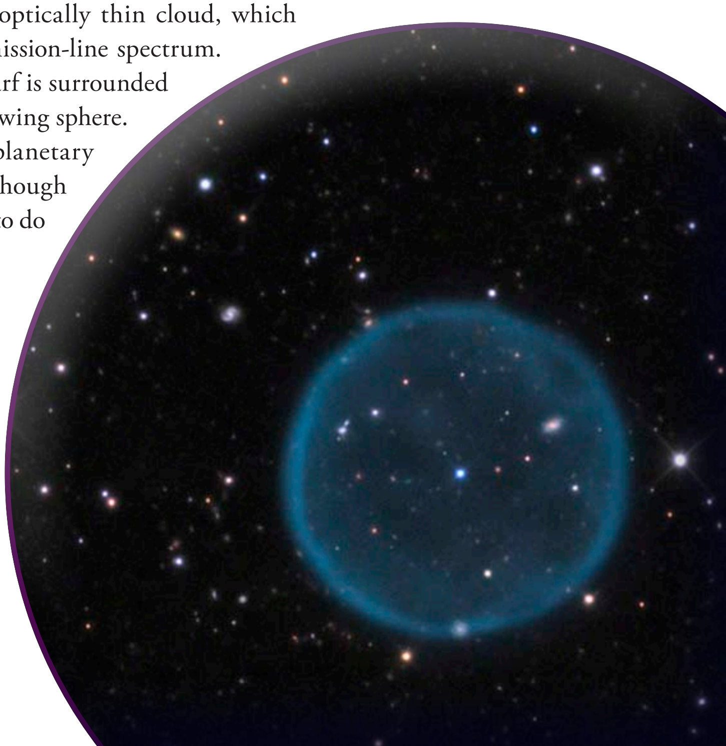
This is a remarkable and seemingly paradoxical property of all stars: When they lose energy, their cores heat up. (Usually, adding energy heats something up.) The jargon is that gravitationally bound systems, like stars, have a negative heat capacity.

- At a temperature of around 10^8 Kelvin, helium nuclei have enough kinetic energy to collide and make beryllium-8—which is a very unstable nucleus, but if there are enough helium nuclei around, there's a chance that one of them will hit the beryllium nucleus hard enough to make carbon-12. So, carbon starts accumulating and sometimes fuses with helium to make oxygen-16.

- Meanwhile, strange things are happening higher up in the star. The outer part of the star swells up to 100 times its usual size—the star becomes a giant. Eventually, the outer layers rise so high and become so weakly bound to the star that they get pushed out into space by radiation pressure. The hot, dense core becomes fully exposed to the universe after billions of years of being hidden away.

- By this point, the core is so dense that degeneracy pressure is sufficient to oppose gravity, so it stops contracting and there's no way to ignite any further nuclear reactions. Once the helium runs out, the core becomes an inert ball of carbon and oxygen. That's a white dwarf. It starts off white-hot but gradually cools and fades to black over billions of years. This is the fate that awaits the Sun.

- While the white dwarf is still hot, all those ultraviolet and x-ray photons stream out and light up the surrounding gas. What used to be the star's outer layers now form an optically thin cloud, which displays an emission-line spectrum. The white dwarf is surrounded by a wispy, glowing sphere. It's called a planetary nebula (even though it has nothing to do with planets).



WHITE DWARF STRUCTURE

- There was a flaw in our earlier calculation regarding the theory of white dwarf structure. An assumption was made that's not always true.

- We started by saying that when Heisenberg's principle starts to become relevant, the uncertainty in momentum is $\hbar$ times the cube root of the number density. And because Δp equals $m\Delta v$, then Δv also varies as $n^{1/3}$.

$$\Delta p \sim \hbar n_{\mathrm{e}}^{1/3}$$

$$\Delta v \sim \frac{\hbar}{m_{\mathrm{e}}} n_{\mathrm{e}}^{1/3}$$

- This equation implies that if we keep increasing the density, we can make Δv as high as we want. But that can't be true. The universe has a maximum speed: the speed of light. So, as Δv approaches c, our formula must break down.

- Let's figure out when this problem occurs. We'll set Δv equal to, let's say, half the speed of light and solve for the critical value of n.

$$\frac{c}{2} \sim \frac{\hbar}{m_{\mathrm{e}}} n_{\mathrm{crit}}^{1/3}$$

$$n_{\mathrm{crit}} \sim \left(\frac{m_{\mathrm{e}} c}{2\hbar}\right)^3$$

- To get the corresponding mass density, we multiply by 2 proton masses, which numerically comes out to be 7.3 million grams per cubic centimeter. Above that density, we need to take relativity into account.

$$\rho_{\mathrm{crit}} \sim 2m_{\mathrm{p}} \left(\frac{m_{\mathrm{e}} c}{2\hbar}\right)^3 \sim 7.3 \times 10^6 \,\mathrm{g\,cm^{-3}}$$

- Because density increases with mass, there must be some critical mass for a white dwarf above which the electrons start moving close to the speed of light.

- To calculate that critical mass, we'll start with the mass-radius relationship we previously derived.

$$R \propto M^{-1/3}$$

- This implies that density is proportional to M/R^3. And for a white dwarf, R is proportional to $M^{-1/3}$. Together these imply that density goes as mass squared.

$$\rho \propto \frac{M}{R^3} \propto \frac{M}{(M^{-1/3})^3} \propto \frac{M}{M^{-1}} \propto M^2$$

- The density of a 1-solar-mass white dwarf is about 3 times lower than the critical density. To raise the density by a factor of 3, we need to increase the mass by a factor of 3, or 1.7. We expect the critical mass to be around 1.7 solar masses, which is not that high. It seems realistic that the core of a star could exceed 1.7 solar masses. To repair our calculation, we need a new formula for degeneracy pressure for the case in which the particles are moving close to the speed of light.

- When we derived the ideal gas law in lecture 8, we started by imagining a gas in which particles move in only 1 direction. Let's do the same thing, but this time all the particles move with same speed, c. What pressure do they exert on the wall? When a particle hits the wall, it reflects, changing its momentum from p to $-p$. Because momentum is conserved, the wall absorbs a momentum of $2p$.

- Now consider the momentum absorbed by an area ΔA of the wall. In a time Δt, it absorbs a momentum of

$$\text{momentum absorbed by wall} = (2p)\left(\frac{n}{2}\right)(c\,\Delta t\,\Delta A).$$

momentum transferred per particle

volume within $c\Delta t$ of wall

density of right-moving particles

- The first factor, $2p$, is the momentum transferred per collision. The second factor, $n/2$, is the number density of rightward-moving particles. And the third factor is the volume of space within which a particle is close enough to hit the wall within a time Δt.

- Pressure is force per unit area, or momentum per unit time and area, so we divide by $\Delta t \Delta A$, giving pnc, where p is the momentum of each particle.

$$\text{pressure } P = pnc$$

- For degenerate particles, we know from Heisenberg's principle that Δp is of order $\hbar$ times the cube root of n. Inserting that for p, we find that the degeneracy pressure is of order $\hbar c n^{4/3}$. That's the formula we need for particles moving near the speed of light.

$$\Delta p \sim \hbar n^{1/3} \;\rightarrow\; P_{\text{deg}} \sim \hbar c n^{4/3}$$

- The nonrelativistic formula went as $n^{5/3}$. This new one goes as $n^{4/3}$; the pressure doesn't increase as rapidly with density. This will be turn out to be bad news for a massive white dwarf.

- When you pile more mass onto a white dwarf, increasing the gravitational force, the only way it can reestablish hydrostatic balance is to become denser. That way, the particles are pressed closer together and Heisenberg's principle leads to higher velocities and higher degeneracy pressure.

- But above the critical mass, the particle speeds can't increase any more; they're already at the speed of light. The degeneracy pressure starts rising more slowly with density. The white dwarf has a harder time fending off gravitational collapse.

- Let's see if we can still achieve hydrostatic balance. We'll take our equation expressing the condition of hydrostatic balance,

$$P_c \sim \rho_{\text{avg}} \frac{2GM}{R},$$

and insert the new formula for degeneracy pressure.

$$P_{\text{deg}} \sim \hbar c n_{\text{e}}^{4/3}$$

- We'll aim straight for the scaling relation between M and R, ignoring all the 2s and πs and neglecting the difference between ρ_c and ρ_{avg}. As before, we'll replace the electron number density with $\rho/2m_p$, but we won't bother with the 2, and then we'll bring all the physical constants over to one side and the stellar properties to the other. And because we're interested in the mass-radius relation, we'll replace ρ with M/R^3.

$$\hbar\, c\, n_e^{4/3} \sim \rho \frac{GM}{R}$$

$$\hbar\, c \left(\frac{\rho}{m_p}\right)^{4/3} \sim \rho \frac{GM}{R}$$

$$\frac{\hbar c}{G m_p^{4/3}} \sim \frac{M}{R\rho^{1/3}} \sim \frac{M}{R}\left(\frac{R^3}{M}\right)^{1/3} \sim M^{2/3}$$

- The Rs cancel out. There's no radius anymore in the equation—just mass. Apparently, there's only 1 possible mass for which we can achieve hydrostatic equilibrium.

- Solving for M leads to an elegant collection of fundamental constants. Plugging in the values of those constants, the number turns out to be 2.2×10^{57}, and the mass is 1.8 solar masses.

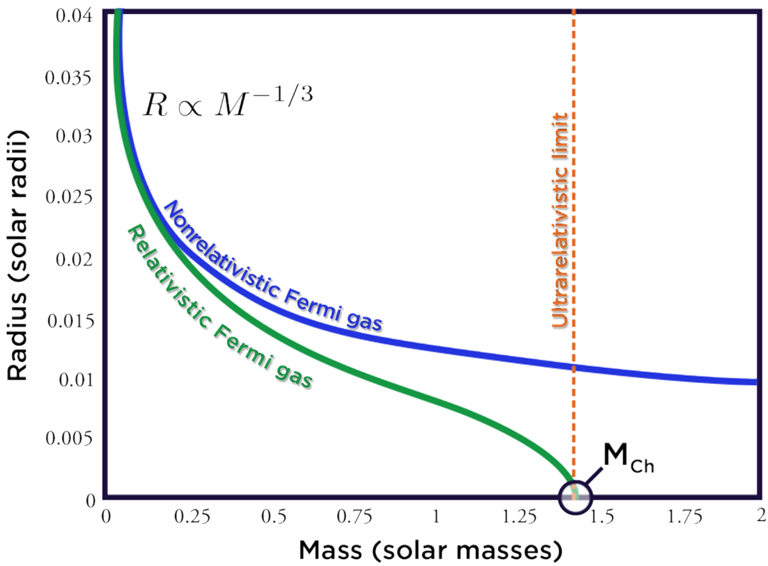
$$M \sim \left(\frac{\hbar c}{G m_p^{4/3}}\right)^{3/2} \sim \underbrace{\left(\frac{\hbar c}{G m_p^2}\right)^{3/2}}_{2.2 \times 10^{57}} m_p \sim 1.8\, M_\odot$$

- A more rigorous calculation that keeps track of all the 2s and πs gives a somewhat smaller answer of 1.4 solar masses. This glorious combination of fundamental constants and the critical mass of 1.4 is called the Chandrasekhar limit, named after astrophysicist Subrahmanyan Chandrasekhar.

$$M_{Ch} = 1.4\, M_\odot$$

- The Chandrasekhar mass means that relativistic degeneracy pressure can only balance gravity for that 1 specific mass.

- On this chart of radius versus mass, the curves are exact solutions of the equation of hydrostatic equilibrium for a cold degenerate gas. One curve shows the relationship we derived for low-mass white dwarfs: R goes as $M^{-1/3}$. But this curve ignores the effects of relativity. When we include relativistic effects in the theory, we get the other curve.



- At low masses, the 2 curves are similar, but as the mass approaches the Chandrasekhar limit, the relativistic curve starts dropping. The radius is shrinking; gravity is winning. And at 1.4 solar masses, the radius drops to 0!

What actually happens if the core of a massive star exceeds 1.4 solar masses?

Here's a quote from the end of Chandrasekhar's most famous paper:

> It must be taken as well established that the life-history of a star of small mass must be essentially different from the life-history of a star of large mass. For a star of small mass, the natural white-dwarf stage is an initial step towards complete extinction. A star of large mass cannot pass into the white-dwarf stage, and one is left speculating on other possibilities.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 16.

Choudhuri, *Astrophysics for Physicists*, chaps. 3 and 4.

Maoz, *Astrophysics in a Nutshell*, chap. 4.

Ryden and Peterson, *Foundations of Astrophysics*, chaps. 14–18.

Shapiro and Teukolsky, *Black Holes, White Dwarfs, and Neutron Stars*.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 18

WHEN STARS GROW OLD

Dark nebulas—clouds of hydrogen and helium, along with a small percentage of heavier elements—are the raw material for stars. Dark clouds are stable for millions or billions of years, but occasionally something causes part of a cloud to get a little too dense—maybe from a collision with a different cloud or a nearby supernova explosion. When that happens, the increased gravity of the dense region attracts the surrounding gas, amplifying the density still further. The cloud begins collapsing. It fragments into clumps of different sizes, which eventually become stars with different masses.

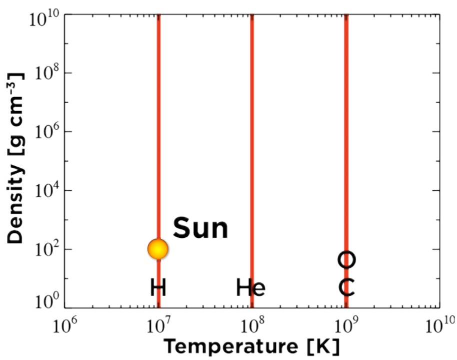
THE CORE OF A STAR

- When examining the core of a star, the 2 properties we need to keep track of are its temperature and density. To do so, let's make a logarithmic chart with central density plotted against central temperature.

- The Sun's core has a density of around 100 grams per cubic centimeter and a temperature of 10^7 Kelvin. That's the ignition temperature for hydrogen fusion. We expect all main-sequence stars to have that same core temperature because they're all fusing hydrogen into helium. On the chart, the hydrogen-fusion line is a vertical line going through the Sun. Any main-sequence star will have core properties falling somewhere on the line.

- The ignition temperature of helium is higher, about 10^8 Kelvin. That's because helium has a higher electric charge than hydrogen, so it takes more energy to overcome the electrical repulsion and get the nuclei to fuse into carbon and oxygen. On the chart, any star that's fusing helium in its core will appear close to a vertical line at 10^8 Kelvin.

- Likewise, fusing carbon and oxygen to make heavier elements requires an even hotter temperature, around 10^9 Kelvin. In general, the heavier the element, the hotter the ignition temperature. (We could keep drawing vertical lines at higher temperatures for fusing even heavier elements.)



- With these landmarks in place, let's consider the fate of the Sun. It's happily fusing hydrogen into helium at 10^7 Kelvin and will stay at that point while the hydrogen lasts. But what happens when the hydrogen runs out?

- Without fusion to replenish the energy that the star is radiating away, the core begins contracting. Its density increases. Contraction also liberates gravitational potential energy, which is converted to heat, raising the temperature. (Recall that when gravitationally bound systems lose energy, they heat up.)

- As the temperature and density rise, the point on the chart that represents the Sun moves up and to the right. We can figure out the exact trajectory of that point by using the fact that the star is still very nearly in hydrostatic equilibrium.

- The reason for this is that the energy loss is so slow that at any instant, the star is extremely close to equilibrium. The pressure may be changing over millions of years, but the star's mass distribution adjusts itself on a timescale of hours to strike a temporary balance between gravity and pressure. This kind of situation is called a quasi-equilibrium.

- As a result, we can calculate a star's path on the chart by using the equation of hydrostatic balance. In a previous lecture, we got the result that M/R is proportional to central temperature by setting the pressure required for hydrostatic balance equal to ideal gas pressure. We used this equation to show that for stars on the main sequence, which all have the same T_c, R is proportional to M.

$$\frac{M}{R} \propto T_{\mathrm{c}}$$

- Now we're talking about stars that aren't fusing hydrogen anymore; there's no reason for the central temperature to be a constant. But as long as we're near equilibrium, it will still be true that that M/R is proportional to T_c.

- Let's write that in terms of central density instead of radius. We'll start by cubing the equation and writing the left side as $M^2(M/R^3)$—because that way, we recognize that second factor as being proportional to density.

$$\frac{M^3}{R^3} \propto T_{\mathrm{c}}^3$$

$$M^2\left(\frac{M}{R^3}\right) \propto T_{\mathrm{c}}^3$$

- Technically, M/R^3 is proportional to the average density, not necessarily the central density, but it's a decent approximation to assume that the central density is always higher than average by the same factor, in which case we can rewrite the equation as

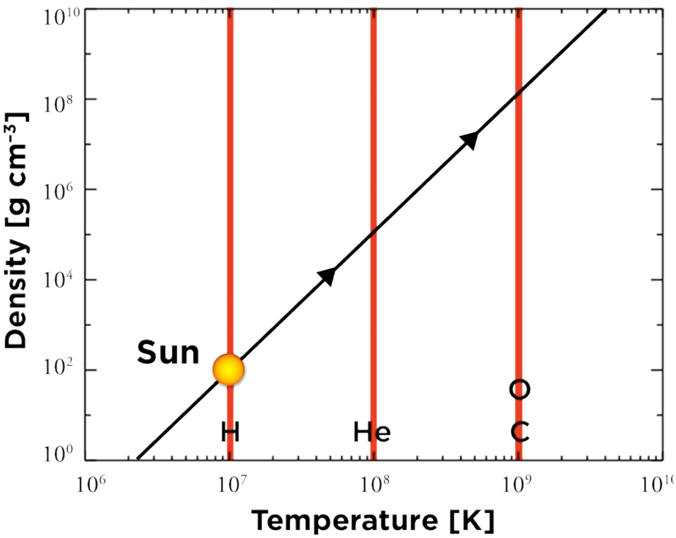
$$M^2 \rho_c \propto T_c^3.$$

- Solving for ρ and taking the log, we get a straight line with a slope of 3.

$$\rho_c \propto T_c^3 M^{-2}$$

$$\log \rho_c = 3 \log T_c - 2 \log M + \text{const.}$$

- That's the line the core of the Sun will follow once there's no hydrogen left—toward a higher density and a hotter temperature.



STELLAR EVOLUTION AT THE CORE

- When a star condenses from a glob of galactic gas, it starts out with low density and low temperature. Gravity draws it together. The core contracts and heats for millions of years, until the center reaches 10^7 degrees. Then, the hydrogen ignites. The energy erupting from fusion in the core comes into balance with the energy being lost up at the surface. The star joins the main sequence. That's where the Sun is now, and where it will remain for another 5 billion years.

- Then, the core runs out of hydrogen; it's nearly pure helium. Fusion fizzles out. So, the core starts contracting again, proceeding farther up a diagonal line with a slope of 3.

- The temperature rises until it hits 10^8 degrees. That's hot enough to ignite the helium. This new power source halts the contraction and allows the star to linger a while—maybe for tens of millions of years.

- The helium-fusing phase doesn't last as long as the hydrogen-fusing phase because helium fusion isn't as efficient, so there's isn't as much energy to be liberated by fusing helium—only a tenth as much energy per unit mass as from fusing hydrogen.

- Next, the helium gets used up. The core becomes pure carbon and oxygen. Fusion fizzles out. Gravitational contraction resumes. The star moves farther up the diagonal line.

- If it reaches a billion degrees, the carbon and oxygen ignite, again pressing the pause button on the collapse. The core fuses heavier and heavier nuclei, each one lasting for less and less time, burning with increasing desperation.

- But there's something important missing from the log chart, something that will alter the fate of the Sun: degeneracy pressure.

- The equation governing the Sun's trajectory—that diagonal line with a slope of 3—came from assuming that the star is supported by ideal gas pressure. As the density keeps rising, though, at some point degeneracy pressure becomes more important than gas pressure.

- When does that happen? We can find out by setting the equation for degeneracy pressure equal to the equation for ideal gas pressure. To keep things simple, we won't keep track of all the constants; we'll just keep track of how ρ depends on T.

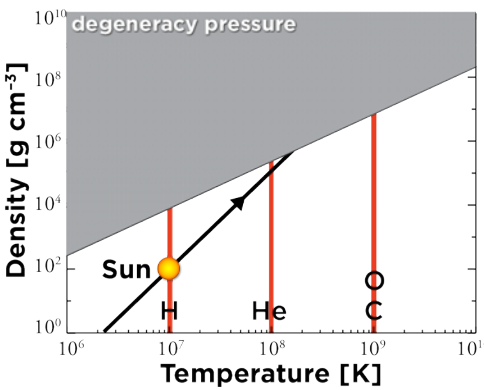
$$P_{\text{deg}} \sim \frac{\hbar^2}{m_{\text{e}}} \left(\frac{\rho}{2m_{\text{p}}} \right)^{5/3} \qquad P_{\text{gas}} = \frac{\rho k T}{m_{\text{avg}}}$$

$$\rho^{5/3} \propto \rho T$$

$$\rho^{2/3} \propto T$$

$$\rho \propto T^{3/2}$$

- After some algebra, we see that on the boundary between gas pressure and degeneracy pressure, ρ varies as $T^{3/2}$, which corresponds to a straight line with a slope of 3/2 on the log chart. To find the *y*-intercept, we'd need to keep track of all the constants; if you do that, you find that the boundary line runs from about 10^2 to 10^8 grams per cubic centimeter as the temperature runs from 10^6 to 10^{10}. Above that boundary, degeneracy pressure dominates.



- However, in the previous lecture, we learned that degeneracy pressure doesn't always go as $\rho^{5/3}$. When the density approaches the critical density—several million grams per cubic centimeter—the pressure starts going as $\rho^{4/3}$. The change in behavior is because the electrons are squeezed so tight that they're moving close to the speed of light.

- So, we need to modify the boundary line. Above the critical density, the slope of the boundary is defined by setting $\rho^{4/3}$ to be proportional to ρT, the ideal gas pressure. In that case, when we solve for ρ, we get that it goes as T^3.

$$P_{\text{deg}} \propto \rho^{4/3} \qquad P_{\text{gas}} = \frac{\rho k T}{m_{\text{avg}}}$$

$$\rho^{4/3} \propto \rho T$$

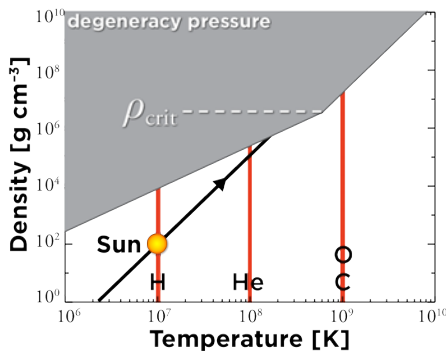
$$\rho^{1/3} \propto T$$

$$\rho \propto T^3$$

$$\log \rho = 3 \log T + \text{const.}$$

- The slope is twice as high as before. The effect is to make a kink in the boundary line at high density.

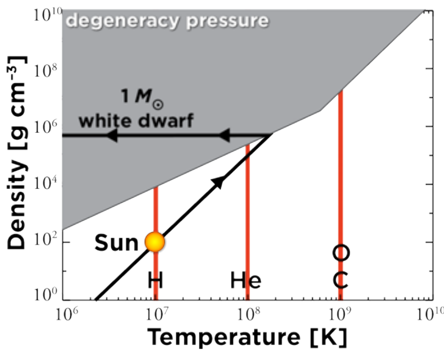
- Now that we've established the domain of degeneracy pressure, let's restart our solar-mass star at low density and temperature. It starts contracting and works its way up the line, pausing to fuse hydrogen and then contracting until the helium ignites.



- But now we see that after the helium runs out, it doesn't advance to carbon ignition. Contraction brings it into the degeneracy zone, and degeneracy pressure halts the contraction. Because degeneracy pressure doesn't depend on temperature or the energy content of the gas, the core can achieve a stable and eternal balance between pressure and gravity—no fusion required. The star becomes a white dwarf.

- The star is still losing energy through radiation, so the temperature decreases, meaning the point on the chart moves straight leftward. The white dwarf cools off and goes dark over billions of years.

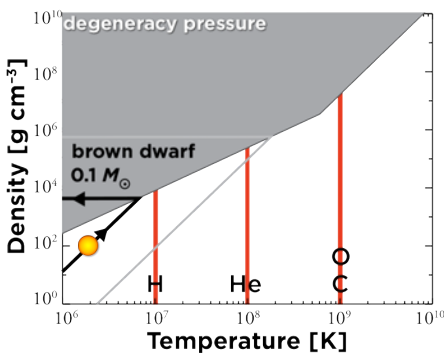
- All of that was for a star of 1 solar mass. Now let's back up to the beginning and follow the progress of a less massive star—for example, one that is a tenth of a solar mass. What changes?



- The trajectory of a star on this chart was dictated by the equation representing hydrostatic balance.

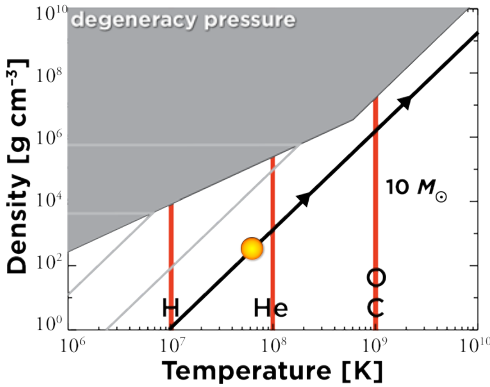
$$\log \rho_c = 3 \log T_c - 2 \log M + \text{const.}$$

- The $-2\log M$ is part of the y-intercept; the lower the mass, the higher the intercept. So, when we decrease M, the line shifts upward.



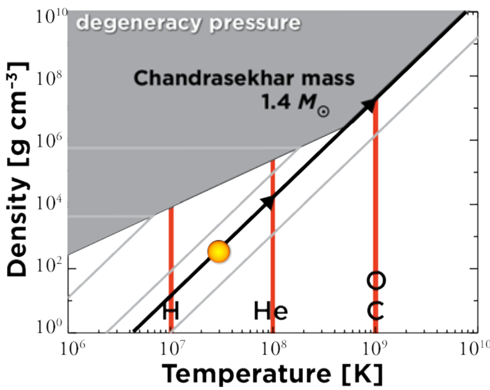
- Let's follow that new line. The star forms with low density and temperature and contracts, getting hotter and denser, but it crosses into the degeneracy zone even before hydrogen can ignite. It ends up as a sort of failed star, which cools off into oblivion—like a white dwarf, but without ever having fused hydrogen. Such an object is called a brown dwarf, a name that has nothing to do with the color of the star.

- So, the fate of an object of low mass is to become a brown dwarf or a white dwarf. What about a high-mass star—for example, one that is 10 solar masses? This time, relative to the Sun, the trajectory will be shifted down by 2 units.



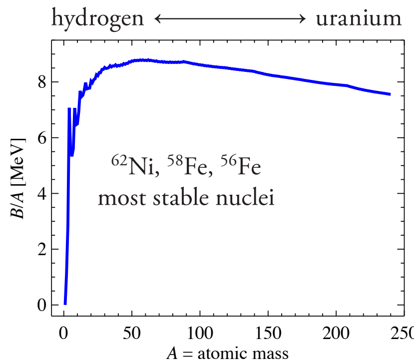
- Shifting down takes us away from the degeneracy zone. In fact, because of the kink in the boundary, the trajectory of a 10-solar-mass star never enters the degeneracy zone. The reason is that the core of a massive star is hotter at any given density; it must be hotter to generate enough pressure to maintain equilibrium. The high temperature ensures that gas pressure overwhelms degeneracy pressure, all the way up to the critical density and beyond.

- If we dial the mass back down, somewhere in between 1 and 10 solar masses, there's a special case in which the star's path just skims the boundary. This special mass turns out to be the Chandrasekhar mass—the maximum-possible mass of a white dwarf.

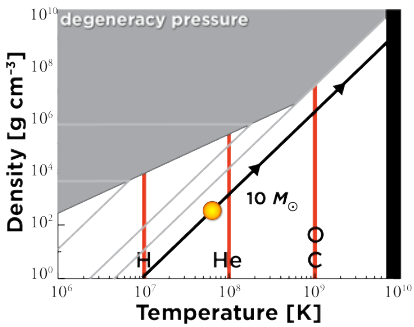


- A star of higher mass will keep contracting, igniting heavier and heavier nuclei, burning ever hotter and trying ever more desperately to prevent gravitational collapse. But it can't hold out forever.

- Think about the binding energy curve. Fusion only releases energy for nuclei on the left side of the peak. More massive nuclei prefer fission to fusion. So, once you fuse your way up to iron and nickel, at a temperature approaching 10 billion Kelvin, nuclear fusion is spent.



- Once the star crosses that boundary, which is marked in the log chart with a thick black line, gravity acts unopposed. In less than a second, the core collapses.

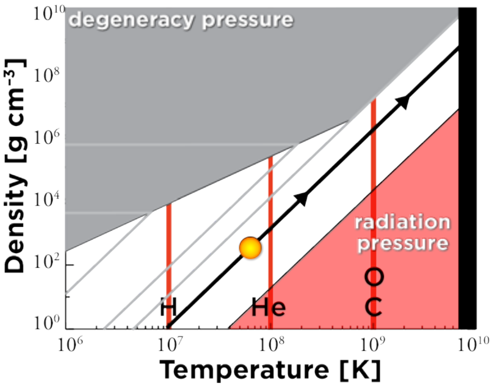


- Another possible source of pressure is radiation pressure, which is the pressure arising from photons, which, like gas particles, carry momentum and thereby exert a pressure. But it obeys a different formula than the ideal gas law: Radiation pressure is proportional to temperature to the fourth power, a much stronger dependence on temperature.

$$P_{\text{rad}} = \frac{4\sigma}{3c} T^4$$

- This means that even though radiation pressure is negligible in Sunlike stars, if we make the core hot enough, radiation pressure will dominate. To figure out where that happens in the chart, we should set the ideal gas pressure equal to radiation pressure and solve for ρ versus T.

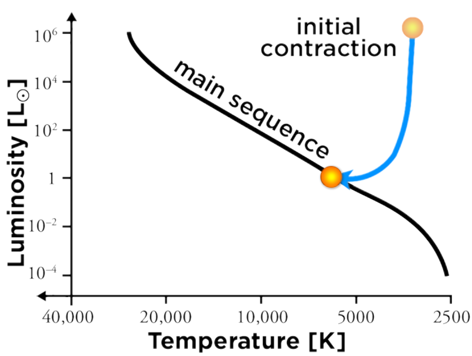
- The result is that the zone where radiation pressure dominates is in the lower-right portion of the chart, with a boundary of slope 3. A star supported by radiation pressure throughout its interior turns out to be unstable; the intense radiation from such a star would push itself apart. That helps explain why we almost never find stars more massive than about 100 solar masses. Such stars are in the radiation zone.



STELLAR EVOLUTION AT THE SURFACE

- We start the story of a solar-mass star, as before, with a clump of low-density gas being pulled together by gravity. It starts out large, in the upper right of a Hertzsprung-Russell diagram, which plots luminosity versus effective temperature. For about 10 million years, it contracts while maintaining a nearly constant effective temperature of 3000° or 4000°.

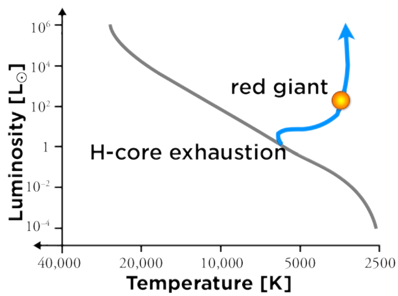
- Then, it swings to hotter temperatures. The underlying reason is that convection stops being as important—a nonobvious consequence of the fact that deep down inside, hydrogen has begun fusing. The star reaches the main sequence. It settles in for 10 billion years of steady hydrogen fusion.



- When the innermost 10% of the star's mass has been converted to helium, the star leaves the main sequence. It starts the shell-burning phase, in which fusion occurs only within a layer of hydrogen surrounding the helium core. Because the shell is sitting on top of a contracting core, the shell gradually becomes denser and hotter, causing it pump tons of power into the outer layers of the star, causing them to expand—a lot. The star swells up to become a red giant.

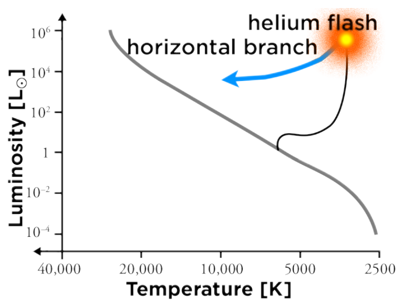
At its maximum size, the Sun will grow 200 times larger than it is today, engulfing Venus and roasting the Earth. Its luminosity will be 2000 times higher than it is today.

- Meanwhile, the core is still contracting. So, the star develops a split personality. It becomes like a star within a star: a very dense, hot core within a much larger puffball of cooler gas.

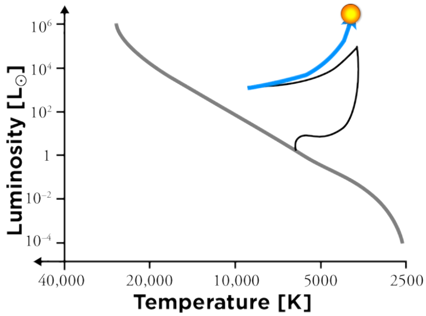


- Eventually, the core contracts enough for helium to ignite. In stars like the Sun, this is an explosive process called the helium flash—a sudden burst of fusion, related to the fact that the core is nearly degenerate when this happens.

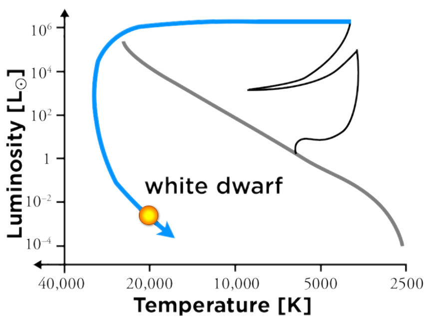
- Then, the star settles in for a steady phase of helium fusion. Now that fusion is happening in the core again, the size and luminosity go back down. The star takes up a position on what could be considered the helium main sequence, officially called the horizontal branch. It's like the hydrogen-burning main sequence but at higher luminosity.



- The star hangs out on the horizontal branch until the core runs out of helium. Then comes a replay of the same events as before. The star starts shell-burning again, this time with a helium shell surrounding an inert core of carbon and oxygen. This causes the star to swell up and become a giant star for the second time.



- At this point, the star also starts shedding. The radiation pressure in the outer layers gets intense enough to push the gas out into space—a strong wind erupts from the surface. A star like the Sun loses about half its mass during this stage. This means that by the time nuclear fusion halts, the star has expelled its outer layers and revealed the hot, dense core inside: a white dwarf.



The mass of a star is what determines its fate. A white dwarf can't exist with a mass larger than 1.4 solar masses, the Chandrasekhar mass. It would collapse under its own weight. But because of all the mass shedding that happens during the giant phases, a white dwarf can emerge from a star that was initially much larger than the Chandrasekhar mass.

That makes for a complicated relationship between a star's initial mass and its ultimate fate.

READINGS

Choudhuri, *Astrophysics for Physicists*, chaps. 3 and 4.

Prialnik, *Theory of Stellar Structure and Evolution*.

Ryden and Peterson, *Foundations of Astrophysics*, chaps. 14–18.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

QUIZ

LECTURES 13–18

1 For some binaries, we can observe both stars move around the center of mass. For others, we can see only one star moving and the other star is too faint. Which properties of the stars can we learn from each type of system? [LECTURE 13]

2 DI Herculis is an eclipsing binary star system with a period of 10.55 days. The radial velocity amplitudes of the stars are 110.7 and 126.6 km/s. What are the masses of the 2 stars? [LECTURE 13]

3 Which is more interesting: finding planets that resemble Earth or finding planetary systems that are fundamentally different from the solar system? [LECTURE 14]

4 Pretend you are viewing the solar system from 10 pc away in a random direction. How likely would it be that you could observe transits of Venus? How often would they occur, and how long would each transit last? How much would the Sun appear to fade during the transits? [LECTURE 14]

5 What are the useful aspects of a mental model of a nucleus as a cluster of marbles? What are the main limitations? [LECTURE 15]

6 Suppose, contrary to fact, that the Sun's observed luminosity came from the combustion of coal, which releases 30 MJ/kg. How many kilograms of coal would need to be burned each second? By what fraction would the Sun's mass decrease each year? [LECTURE 15]

7 The lifetime of an energy-releasing process is equal to the amount of fuel divided by the rate at which the fuel is burned. How does the lifetime of a main-sequence star depend on its mass? Which live longer: low-mass or high-mass stars? [LECTURE 16]

8 Make an order-of-magnitude estimate of the central pressure of the Earth in N/m^2. [LECTURE 16]

9 Describe the significance of the Chandrasekhar mass in your own words. [LECTURE 17]

10 The stars known as 40 Eridanus B and C form a binary with a period of 247.9 years and a semimajor axis of 34.3 AU. The ratio of distances of B and C from the center of mass is 0.37. The absolute magnitude of star B is fainter than the Sun by 4.77 units, and its effective temperature is 16,900 K. Find the mass, radius, and luminosity of star B. [LECTURE 17]

11 Why is the helium-burning phase so much shorter in duration than the hydrogen-burning phase? Why is the carbon-burning phase even shorter in duration? [LECTURE 18]

12 Make your own Hertzsprung-Russell diagram with logarithmic axes ranging from 100,000 to 1000 Kelvin and 10^{-4} to 10^{6} solar luminosities. Locate the Sun. Draw straight lines to represent 0.1, 1.0, and 10 solar radii. Sketch the future evolutionary track of the Sun. [LECTURE 18]

Go to page 338 for solutions.

Lecture 19

SUPERNOVAS AND NEUTRON STARS

The Magellanic Clouds—2 patches of light that can be seen on a dark night in the Southern Hemisphere—are neighboring galaxies caught by the gravitational pull of our galaxy. On February 22, 1987, someone pointing a telescope at part of the Large Magellanic Cloud would have seen a star named Sanduleak −69 202 with only 1 day to live.

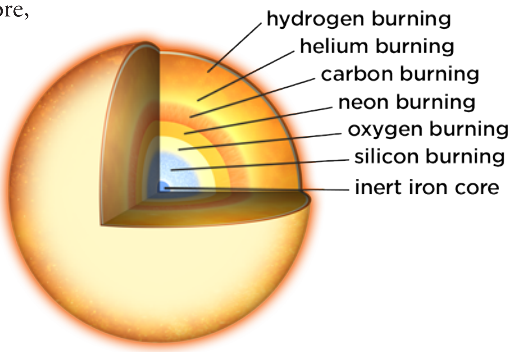
CORE-COLLAPSE SUPERNOVAS

- Sanduleak –69 202 is a star of nearly 20 solar masses that was shining for 10 million years. Then, the core ran out of hydrogen and started contracting, surrounded by a shell of burning hydrogen. Soon, the core became hot and dense enough to fuse helium into carbon.

- The helium lasted 500,000 years. But eventually the core was completely converted into carbon and oxygen, surrounded by a shell of burning helium, which was itself surrounded by a shell of burning hydrogen.

- The core contracted enough to ignite the carbon, but fusing heavier elements doesn't produce as much power. The nuclear binding energy curve starts flattening out as the peak is approached. So, the supply of carbon lasted only a few hundred years.

- The core contracted some more, igniting heavier elements, getting ever denser and hotter, with the ashes from one reaction serving as the fuel for the next one. The core started to look like an onion, with concentric shells of fusion.



- After just a few years, the core worked its way up to silicon, forging it into iron. This last, most desperate phase of fusion lasted only 1 day. Now the core is nearly all iron, one of the most stable nuclei. Fusing iron doesn't release energy; it requires energy. So, there's no more nuclear power available to prevent the core from gravitational collapse.

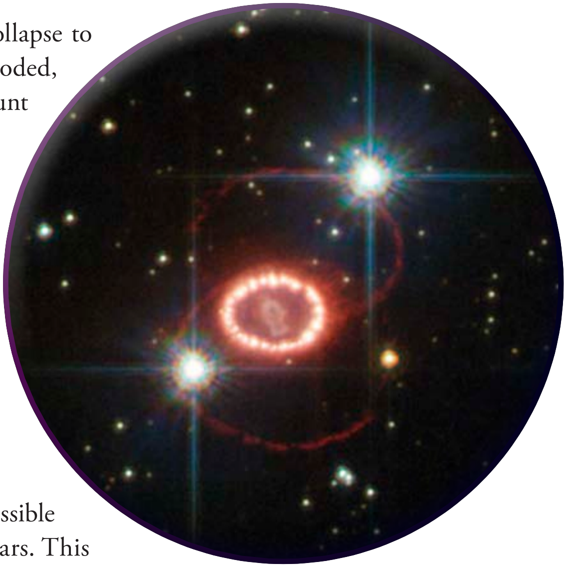
- And despite shedding lots of material up at the surface from the star's radiation pressure, the core is still more massive than the Chandrasekhar limit of 1.4 solar masses—which means electron degeneracy pressure is powerless to prevent collapse.

- So, the floor drops out. All the material in the core falls freely toward the center. The free-fall time is about half a second for an Earth-sized object.

$$\text{free-fall time}\ \ T \sim \sqrt{\frac{R^3}{2GM}}$$

$$T \sim \sqrt{\frac{R_\oplus^3}{2G \cdot 5\,M_\odot}} \sim 0.5\,\text{sec}$$

- But the star didn't just collapse to form a black hole. It exploded, releasing a staggering amount of energy.



- This kind of event, a core-collapse supernova, only happens about once per century in a galaxy like the Milky Way. So, the astronomers of 1987 were excited to see one so relatively nearby.

- They observed it at all possible wavelengths over many years. This allowed them to estimate the total energy released by the explosion in all detectable forms, including light, x-rays, and the kinetic energy of the shock waves that slammed into the surrounding gas. They came up with a figure of order 10^{44} joules. Astronomers had already seen lots of core-collapse supernovas in more distant galaxies—not in nearly as much detail—but the total energy always tends to be of order 10^{44} joules.

Within a few seconds, a core-collapse supernova releases about as much energy as is radiated by the Sun over billions of years, and it releases even more energy in a hidden form.

NEUTRINOS AND NEUTRON STARS

- By 1987, experimental physicists had built a new generation of neutrino detectors—huge underground vaults of liquid surrounded by equipment that can spot the tiny flashes of light that happen whenever a high-energy neutrino makes a rare collision with an atomic nucleus somewhere in the liquid.

- These new detectors weren't designed to study neutrinos from the Sun. The new purpose was to measure the decay of protons, which—if it happens—would produce neutrinos. Protons are very stable, but some of the particle physics theories that were in vogue predicted that very rarely a proton would decay.

- None of these experiments ever spotted a decaying proton. But some of them did detect a burst of neutrinos on February 23, 1987. The total number of neutrino detections that could be attributed to the supernova was 25.

- After correcting for how many neutrinos weren't detected because they sailed right through the Earth, physicists estimated that the total energy of all the neutrinos produced in the supernova explosion was of order 10^{46} joules—100 times more than all the light and heat!

- This teaches us that we should think of a core-collapse supernova as a neutrino explosion that, as an incidental by-product, produces a small fireworks display. Where do all these neutrinos come from?

- Whenever you see neutrinos, you know the weak nuclear interaction has been up to something. What happened in Sanduleak –69 202—and what happens in all core-collapse supernovas—is that the core of the star contracts so much that the positive ions and negative electrons are crushed together, packing so close that even the weak interactions, with their very short range, can mix them up and transmute them into neutrons.

- As it turns out, the end point of those transmutations—the most stable mixture at those high densities—is mainly neutrons, with only a minority of protons and electrons. The star becomes a neutron star.

- When nucleons are pressed together tighter than a femtometer, the strong nuclear interactions become important, preventing further compression. If we think of a nucleus as a cluster of rigid marbles, a neutron star is a cluster of 10^{57} marbles, and the gravity is so strong that the marbles are noticeably squished.

- If the mass of the core of the collapsing star is too large, more than about 2 solar masses, the marbles shatter. The strong force gives way, and the core collapses all the way to become a black hole—a point with 0 volume endowed with the mass of the star.

- But less than 2 solar masses and the strong force is just strong enough to prevent this fate. That's when you get a neutron star.

- We can estimate the size of a neutron star by analogy with our previous work on white dwarfs. What we did then was set the pressure required for hydrostatic balance equal to the pressure from electron degeneracy. That led to a relationship between mass and radius, with radius being proportional to 1 divided by the cube root of mass.

$$R_{\mathrm{wd}} \sim \frac{\hbar^2}{G\,\underset{\substack{\uparrow \\ m_{\mathrm{n}}}}{m_{\mathrm{e}}}\,m_{\mathrm{p}}^2}\left(\frac{M}{m_{\mathrm{p}}}\right)^{-1/3}$$

- The formula for the radius has a factor of the electron mass (m_e) in the denominator. That traces back to our decision to consider only the electrons and ignore the degeneracy pressure from the nucleons. This made sense at the time, because degeneracy pressure varies inversely with particle mass and a nucleon is 1800 times more massive than an electron.

- But if most of the electrons are gone because they merged with the protons, then it's the neutrons that provide the degeneracy pressure. If we repeated our calculation for neutron stars, the electron mass in the denominator would be replaced by the neutron mass (m_n).

- The consequence is that the radius is lower, by a factor of 1800, for a given mass. We saw that electron degeneracy leads to stars the size of Earth, which has a radius of 6400 kilometers. Neutron degeneracy should lead to stars the size of just 3 or 4 kilometers.

- Although this line of reasoning does give the right order of magnitude, in reality most of the internal pressure in neutron stars is from the strong nuclear force, not just degeneracy pressure. That's something we didn't have to worry about for white dwarfs. But for neutron stars, it's a big worry, and it makes calculating the mass-radius relationship difficult.

- Most theorists who have tried arrive at a radius in the neighborhood of 10 kilometers for a neutron star with a typical mass of 1.5 solar masses.

A neutron star is a city-sized object that packs the mass of 1.5 Suns.

- A neutron star has a comparable mass to a white dwarf but a radius 1000 times smaller, so the density of a neutron star is higher by a factor of 1000^3, or 10^9.

While a cubic centimeter of a white dwarf has the mass of several metric tons, a cubic centimeter of a neutron star has the mass of a billion tons.

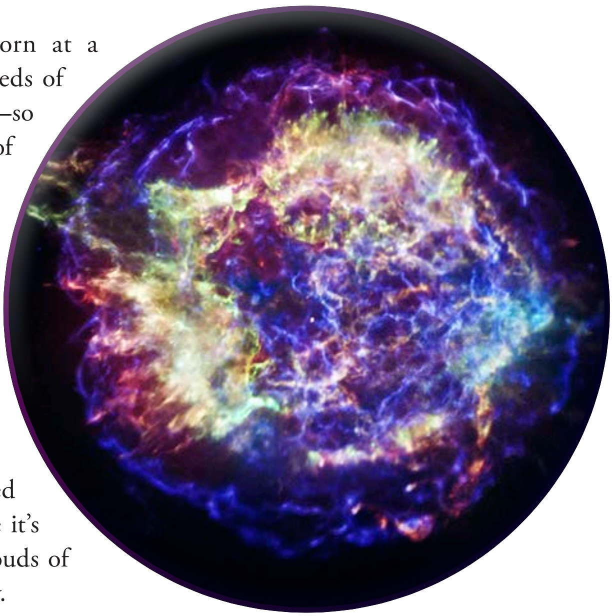
SUPERNOVA REMNANTS

- Once the collapse of a star's core has been halted by the strong force, there are still several solar masses of material in the outer layers of the star—material that is now falling directly onto the surface of the newborn neutron star. It rebounds from the surface, producing a shock wave that travels out at thousands of kilometers per second, driving the gas outward.

- This produces a lot of heat and a burst of nuclear reactions. And these reactions are not like the steady reactions at the center of the Sun, which gradually work their way up to heavier elements over billions of years; these are sudden, violent, uncontrolled reactions that reach all the way up to iron and even a little beyond.

- The elements that are forged range across the periodic table from carbon, nitrogen, and oxygen up to yttrium and zirconium. And the explosion sprays them all over the galaxy.

- A neutron star is born at a temperature of hundreds of thousands of degrees—so hot that there's hope of detecting its thermal radiation even though it's so tiny.



- Cassiopeia A is the fading remnant of a supernova explosion in the 17th century, although nobody seems to have noticed it at the time because it's behind some thick clouds of dust in the Milky Way.

- But we can see it clearly in x-ray images, which show all the gas that was pushed out into space and heated to hundreds of thousands of degrees. And right in the center is a tiny dot, representing x-rays from the neutron star.

- The Crab Nebula, 6500 light-years away in the constellation of Taurus, is a different kind of supernova remnant. The neutron star isn't just quietly cooling off. It's ejecting energetic particles that energize the surrounding gas, producing a blue-green haze that permeates the nebula, along with lots of x-rays.

THE DISCOVERY OF NEUTRON STARS

- The neutron particle was discovered by physicist James Chadwick in 1932. Just 2 years later, astronomers Walter Baade and Fritz Zwicky dreamt up the idea of a neutron star and proposed that a supernova is the transformation of an ordinary star into a neutron star. Then, in 1967, neutron stars were detected!

- But the way in which neutron stars revealed themselves wasn't anticipated by anyone and remains to this day a poorly understood phenomenon. They pulse.

- Jocelyn Bell (now Bell Burnell) and Antony Hewish discovered radio sources that emit regular pulses, like a clock. The Crab pulsar, for example, emits a pulse every 33 milliseconds.

- What we think is happening is that the neutron star is shining narrow beams of radiation from points on its surface—probably the 2 points along the axis of the star's magnetic field. And the magnetic axis need not be aligned with the rotation axis. So, as the star rotates, the beams of radiation swing around, like the beam of a lighthouse. If one of those beams happens to sweep across the Earth, we see a pulsar.

- While the Sun only rotates once a month and white dwarfs tend to rotate about once a day, the Crab pulsar spins around 30 times per second! The reason neutron stars spin so fast is the conservation of angular momentum. The angular momentum of a spinning body is $I\omega$, where ω is the angular frequency—how many radians per second—and I is the moment of inertia, which is MR^2 times some constant that depends on how the mass is distributed. For a sphere of uniform density, the constant is 2/5.

Imagine a city-sized sphere, more massive than the Sun, spinning as fast as the tires of an Indy 500 race car.

$$L = I\omega = \frac{2}{5}MR^2\omega$$

- Before it collapses, the core has an initial radius of R_{init} and spins with angular velocity ω_{init}. Then, gravity compresses the core to a radius of R_{final}.

$$R_{\text{init}}^2\omega_{\text{init}} = R_{\text{final}}^2\omega_{\text{final}}$$

- What happens to its angular velocity? We can figure out by appealing to the conservation of angular momentum. During the collapse, the star doesn't lose mass and it's not getting torqued from outside, so we can equate the initial and final values of $R^2\omega$. Therefore,

$$\frac{\omega_{\text{final}}}{\omega_{\text{initial}}} = \left(\frac{R_{\text{init}}}{R_{\text{final}}}\right)^2.$$

- Because R shrinks by a factor of 1000, the angular velocity increases by 1000^2, or 1 million.

$$\left(\frac{10^4\,\text{km}}{10\,\text{km}}\right)^2 \sim 10^6$$

- A rotation period of 1 day becomes a period of 100 milliseconds.

$$P_{\text{init}} = 1\,\text{day} \longrightarrow P_{\text{final}} = 100\,\text{ms}$$

- If you observe a pulsar long enough, you'll usually see that the rotation is slowing down very slightly. Every year, the Crab pulsar's rotation period grows by 13 microseconds. If it's slowing down, it must be losing energy. Let's figure out how much.

- The kinetic energy of a rotating object is

$$E_{\mathrm{k}} = \frac{1}{2} I\omega^2 ,$$

which we can write in terms of the rotation period, P, using $\omega = 2\pi/P$.

$$E_{\mathrm{k}} = \frac{1}{2} I\omega^2 = \frac{1}{2} I \left(\frac{2\pi}{P}\right)^2$$

- If P changes, then E must be changing. Let's take the time derivative to see the connection. Then, let's plug in the parameters of the Crab pulsar: M is 1.5 solar masses, R is 10 kilometers, P is 33 milliseconds, and dP/dt is 13 microseconds per year. After converting units and multiplying through, the answer is 5×10^{31} joules per second—which is a good match to observations of the total luminosity of the Crab Nebula.

$$\frac{dE_{\mathrm{k}}}{dt} = -\frac{4\pi^2 \overbrace{I}^{\frac{2}{5}(1.5\,M_\odot)(10\,\mathrm{km})^2}}{\underbrace{P^3}_{33\,\mathrm{ms}}} \underbrace{\frac{dP}{dt}}_{13\,\mu\mathrm{s\,year}^{-1}} = -5 \times 10^{31}\ \mathrm{J\,s^{-1}}$$

- The rotating pulsar is flinging away high-energy particles from its surface that fly off and energize the surrounding nebula. The neutron star is acting like a giant flywheel, a reservoir of rotational kinetic energy that's being tapped to light up the surrounding gas. In that sense, the Crab is a rotation-powered nebula.

- Once the period and the period derivative have been measured for a pulsar, that's usually enough to predict future pulse times very accurately. In that sense, pulsars make good clocks.

Some pulsars rival the accuracy of the world's best atomic clocks.

A category of pulsars called millisecond pulsars spin unusually rapidly even for neutron stars and are incredibly stable. By timing them, we can measure all sorts of subtle physical effects, including the slight warpages of space and time predicted by general relativity.

SUPERNOVA NOMENCLATURE

- Originally, the word "supernova" simply meant a really energetic explosion—to be distinguished from a "nova," a milder but still impressive explosion.

- Over time, different flavors of supernovas were recognized. A spectrum of the explosion at its brightest phase sometimes reveals evidence for hydrogen; others have no hydrogen. Those without hydrogen were called Type I supernovas, and the ones with hydrogen were called Type II.

- The Type I supernovas turned out to show some variety, too. Some of them have lots of silicon in their spectra while others don't. Type I was subdivided into Ia, Ib, Ic, and so on, depending on which elements were on display.

- The most meaningful physical distinction is not between Type I and Type II, but between Type Ia supernovas and everything else.

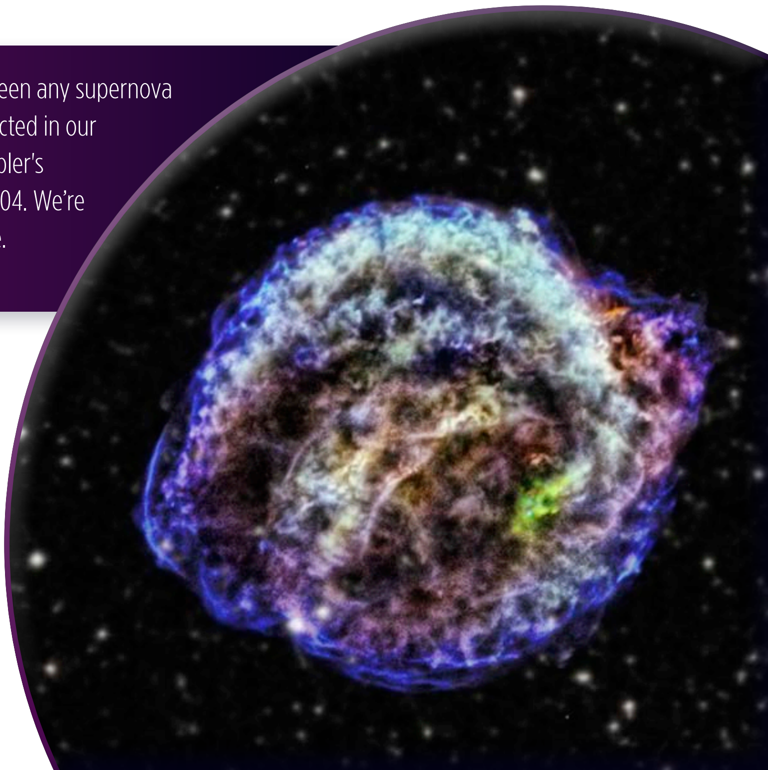
- The transformation of an ordinary star into a neutron star—called a core-collapse supernova—can be Type II, Ib, or Ic, depending on the details of how the massive star shed its outer envelope before the core collapsed.

- Type Ia supernovas have a fundamentally different origin, one that's less well understood. All the theories involve a white dwarf that for some reason becomes too massive for its own good.

- A striking thing about Type Ia supernovas is that they all have nearly the same peak luminosity; they're much more standardized than core-collapse supernovas. That's what makes Type Ia supernovas useful for measuring distances to faraway galaxies. They function as standard candles, or "standard explosions." Measure the maximum flux and infer the distance. This commonality of Type Ia supernovas might be because the exploding mass is always nearly the same—the Chandrasekhar mass.

- Type Ia supernovas also leave behind remnants that are just as impressive as core-collapse supernovas.

There haven't been any supernova explosions detected in our galaxy since Kepler's Supernova in 1604. We're overdue for one.



READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 16.

Maoz, *Astrophysics in a Nutshell*, chap. 4.

Ryden and Peterson, *Foundations of Astrophysics*, chaps. 14–18.

Shapiro and Teukolsky, *Black Holes, White Dwarfs, and Neutron Stars*.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 20

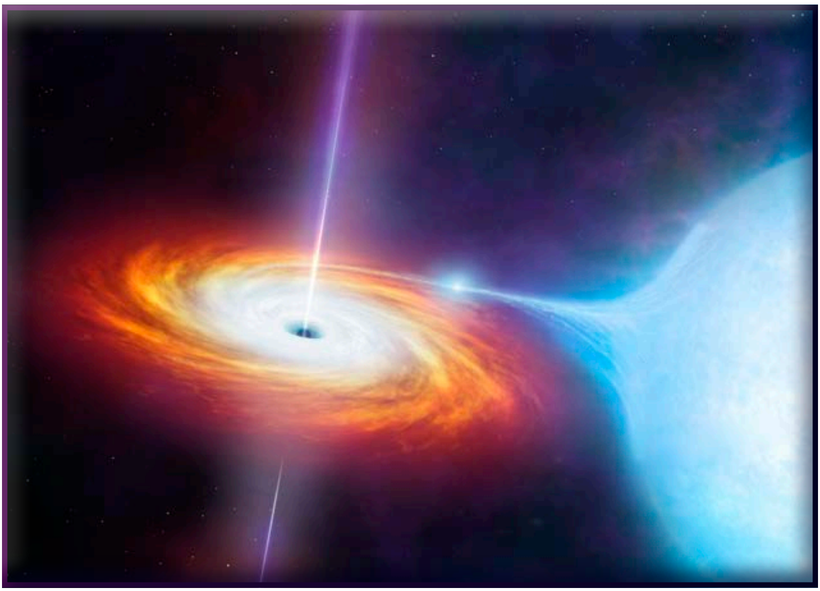
GRAVITATIONAL WAVES

One way that neutron stars announce themselves to astronomers is as radio pulsars. Another way is through x-rays. In the early days of x-ray astronomy, in the 1970s, some of the strongest sources were found to be binary stars—or, at least, Doppler spectroscopy revealed 1 star that was being pulled around by a massive companion, but there was no light from any second star, just x-rays.

ACCRETION DISKS

- Imagine an ordinary star and a neutron star in a tight orbit. The normal star swells up into a giant star and gets distorted by the strong tidal forces from the neutron star until it violates the Roche limit.

- Material from its outer layers spills out, but it can't fall directly onto the neutron star; it has too much angular momentum. So, it swirls around the neutron star, forming a vortex, like the water swirling down the drain of a bathtub. In this context, the vortex is called an accretion disk.



- Gradually, the material loses angular momentum and falls toward the neutron star, releasing a lot of gravitational potential energy along the way. The gas heats up to tens of millions of degrees and glows with x-rays. We can calculate how much energy is released by following the progress of a small amount of gas, with mass m, as it spirals down the drain.

- The spiraling is a slow process, so at any moment the gas follows a nearly circular orbit. If the gas starts at some large distance, where the orbital energy is practically 0, and descends to the neutron star of radius R, the energy released is given by the usual formula for orbital energy,

$$\Delta E = \frac{GMm}{2R},$$

where M is the mass of the neutron star.

- Accretion disks are widespread throughout astrophysics. You find them around not just neutron stars but also white dwarfs and black holes. Even a normal star can accrete mass from a companion, though it doesn't release nearly as much energy because the material doesn't have as far to fall.

In terms of energy, gravitational accretion outdoes nuclear fusion by more than an order of magnitude.

- For normal stars and white dwarfs, the accretion disk glows in the visible and ultraviolet range. For neutron stars and black holes, it's x-rays.

MERGING PAIRS OF BLACK HOLES

- Until the early 20th century, gravity was conceived as a force that one mass exerts on another. A mass can just reach across space and pull on other masses. But Einstein taught us that masses don't directly pull on each other. Instead, a mass distorts the surrounding space; it curves it, the way a bowling ball curves the surface of a trampoline.

- Then, other masses follow the curvature of space instead of just coasting by. If you roll a marble on the trampoline, it curves around the bowling ball. From high above, it might look like the bowling ball is pulling on the marble. But the connection is indirect.

- One implication is that gravity doesn't act instantly. If a mass changes location, there's a slight delay as the stretching of space is transmitted outward. That allows for the possibility of gravitational waves.

- If you're in a large swimming pool and someone does a cannonball dive on the other side of the pool, it takes a while for the waves to reach you. Likewise, when black holes crash into each other, there's a delay before the resulting distortions in space reach us. But while a water wave might travel a few meters per second, gravitational waves travel at 300 million meters per second—the same speed as light.

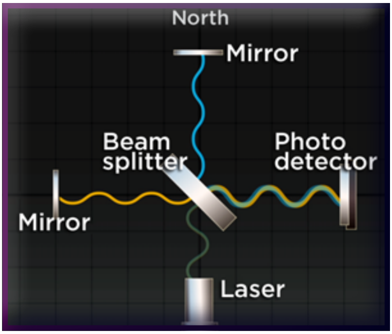
- Another difference is what happens when the wave reaches you. If you're in the water and a wave reaches you, you bob up and down. But if you're in space and a gravitational wave reaches you, you get stretched—first vertically, then horizontally, then back to vertical, and so on. You feel oscillating tidal forces.

- The stretching motion is unimaginably small. Even the strongest gravitational waves that pass through the Earth stretch things by only a few parts in 10^{21}.

- But even though gravitational waves are incredibly weak, they have been detected by a team led by scientists from MIT and Caltech at the Laser Interferometer Gravitational-Wave Observatory (LIGO). LIGO takes advantage of the wave nature of photons to sense the tiny changes in length induced by a gravitational wave. They use electromagnetic waves, from a laser, to detect gravitational waves.

- They built tunnels in 2 perpendicular directions; let's call them north and west. A gravitational wave changes the relative lengths of the tunnels: One gets stretched while the other gets scrunched, and vice versa. So, the goal is to continuously monitor the relative lengths of the tunnels.

- The way it's done is to send laser light through a beam splitter, essentially a half-silvered mirror that reflects half the light west and transmits the other half north. At the end of each tunnel is a highly reflective mirror that bounces the light back to the beam splitter. From there, the light either goes back toward the laser or bounces to a photodetector.



- If the 2 laser beams arrive at the detector in phase—that is, if each crest arriving from the west is met by a crest arriving from the north—then the beams interfere constructively. The combined light has a higher amplitude than either beam separately, and the detector registers a large signal.

- If the 2 optical paths differ by half a wavelength, then the beams interfere destructively. Each crest from the west is met by a trough from the north and the signals cancel out—no light gets to the detector.

- The experimenters arrange for destructive interference and work as hard as they can to maintain that condition, keeping all the equipment stable. They call this fringe lock.

- Then, a gravitational wave comes through, stretching one tunnel and squishing the other, back and forth. This breaks the fringe lock. Some light reaches the detector, oscillating in intensity at the same frequency as the incoming gravitational wave.

- But the stretching effect is minuscule. The strength of a gravitational wave is quantified by its strain, defined as the fractional change in length it produces, $\Delta L/L$, which is at best 1 part in 10^{21}.

$$\text{strain} = \frac{\Delta L}{L}$$

- One thing that helps is to make the tunnels very long—increase L as much as possible so that ΔL is also larger and easier to detect. The LIGO tunnels are 4 kilometers long. In addition, there are extra mirrors that bounce the light back and forth hundreds of times, which has the same effect as lengthening the tunnels.

- Another challenge is making sure the mirrors are kept isolated from external vibrations. The slightest tremor—from a distant earthquake, a passing truck, or even sound waves—would knock the mirrors out of alignment. So, the experimenters pump all the air out of the tunnels and use the world's best shock absorbers.

- And even then, even in fringe lock, the detector is never totally dark. There are always some vibrations unrelated to gravitational waves that simply can't be filtered out.

- The solution for this was to build 2 interferometers at different sites: one in Hanford, Washington, and the other 3000 kilometers away in Livingston, Louisiana. The idea is that the random fluctuations will be different at the 2 sites. You only get excited when you see the same signal at both locations at the same time—or at least nearly the same time.

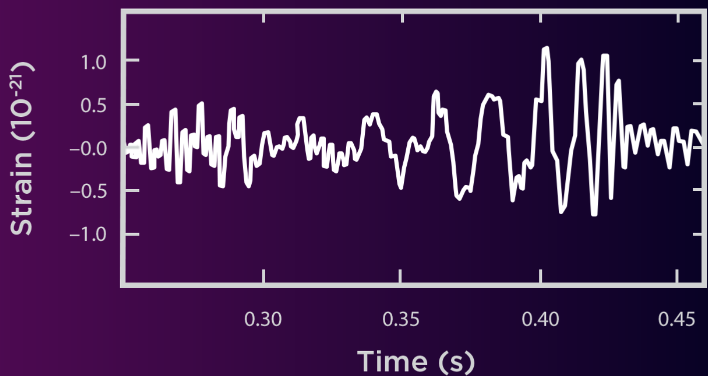
- Depending on where the wave is coming from, it will arrive at one site earlier than the other, but still within 10 microseconds, the time it takes to travel 3000 kilometers at the speed of light.

- In fact, we can use the measured delay to help figure out where the wave is coming from, although not very well. We can't pinpoint the location of the source on the sky; by measuring the delay, you can only restrict the possibilities to a certain circle on the sky.

- The solution for this was to have a third facility, which is called VIRGO and is located in Italy, build an interferometer. With 3 detectors, the position of the source on the sky can be triangulated.

The Historic First Detection of Gravitational Waves

This chart shows the signal of the gravitational wave measured at LIGO's Hanford site in September 2015. The horizontal axis is time. The vertical axis is the strain, expressed as a multiple of 10^{-21}.

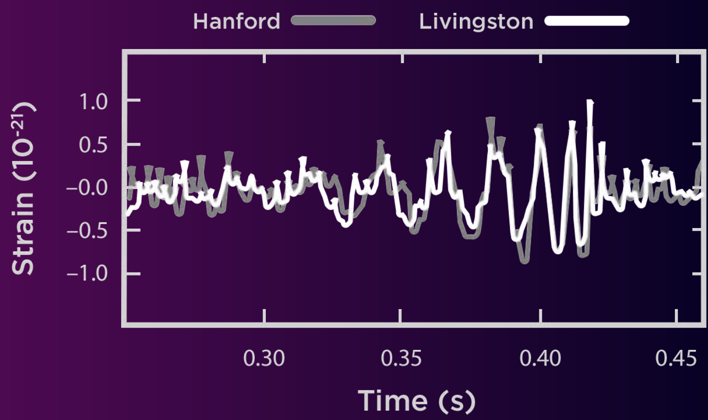


Before about 0.33 seconds, the data look like random noise. Then, an oscillation starts to build, gets stronger and more rapid, and then disappears at 0.43 seconds. That signal represents the ripples of gravitational waves that arrived at the Milky Way from a remote galaxy where billions of years ago a pair of black holes were spiraling together.

Space was churning furiously as the black holes orbited around one another. The resulting gravitational waves carried energy away from the system at the expense of the orbital energy. That made the orbit shrink, causing the black holes to speed up, which increased the rate of energy loss more. This positive feedback loop explains the quickening oscillations in the Hanford signal.

Then, the 2 event horizons merged, the 2 black holes became 1, and the waves stopped.

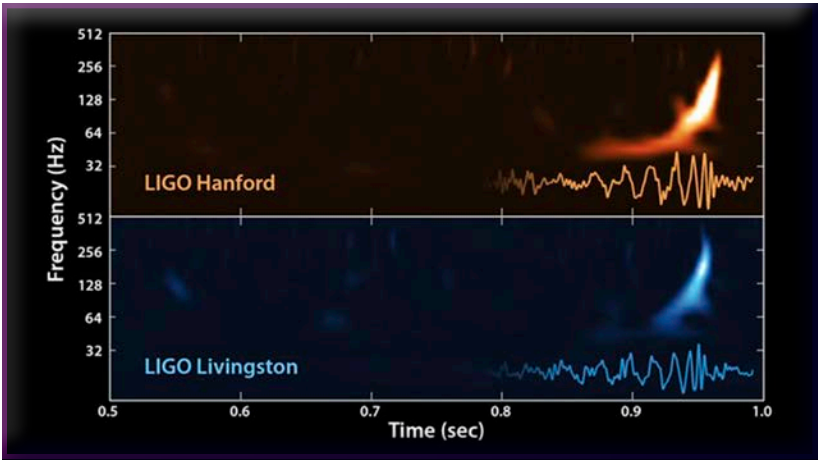
The data from Livingston tell the same story: The signal is consistent with the Hanford data after accounting for the different locations and orientations of the 2 interferometers.



THE MASSES OF THE BLACK HOLES

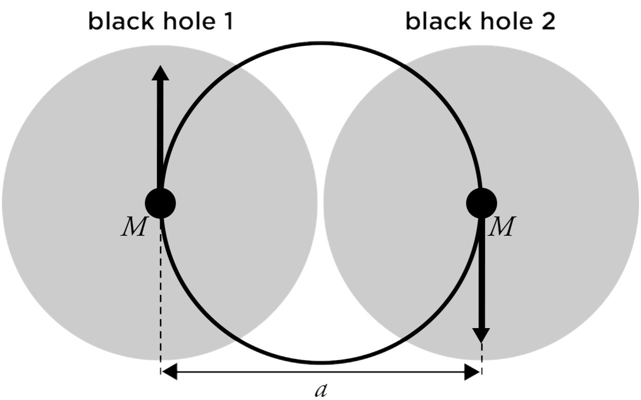
- Another way of viewing the data is a frequency-time diagram, in which the horizontal axis is still time, but the strain is indicated by color and the vertical axis is frequency—the rate of oscillations at a given time.

- Based on the maximum frequency of the signal, we can calculate the masses of the 2 black holes.



- To calculate the masses accurately requires the machinery of general relativity, but we can get an approximation with the help of Kepler's third law, which allows us to calculate the total mass of a binary if we know both the orbital period and the orbital separation.

- To keep things simple, let's consider a pair of black holes with the same mass on a circular orbit with period P and separation a.



- The data show that the gravitational wave rose in frequency from 50 to 300 hertz, which corresponds to wave periods ranging from 1/50 to 1/300 of a second.

- In our scenario, the wave period is equal to half of the orbital period. That's because after half an orbit, the black holes switch places, and because they're identical, the system looks the same again. So, the pattern of gravitational waves must repeat every half an orbital period. We conclude that the orbital period, P, shrunk to a minimum of 1/150 of a second.

- What about the orbital separation? We can't just read that off the chart, but we can reason that the period was shortest just before the final merger, when the 2 event horizons started touching each other. At that moment, a was equal to about twice the Schwarzschild radius.

- Now we can apply Kepler's third law. The total mass, $2M$, is $\frac{4\pi^2}{G}\frac{a^3}{P^2}$.

- For a, we insert $2 \cdot \frac{2GM}{c^2}$.

- Now M appears on both sides of the equation.

$$2M = \frac{4\pi^2}{GP^2}\left(2 \cdot \frac{2GM}{c^2}\right)^3$$

- We solve for M and then run the numbers, using $P = 1/150$ of a second.

$$\frac{4^4\pi^2 G^2 M^3}{P^2 c^6}$$

$$\frac{2}{4^4\pi^2}\frac{c^6 P^2}{G^2} = M^2$$

$$M = \frac{\sqrt{2}}{16\pi}\frac{c^3 P}{G} = 7.6 \times 10^{31}\ \mathrm{kg} = 38\ M_\odot$$

- The mass comes out to be 38 solar masses. This agrees pretty well with the more sophisticated analysis by the LIGO team, which found the 2 black hole masses to be 36 and 29 solar masses.

MAXIMUM STRAIN

- The other key piece of information from the chart is the maximum strain, 10^{-21}, which we can use to determine the distance to the galaxy where the merger took place. But first we need a formula telling us how strain decreases with distance.

- To determine distances, we often rely on the flux-luminosity relation:

$$F = \frac{L}{4\pi d^2},$$

which we derived based on the conservation of energy: All the power from the source gets spread out over a giant sphere of surface area $4\pi d^2$. The same is true of gravitational waves.

- But LIGO is not a flux detector. The response of an interferometer is not based on the flux of energy carried by the waves; it responds to the oscillations themselves. It directly measures the wave amplitude—the strain.

- Because amplitude varies as the square root of flux, the amplitude of a wave varies as 1 divided by distance, not distance squared. So, we expect the formula for the strain of a gravitational wave to be something divided by d.

$$\text{amplitude} \propto \sqrt{\text{flux}} \propto \frac{1}{d}$$

- And what is that something? That's a job for a general relativist. But we can make an educated guess. Gravitational waves are produced whenever a mass is shaking around, so we might expect the amplitude to depend on the mass and on how fast it's shaking.

- When black holes merge—when their event horizons touch—how fast are they moving? It turns out they're moving close to the speed of light, c. Therefore, in the strain formula, we expect the numerator to depend on M and c. And because this is gravity, G should also make an appearance.

$$\frac{\Delta L}{L} = \frac{(??)}{r}$$

- Now let's think about units. The formula we're seeking is for the strain, which is unitless; it's a fractional change in length. And the denominator is distance, so the numerator must also be a distance.

- To build a quantity with units of distance from M, c, and G, we use GM/c^2. That's proportional to the Schwarzschild radius, which has units of distance.

$$\frac{\Delta L}{L} \overset{?}{\sim} \frac{GM}{c^2 d}$$

- In short, we expect the strain to be of order GM/c^2d. We can use this fact to calculate the distance the gravitational waves traveled before they reached the LIGO detectors in September 2015. We solve for d and insert 10^{-21} for the strain and 39 solar masses for M, giving a distance of 1.8 billion parsecs.

$$d \sim \frac{G\overset{\substack{38\,M_\odot \\ \swarrow}}{M}/c^2}{\underset{\substack{\nwarrow \\ 10^{-21}}}{\Delta L/L}} \sim 1.8 \times 10^9 \text{ pc}$$

- A more accurate analysis, based on general relativity, gives 0.4 billion parsecs. We got the right order of magnitude.

- A billion parsecs is a long way away. This tells us 2 things: Mergers between giant black holes must be rare if the first one we found was so far away, and the total energy released in gravitational waves must be tremendous to be detectable at such a distance.

The current best estimate is that a galaxy like the Milky Way hosts 1 merger between giant black holes every 100 million years.

ENERGY RELEASED

- We can calculate the energy released in gravitational waves by asking how much the orbital energy decreases when the black holes start in some wide orbit and spiral inward until their event horizons touch. The answer is $GM^2/2a$, where a, the orbital separation, is twice the Schwarzschild radius.

$$\Delta E = \frac{GMM}{2a} = \frac{GM^2}{2 \cdot 2R_\mathrm{S}}$$

The energy released in gravitational waves is 100 times more than the energy released in a core-collapse supernova.

- We can simplify that, and when dust settles, the answer is $1/8Mc^2$. The efficiency is $1/8$. If we plug in 39 solar masses, ΔE comes out to be about 10^{48} joules.

$$= \frac{GM^2}{4}\frac{c^2}{2GM}$$

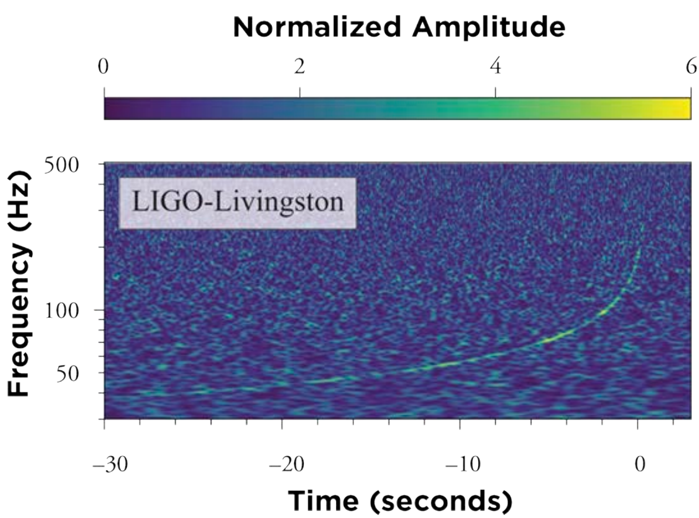
$$= \frac{1}{8}\underset{\substack{\uparrow \\ 39\,M_\odot}}{M}c^2 \sim 10^{48}\ \mathrm{J}$$

- And almost all that energy is released within a tenth of a second, so for that brief interval, the luminosity is of order 10^{49} watts. This can be expressed as 10^{23} times the Sun's luminosity, or 10^{12} times the luminosity of all the stars in the Milky Way Galaxy. These 2 merging black holes were—for a moment—emitting more power than a trillion galaxies!

MERGING NEUTRON STARS

- Another momentous discovery came in August 2017, when LIGO and VIRGO detected the gravitational waves from a pair of merging neutron stars.

- In that case, the frequency-time diagram showed a longer signal, lasting 30 seconds instead of just a fraction of a second. That's because neutron stars are only around 1 solar mass, instead of 40, and it takes longer for them to radiate away all their orbital energy.



- Astronomers detected the afterglow of the merger—a new source of light, x-rays, and radio waves emanating from a distant galaxy.

- Colliding black holes produce no electromagnetic radiation; they're black. But colliding neutron stars produce a fireball. And the observations of the afterglow confirmed the predicted characteristics of the fireball.

- The fireball models also predict that the conditions are so extreme that a series of exotic nuclear reactions takes place—reactions that require a massive dose of neutrons—forging many of the heaviest elements in the periodic table, such as bismuth, thorium, uranium, and gold.

It's now thought that almost all the gold in the universe is thrown off like sparks from colliding neutron stars. Each collision produces a few Earth masses of gold.

READINGS

Abbott, et al., "Observation of Gravitational Waves from a Binary Black Hole."

Bartusiak, *Einstein's Unfinished Symphony*.

Levin, *Black Hole Blues and Other Songs from Outer Space*.

Thorne, *Black Holes and Time Warps*.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

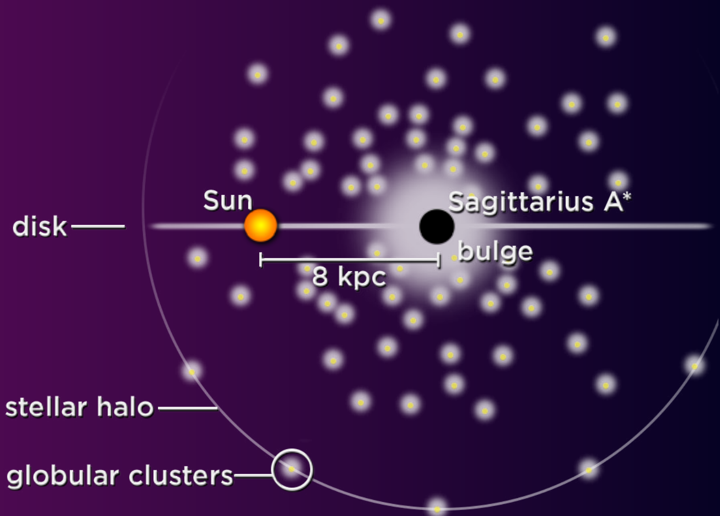
Lecture 21

THE MILKY WAY AND OTHER GALAXIES

From a given point on Earth, we only see a section of the Milky Way—which our ancestors thought looked like spilled milk—but we can piece together an all-sky view, which shows a flattened distribution of stars, like a disk. When we look along the "galactic equator," we're looking within the plane of the disk at multitudes of distant stars. When we look above and below the disk, we hardly see any stars by comparison. We also see a brightening in the middle of the map. That's the center of the galaxy, where the disk thickens into a bulge, a sort of elliptical blob of stars.

MILKY WAY

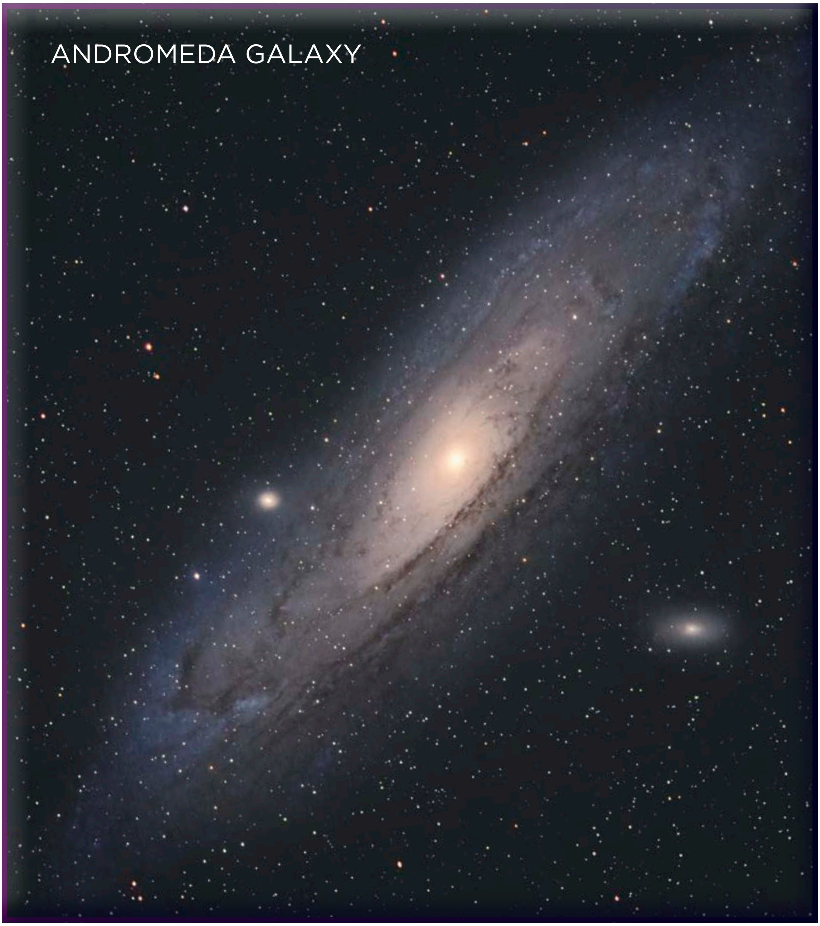
At the center of the Milky Way is the black hole Sagittarius A*, with the mass of 4 million suns. It's surrounded by the stellar bulge, which is a few kiloparsecs wide and has a mass of 20 billion suns. Then there's the disk, with a total mass of around 70 billion suns. The Sun is in the disk about 8 kiloparsecs from the center.



Outside the plane of the disk there aren't as many stars, but there are some, and they're arranged in a more spherical distribution, called the stellar halo. That's also where we find the globular clusters, each with up to a million stars packed into a tight ball.

WHAT DOES OUR GALAXY LOOK LIKE?

- We have no way to obtain a picture of the Milky Way from afar, but we can look at neighboring galaxies, which are probably like ours. Andromeda is the nearest big galaxy to ours. It's "only" 2.5 million light-years away, or 0.8 megaparsecs. The Milky Way probably looks something like Andromeda.



- Looking farther away, we can find galaxies that we're viewing from higher above the disk, giving us a better view of the spiral patterns of stars, gas, and dust.

- Very often, the spiral arms appear bluer than the bulge. That's because the disk is where new stars are formed—including blue, massive stars. The bulge is mainly older stars, so all the blue, massive stars are gone; they already evolved into red giants. The spiral arms are also dotted with colorful nebulas, marking the locations of star-forming regions.

> The Sun makes a complete orbit of the Milky Way every 200 million years, traveling at about 220 kilometers per second.

- A spectrum of a disk galaxy reveals that the disk is rotating. You can tell from the Doppler effect that half of the galaxy is approaching us, so it's blueshifted, while the other half is receding, so it's redshifted.

- In addition to disk galaxies, we find a lot of galaxies that look like featureless blobs of stars—all bulge and no disk. They're called elliptical galaxies. The stars in elliptical galaxies tend to be older and redder than in spirals, and they don't all rotate together or show much coordinated motion. Instead, the stars are orbiting more randomly in all directions.

- Those are clues that elliptical galaxies are older than disk galaxies. They've had more time for stars to age and for their orbits to randomize. Current thinking is that when 2 disk galaxies collide, the resulting train wreck becomes an elliptical galaxy.

SPIRAL AND ELLIPTICAL GALAXIES

- There are elliptical galaxies that are cut through with dark dust lanes, which are more characteristic of spirals. There are disk galaxies that don't have any spiral arms. There are disk galaxies in which the spirals attach to the bulge through an elongated bar of stars. And there are some spirals that don't seem to have any bulges at all.


NGC 5128

- Galaxies have more diverse forms than stars. They're like the orchids of the universe. Both their formation and their evolution are more contingent on circumstances than for stars—and we shouldn't expect to be able to understand them with simple equations. But we can make some progress.

- The first step is to think of a galaxy as a fluid, or a gas, but instead of being made of atoms or molecules, it's made of stars. That's a mind-bending idea. A star is millions of times more massive than the entire Earth, but galaxies are trillions of times larger still. So, when we zoom out to galactic scales, we can treat stars as the microscopic particles of a fluid. In a way, we're reverting to the original description of the Milky Way as a stream of milk!

- The fluidlike nature of galaxies is especially clear in images where 2 galaxies are interacting with each other. In some cases, the tidal gravity from one galaxy causes stars to spray out of the other galaxy, forming long, thin streams, like a fountain.

- In other cases, the collision of 2 galaxies launches spherical waves in the pattern of stars, which look like wispy bubbles and curtains.

- But there are major differences between an ordinary fluid and a fluid of stars. One of them relates to the mean free path, the average distance a particle travels before it collides with another one.

- The mean free path of a nitrogen or oxygen molecule in air is about 0.1 microns. In the Sun, the mean free path of a photon is about a millimeter. In both cases, the mean free path is much smaller than the scales we're usually interested in—that's why we can assume that the particles rapidly exchange energy and achieve thermal equilibrium and that each position can be associated with a well-defined temperature.

- What about galaxies? How far does a star travel, on average, before it hits another star?

- The defining criterion for the mean free path is $n\sigma\ell = 1$, where ℓ is the mean free path, n is the number density of the particles, and σ is the cross section, the area of the "target" a particle has to hit for an interaction to occur.

- In our neighborhood of the Milky Way, the density of stars is about 1 star per cubic parsec. What about the cross section, σ?

- Imagine throwing a star at another star of the same radius, R. For a collision, the centers of the stars must come within a distance of $2R$ of each other. So, the target area is a circle of radius $2R$, and the collision cross section is π times $4R^2$.

$$\sigma = 4\pi R^2$$

- Using the radius of the Sun, the mean free path comes out to be 2×10^{14} parsecs.

$$\ell = \frac{1}{n\sigma} \sim \frac{1}{(1\,\mathrm{pc}^{-3})(4\pi R_\odot^2)} \sim 2 \times 10^{14}\,\mathrm{pc}$$

- The size of the Milky Way is "only" 2×10^4 parsecs. According to this calculation, the Sun could cross the Milky Way 10 billion times before it hit another star!

- This calculation is a little unfair, because stars don't have to physically touch in order to deflect each other and exchange energy. So, the true cross section for stellar interactions is larger than $4\pi R^2$. But even when you take that into account with a more complex calculation, the fact remains that a typical star in the Milky Way is unlikely to ever hit another star.

- Stellar collisions only occur in unusually dense groupings, such as globular clusters, or when they're caused by effects other than long-range gravitational forces, such as energy loss due to tidal forces or gravitational radiation.

Wouldn't you love to see a galaxy collision up close?

The Andromeda Galaxy, the nearest big spiral galaxy, is headed straight for the Milky Way at 110 kilometers per second. Over the next 4 billion years, Andromeda will loom larger and larger in the sky until the 2 galaxies begin merging into a single galaxy.

But there's no reason to be concerned. Even during this galactic pileup, the probability that the Sun will have a close encounter is extremely low.

IDEAL VERSUS COLLISIONLESS GAS

- So, we can think of a galaxy as a "gas" of stars—but it's not an ideal gas. It's a collisionless gas. A star's trajectory is determined not by close encounters with other stars, which are vanishingly rare, but rather by the combined gravitational field of the entire galaxy.

- The spatial distribution of all the stars determines the galaxy's gravitational field, which then determines the motions of the stars and changes their spatial distribution. Everything is tangled up on large scales rather than small scales. That allows for more complicated behavior than a gas, including all the wonderful patterns and instabilities we see, such as spiral arms, bulges, and bars.

- Another big difference between an ideal gas and a collisionless gas relates to temperature.

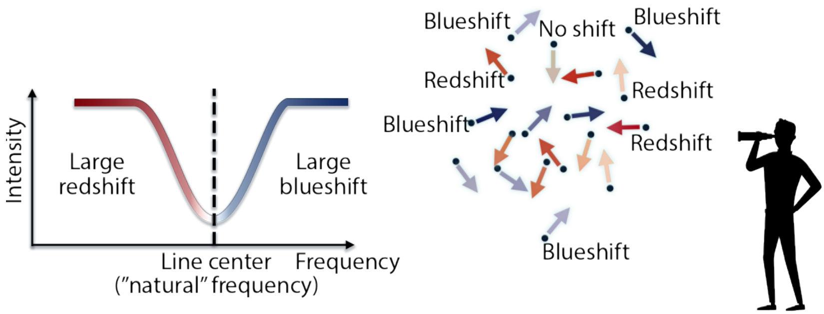
- The temperature of an ideal gas is defined to be proportional to the average kinetic energy associated with the random motions of the gas particles. The average value of $1/2mv^2$ is $3/2kT$, where k is Boltzmann's constant.

$$\left\langle \frac{1}{2}mv^2 \right\rangle = \frac{3}{2}kT$$

- We can try to do something similar for stars in a galaxy. We can define the dynamical temperature (T_{dyn}) to be proportional to the average of $1/2m\sigma_v^2$, where m is the mass of a star and σ_v is the velocity dispersion, or the spread in velocities of the stars at a given location.

$$\left\langle \frac{1}{2} m \sigma_v^2 \right\rangle \propto T_{\text{dyn}}$$

- We can measure the velocity dispersion by obtaining a spectrum of all the starlight from a given location in the galaxy and looking at the width of the absorption lines—the spread in wavelength. This "blurring" of the absorption lines is produced by all the different Doppler shifts of individual stars at that location.



- The concept of dynamical temperature gives us a different way to think about galaxies. The stars in both elliptical and disk galaxies have speeds of hundreds of kilometers per second. The difference is that ellipticals are dynamically hot: The stars fly around randomly, with a large dispersion. Disks are dynamically cold: The stars rotate all together, with less variation in velocity between neighboring stars.

- A spiral galaxy, then, is analogous to a cold, swirling glass of milk, while an elliptical galaxy is more like a hot puff of smoke.

- But the analogy with ordinary temperature breaks down if you push it too far. For example, in a highly elongated elliptical galaxy, the velocity dispersion is higher when you measure the components of velocity along the long axis as opposed to the short axis. So, the dynamical temperature at a given location depends on direction! That's pretty weird.

- And with galaxies, there's no exact equivalent of thermal equilibrium. If you stir up an ordinary gas, it eventually returns to a steady state of constant temperature with the velocities of the gas particles approaching a Maxwell-Boltzmann distribution.

- But in a galaxy, there is no steady state. In the fullness of time, a system of gravitationally interacting particles that can exchange energy with its surroundings will tend to eject some particles and all the rest collapse toward the center. It's a mathematical phenomenon called the gravothermal catastrophe.

THE VIRIAL THEOREM

- Imagine a bunch of particles flying around with different speeds and trajectories but held together by their mutual gravity so they can't escape. The virial theorem says that if you watch the system for a long time, then on average, regardless of the details, the total kinetic energy will be equal to –1/2 of the total potential energy.

$$\langle E_{\mathrm{k}} \rangle = -\frac{1}{2} \langle E_{\mathrm{g}} \rangle$$

The virial theorem is derived in the video lecture.

- The virial theorem is a mathematical expression of the concept that stars have a negative heat capacity—that when they lose energy, they heat up. And it applies to any isolated, gravitationally bound collection of particles, whether it's atoms in a star, stars in a galaxy, or a cluster of thousands of galaxies.

- Let's say we're studying an elliptical galaxy. We'd love to know the orbits of all the stars and how they change with time, but that's a job for a trained galactic dynamicist.

- Thanks to the virial theorem, we know that regardless of the details, the kinetic energy is –1/2 of the gravitational potential energy on average.

$$\langle E_{\mathrm{k}} \rangle = -\frac{1}{2} \langle E_{\mathrm{g}} \rangle$$

- From that, we can write a useful order-of-magnitude expression. For the kinetic energy, we'll write $1/2M\sigma_v^2$, where M is the total mass of the galaxy and σ_v is the velocity dispersion.

- For the potential energy, we'll write GM^2/R.

- In general, there will also be some numerical factor like 3/5 or $\pi/8$ that depends on how the stars are arranged, but our expression captures the right order of magnitude.

- Then, we can solve for M, giving σ_v^2R/G.

- This is useful because we can measure R based on the galaxy's angular size and its distance, and we can measure σ_v by looking at the widths of the absorption lines in the spectrum. Then, this equation allows us to estimate M, the mass of the galaxy—which we otherwise wouldn't be able to measure.

For any gravitationally bound system, if we measure the overall size of the system and the spread in velocities, we can estimate the total mass.

To do that, we do have to assume that the observed state of the system is representative of its long-term average. This would be false if we were observing the initial formation of the system or if it were about to collapse. But subject to that caveat, the virial theorem can be used to measure the masses of elliptical galaxies, the bulges of disk galaxies, star clusters, and even the black holes that reside at the centers of galaxies.

GALAXY DYNAMICS

- We now know to think of a galaxy as a collisionless gas of stars. The stars hardly ever have close encounters; they interact through long-range, collective forces. But what about the interactions of galaxies with each other?

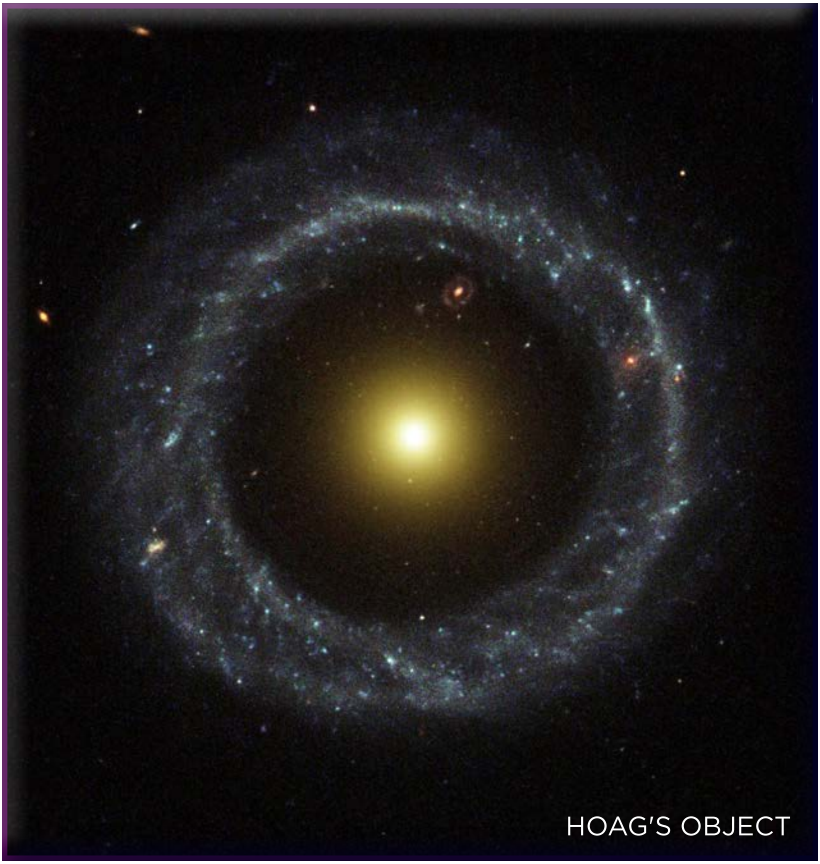
- The Hubble Space Telescope shows us that when we look in any direction in the sky, we see galaxies galore—about 1 big galaxy per cubic megaparsec. So, when we zoom way out to the gigaparsec scale, entire galaxies play the role of the fundamental particles of the universe. So, should we think of the universe as a collisionless gas of galaxies?

- It turns out the answer is no. Galaxies often interact and collide. The typical spacing between galaxies is about 50 times the size of an individual galaxy. That's much more closely packed than the stars within an individual galaxy, where the spacing between stars is millions of times larger than an individual star. Calculations of the mean free path show that over a billion years, a typical galaxy has a few-percent chance of smacking into another galaxy.

- This implies that if you look at hundreds of images of galaxies, you'll see some close encounters. And we do. We see pairs of galaxies in the midst of collisions that last hundreds of millions of years. The tidal force of each galaxy on the other causes their stars to spill out in curved arcs or to get distorted into blobby, cometlike shapes.

PORPOISE GALAXY

- In galaxies that look like elliptical galaxies surrounded by a ring of stars, the ring is probably a spiral galaxy that got too close and the tidal gravitational forces strung it out into a complete circle—sort of like a galactic-scale version of the rings of Saturn.



HOAG'S OBJECT

- The bubbles and shells we see around some galaxies are thought to be an aftereffect of galaxy mergers. The smaller galaxy oscillates inside the bigger one for a while, causing stars to be ejected in spurts before the 2 galaxies merge completely.

- Mergers were ubiquitous early in the history of the universe, when galaxies first formed. All large galaxies, including ours, formed through the mergers and accretion of smaller galaxies.

- Many elliptical galaxies were probably created by the mergers of 2 disk galaxies. Computer simulations can show the progress of simulated galaxies over billions of years as 2 disk galaxies approach each other, collide, emit streams and shells of stars, and then gradually come together and sink in their combined gravitational well, forming an elliptical.

- In other words, when 2 dynamically cold galaxies hit each other, the collision converts all that galactic-scale kinetic energy into the random motions of individual stars, resulting in a dynamically hot elliptical galaxy. In that sense, galaxy collisions are like 2 globs of cold milk crashing together, splattering, and heating up.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chaps. 2, 25, and 26.

Choudhuri, *Astrophysics for Physicists*, chaps. 7.

Maoz, *Astrophysics in a Nutshell*, chaps. 6 and 7.

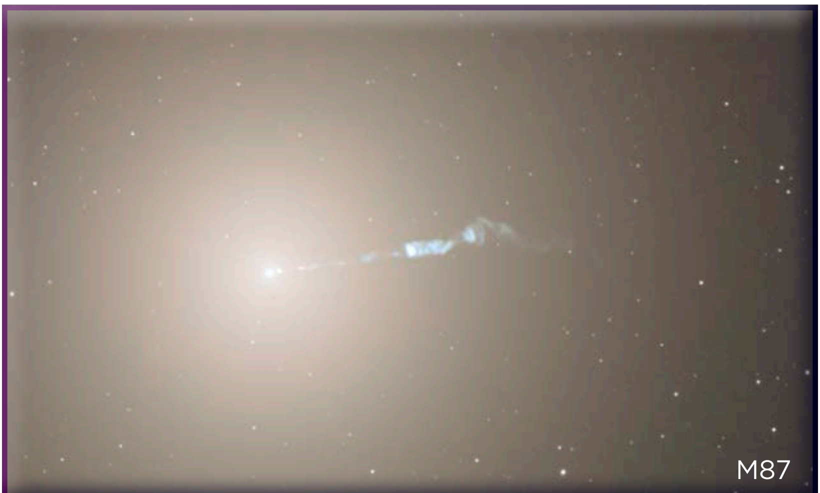
Ryden and Peterson, *Foundations of Astrophysics*, chap. 20.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 22

DARK MATTER

M87 is an elliptical galaxy with a tiny blue streak near its center. A close look shows that there's something extraordinary about this galaxy: The center is glowing brightly, and it's emitting a beam of light.

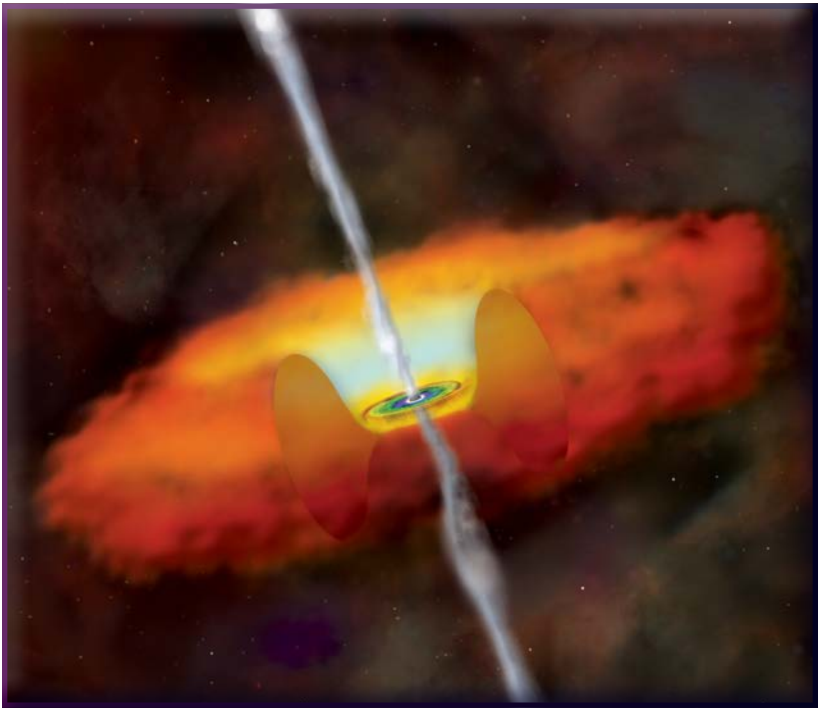


M87

ACTIVE GALAXIES

- M87 is an example of an active galaxy. There's a supermassive black hole at the center, just as there is in all big galaxies, but what's different here is that the black hole is actively accreting gas. Gas is funneling toward the black hole. The gas slowly loses energy and angular momentum, causing it to spiral inward, speed up, and heat up. By the time it's within a fraction of a parsec of the black hole, it's glowing brightly.

- After it crosses the event horizon, we never see it again—and the black hole gets slightly more massive. But just before that, at the innermost edge of the accretion disk, 2 powerful jets of plasma are being launched up and down, perpendicular to the disk.



- Why this happens is only dimly understood.

- One thing that's clear is the force propelling the jets is electromagnetic. By this point, the gas is a hot, ionized plasma, and like many plasmas, it's prone to developing a tangled internal magnetic field. As the plasma and the magnetic field approach the black hole, they whirl around in a frenzy, inducing electric fields that can accelerate charged particles vertically.

- The electromagnetic field becomes so strong that electrons and positrons emerge spontaneously out of pure energy, a phenomenon called pair production. All the newborn charged particles fly away from the disk at nearly the speed of light. And the underlying energy supply for all these fields and particles might be the black hole.

Positrons are the antimatter equivalent of electrons; they have the same mass but opposite charge.

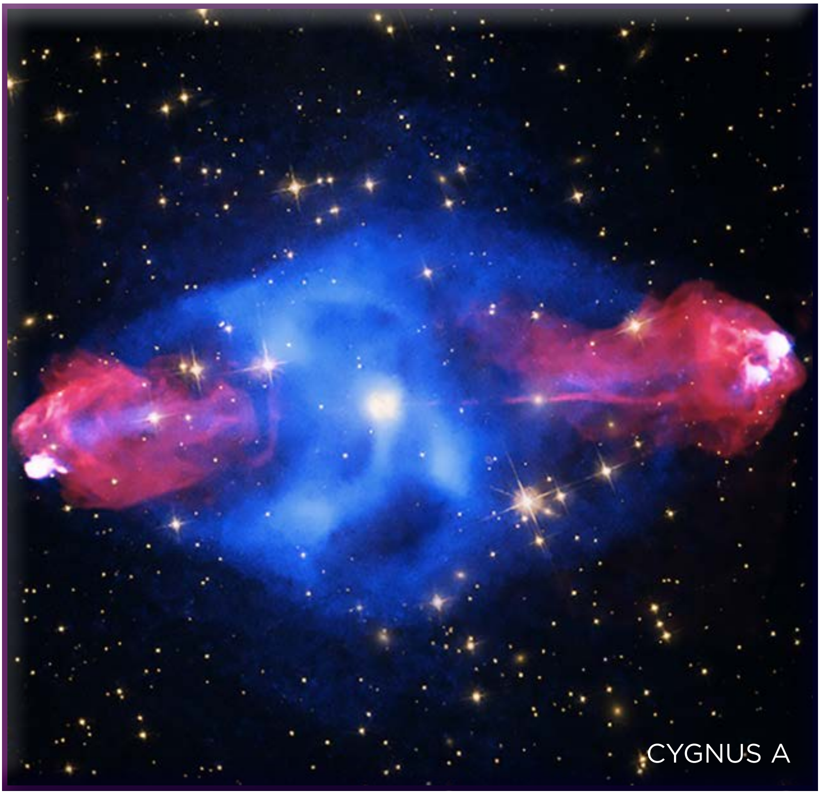
- Black holes can rotate. Accreting black holes absorb so much angular momentum from the spiraling material that they probably rotate close to the speed of light. This leads to a relativistic effect in which the surrounding space starts spinning, too, and if there's plasma there, the rotational energy can be converted to magnetic energy.

- So, the jets of an active galaxy might be rotation-powered, in the same way that the luminosity of the Crab Nebula comes from the rotation of the neutron star at its center.

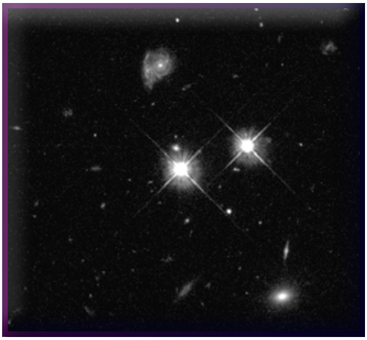
- This so-called Blandford-Znajek mechanism relies on concepts from plasma physics and, by itself, doesn't explain why the ejected particles end up forming such a narrow beam. To explain that, theorists figure that the accretion disk is not like a thin, flat plate surrounding the black hole; it's more like a fat inner tube, or torus, which restricts the escaping charged particles to narrow cones surrounding the black hole.

- There are many different types of active galaxies that go by different names, depending on how luminous they are and what angle we're viewing them from.

- For example, an optical image of Cygnus A shows a galaxy with an irregular shape. When we look with radio and x-ray telescopes, we see that the galaxy is surrounded by a haze of x-rays, and at radio wavelengths, we see 2 jets that puff out into galactic-scale fireballs. This type of active galaxy is called a radio galaxy.


CYGNUS A

- Other active galaxies look more prosaic. This Hubble image shows 2 stars along with some galaxies in the background. But in fact, the star in the middle of the image isn't a star. It's an active galaxy that is millions of times farther away than the star. The giveaway is its spectrum, which shows emission lines from the hot gas in the accretion disk.



- The accretion disk is 100,000 times more luminous than all the stars of the galaxy put together. With that bright light shining in our faces, we can't even see the galaxy. This kind of active galaxy is called a quasi-stellar object, or quasar.

How can the light from a single accretion disk overwhelm an entire galaxy of stars? How long could such a beacon possibly shine before running out of energy?

Accretion disks glow because gravity pulls the gas inward, speeding it up, converting gravitational potential energy into kinetic energy. Then, within the vortex of material, a lot of that kinetic energy is dissipated as heat.

When a small mass dm falls from far away, the energy released is $GMdm/2R$, where M is the black hole mass and R is the inner edge of the accretion disk, which is about 3 times the Schwarzschild radius (R_S)—the location of the innermost stable circular orbit.

To calculate the resulting luminosity, energy per unit time, we divide by the time interval dt over which the mass is falling and simplify.

$$L = \frac{dE}{dt} = \frac{GM}{6}\frac{c^2}{\underbrace{2GM}_{R_S}}\frac{dm}{dt} = \frac{1}{12}\frac{dm}{dt}c^2$$

Then, we solve for *dm/dt*, the rate at which the black hole must be fed to produce a given luminosity.

$$\begin{aligned} \frac{dm}{dt} = \frac{\overset{10^{15}\,L_\odot}{L}}{12c^2} &= 3.5 \times 10^{23}\,\mathrm{kg\,s^{-1}} \\ &= 6\,M_\odot\,\mathrm{year}^{-1} \end{aligned}$$

An entire galaxy has a luminosity of order 10 billion suns, and if the accretion disk is 100,000 times brighter, then L is 10^{15} solar luminosities. The value of *dm/dt* required is 3.5×10^{23} kilograms per second, or 6 solar masses per year.

That's not so much. Just toss in 1 star every few months and the radiation from the accretion disk will overwhelm the light of the surrounding 10 billion stars. Such is the power of gravitational accretion.

- The Milky Way is not currently an active galaxy. Our black hole, Sagittarius A*, is in between meals. But all big galaxies, including ours, go through phases of activity and inactivity, depending on the supply of gas to the central black hole.

- There's evidence that the Milky Way was last active a few million years ago. It's based on a gamma ray image of the entire sky made by the Fermi Gamma-ray Space Telescope. There are 2 regions of excess gamma rays above and below the center interpreted as the fading "bubbles" of high-energy particles that were spewed out by relativistic jets during an episode of accretion sometime during the Pliocene Epoch on Earth.

- Active galaxies might be part of the explanation for the connection between the mass of a black hole and the mass of the surrounding galaxy. When a galaxy grows by accreting a smaller galaxy, the central black hole feasts on the fresh supply of gas.

- Then, the powerful jets from the newly activated accretion disk push back on the infalling gas, expelling it, or at least preventing it from accreting. In this scenario, galaxy growth is self-regulating: If the galaxy starts growing too fast, the black hole gets out of control, producing outflows that put the damper on any further growth. This type of negative feedback between the black hole and the surrounding galaxy could explain why their properties are linked so tightly.

GALAXY CLUSTERS

- Galaxies don't have totally random locations. They're clustered. If you start in one galaxy, you're more likely to find another galaxy within a few megaparsecs than if you'd started at a random point in the universe.

- A galaxy cluster is a gravitationally bound system, with each galaxy following a complex orbit determined by the combined gravitational potential of all the galaxies in the cluster. And the virial theorem can be used to estimate the mass of the entire cluster.

- This was first done by an astronomer named Fritz Zwicky in the 1930s in a study of the Coma cluster. He obtained spectra of individual galaxies and saw that some were redshifted and some were blueshifted relative to the average.

- From the spread in these Doppler velocities, he measured the velocity dispersion to be 1000 kilometers per second. He also measured the cluster radius, so he could estimate the mass using the formula from the previous lecture.

$$M \sim \frac{\sigma_v^2 R}{G} \sim 5 \times 10^{13}\, M_\odot$$

- The answer he got was about 5×10^{13} solar masses. The Coma cluster has about 1000 galaxies, so Zwicky's calculation implied that the average galaxy mass is 5×10^{10} solar masses. But based on the luminosities of the individual galaxies, his best estimate for the average mass was a few times 10^8.

- In other words, the mass of the cluster seemed to exceed the sum of the masses of all the stars inside the galaxies by a factor of 100.

- This was the first clue that the overwhelming majority of the mass in the universe is not luminous. There's something in the cluster that is exerting gravitational forces on the galaxies, making them move fast. But we don't know what. We do know it's invisible at all wavelengths. And it's dark. This is the famous dark matter problem.

- The numbers have changed since Zwicky's day; these days we think the dark matter outweighs normal matter by a factor of 5 or 6. And we have lots of other evidence for dark matter.

- Starting in the late 1970s, it became clear that dark matter pervades individual galaxies, too—not just the space between them.

DARK MATTER

- For an elliptical galaxy, we can estimate the total mass in 2 ways: by measuring the velocity dispersion of the stars and using the virial theorem or by measuring the total luminosity of the galaxy and calculating what total mass of stars is needed to produce that much light. The virial mass always exceeds the luminous mass by a large factor.

- We can perform a similar test with disk galaxies. We perform Doppler spectroscopy of the starlight at different distances from the center of the galaxy. That way, we determine the rotation velocity as a function of radius. That function, $V(r)$, is called the galaxy's rotation curve.

- What would we expect the rotation curve to look like? Let's pretend, contrary to fact, that the galaxy is a point mass M. Then, the situation is just like a planet going in a circle around a star: We set the centripetal acceleration equal to the gravitational acceleration and solve for V, giving the square root of GM/r. So, we'd expect the rotation velocity to decline with increasing radius.

$$\frac{V^2}{r} = \frac{GM}{r^2} \longrightarrow \quad V(r) = \sqrt{\frac{GM}{r}}$$

- A real galaxy, though, does not have a central dominant point mass; the black hole is tiny compared to the combined mass of the stars. So, we need to calculate the gravitational force from the whole distribution of mass within the galaxy.

- Newton taught us that if we're at a distance r within a spherical mass distribution, we're allowed to ignore all the exterior mass and pretend that all the interior mass is lumped together as a single point at the center. But disk galaxies aren't spheres, and Newton's theorem doesn't apply to flat disks.

- For a disk, the formula for the rotation curve is much messier, so what physicists tend to do is assume the galaxy is a sphere anyway and trust that the answer will resemble the right answer even if it differs in detail.

- We'll describe the mass distribution with a function, M_r, that tells us how much total mass is enclosed within a sphere of radius r. As r increases, M_r rises, too, because we're enclosing more and more material until we get far enough away that we're enclosing all the mass. Then, in the limit of R, M_r levels off to a constant, the total mass of the galaxy.

$$\lim_{r\to\infty} M_r = M_{\text{tot}}$$

- To find the velocity of an orbiting star, we appeal to Newton's theorem. We simply take our previous result and replace the constant M with the function M_r.

$$V(r) = \sqrt{\frac{GM_r}{r}}$$

- In the 1970s, measurements of galaxy rotation curves became more accurate and extended to larger r. Nearly everyone expected that M_r would level off once r was larger than 10 kiloparsecs, outside the visible disk of stars. That far away, M_r should stop increasing and the rotation velocity should start declining as 1 divided by the square root of r.

- But that's not what was observed. Instead, V was found to keep rising! In many galaxies, it levels off to a constant value—but it never decreases, even well outside the disk!

- This means that if V is observed to be a constant, then we can solve for M_r and find that it grows in proportion to r.

$$V(r) = \text{const.} \quad \longrightarrow \quad M_r = \frac{V^2 r}{G}$$

- Way out there, where there are hardly any stars, as we increase r, we still enclose more and more mass. It's the dark matter again.

- More sophisticated analyses using the best-available data show that the dark matter does in fact form a nearly spherical mass distribution and extends out to hundreds of kiloparsecs!

- We are led to the stunning conclusion that a spiral galaxy is just a little bit of flotsam spinning around at the center of a much larger and more massive entity, called the dark matter halo.

- What is the dark matter? That's one of the most important unanswered questions in astrophysics. Astronomers and physicists have tried for decades to detect dark matter in some other way—besides its gravitational influence—and have failed. Theorists have tried to dream up new forms of matter that could avoid detection in all these ways.

- The idea currently in favor is that dark matter is composed of one or more hitherto unknown fundamental particles—particles that feel and exert gravity but that otherwise interact very weakly, if at all, with normal matter.

THE HUBBLE DIAGRAM

- We can measure how fast a galaxy is moving toward or away from us from its Doppler shift. The fractional shift in wavelength is equal to the radial velocity divided by the speed of light.

$$\frac{\Delta\lambda}{\lambda} = \frac{v_\mathrm{r}}{c}$$

- If the wavelengths are being stretched to larger values—toward the red end of the spectrum—$\Delta\lambda$ is positive and the galaxy is moving away from us. Likewise, if the spectrum is blueshifted, the galaxy is coming toward us.

- When we do this for the few dozen brightest galaxies in the sky, we find that the radial velocities are all over the place; some are coming toward us while others are zooming away. But when we do this for fainter galaxies, a strange thing happens: They're almost all redshifted, speeding away from the Milky Way. This was first discovered by Vesto Slipher in the early 20$^{\text{th}}$ century.

- But it gets even more interesting when we measure the distance to each galaxy. That's much harder.

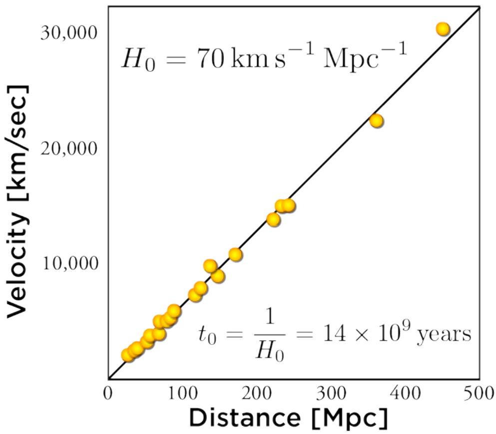
- One way to do it relies on Cepheid variable stars, which pulsate, glowing brighter and then fainter with a regular period. And the period is found to be closely linked to the luminosity, so if we measure the period, we can calculate the luminosity. They act as standard candles. We can use the flux-luminosity relation and solve for the distance, *d*.

$$F = \frac{L}{4\pi d^2} \longrightarrow d = \sqrt{\frac{L}{4\pi F}}$$

- But individual Cepheids can't be detected beyond around 50 megaparsecs. To reach out farther, we use standard explosions: Type Ia supernovas, which always seem to explode with the same energy (or at least nearly the same). They're predictable enough so that if you measure the color and duration of the afterglow, you can calculate its luminosity to within a few percent. And they're bright enough to see all the way across the universe, billions of light-years away.

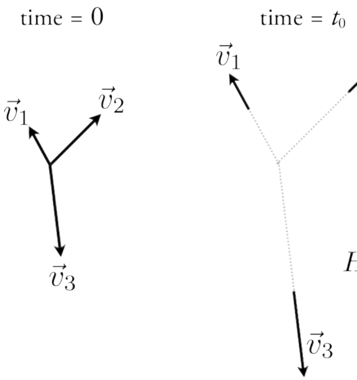
- When we measure radial velocities and distances to lots of galaxies and plot one against the other, we see a straight line: V is proportional to d.

- A diagram of this type, published by Edwin Hubble in 1929, was the first clear evidence for the big bang—because V proportional to d is what you expect from an explosion.



- Imagine a bunch of particles are at a single location at time 0. Then there's an explosion, imparting each particle with a random velocity. Some particles are shot out at high speed; others are shot out at lower speeds. At some later time, where will each particle be? The higher the speed, the farther it will get. Distance equals velocity times time, so at any given time, we'd observe d and V to be proportional.

- More formally, suppose the i^{th} particle has velocity v_i. The distance it travels between time 0 and the present time, t_0, is v_i times t_0. That implies v_i equals r_i divided by t_0. Velocity is proportional to distance, and the constant of proportionality—the slope of the line in Hubble's diagram—is 1 divided by the time since the explosion.



$$\vec{r}_i = \vec{v}_i\, t_0$$

$$\vec{v}_i = \frac{1}{t_0}\,\vec{r}_i = H_0\,\vec{r}_i$$

$$H_0 \approx 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$$

$$\longrightarrow \quad t_0 \approx 14\,\mathrm{Gyr}$$

- The slope is called the Hubble constant, H_0, and its value is about 70 kilometers per second per megaparsec. Taking the reciprocal and converting megaparsecs to kilometers and seconds to years gives a value for t_0 of 14 billion years since the big bang.

- This agrees well with the best estimate of 13.8 billion years that comes from more sophisticated analyses—but that's partly a coincidence. The story isn't as simple as galaxies coasting away from a single point. Their velocities need not be constant; after all, gravity acts to slow the galaxies down and pull them back together.

Why are the galaxies rushing away from us? Does that mean we're sitting at the center of the universe—the site of the big bang?

Even though it might seem like it at first glance, the Hubble diagram doesn't imply that there is any sort of privileged galaxy at the center. Edwin Hubble would have published a similar diagram no matter which galaxy he lived in.

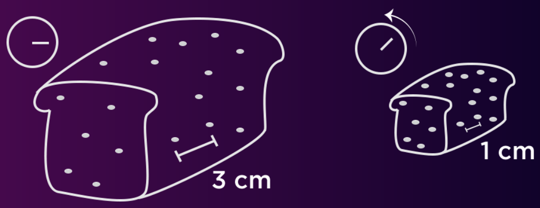
Let's say you're inside a lump of raisin bread dough, sitting on a raisin, with other raisins all around. The oven comes on and the dough expands—the bread rises. You'll see all the raisins receding from you, with velocity proportional to distance. And if you hop over to a different raisin, you'll still see all the other raisins receding from you, with velocity proportional to distance. There's no unique center of the expansion.

This result is proven more formally in the video lecture.

But if we turn back the clock, the galaxies come closer and closer together. Doesn't that mean they'll all land on a single point?

Not necessarily. For one thing, the theory of general relativity describes the big bang as the expansion of all of space from a condition of infinite density, not an explosion that took place at a single location.

In the world of baking, we need to imagine an infinite loaf of raisin bread. If we reverse the clock and watch the bread "unrise," the raisins get closer together, but it's still infinite in all directions. We just bring in more and more raisins into our field of view, and the bread gets denser until at time 0 it's infinitely dense.



READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chap. 28.

Ryden, *Introduction to Cosmology*.

Ryden and Peterson, *Foundations of Astrophysics*, chaps. 19 and 20.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

Lecture 23

THE FIRST ATOMS AND THE FIRST NUCLEI

The universe began as a collection of fundamental particles, hot and dense. Space itself was expanding, causing the density to drop, along with the temperature. With these premises and knowledge of fundamental physics, we can make quantitative predictions about the formation of the first nuclei and the first atoms that are confirmed through observations of light element abundances and the cosmic microwave background radiation.

HUBBLE'S LAW AND THE COSMOLOGICAL SCALE FACTOR

- The observation that the farther away the galaxy, the faster it's receding—Hubble's law—is one of the pillars of evidence supporting the big bang theory.

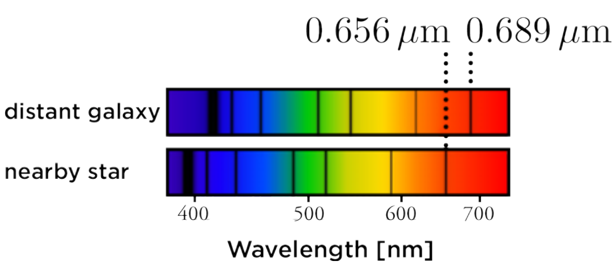
- In equation form, Hubble's law is $v = H_0 d$, where d is the distance to the galaxy, v is the velocity with which it's receding, and the proportionality constant is the Hubble constant (H_0).

$$H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$$

- Hubble's law implies that v increases without bound as we look farther and farther away. But that can't be right. The galaxies can't exceed the speed of light, can they?

- It's tricky. What we observe directly is not velocity but rather the Doppler shift of a galaxy's spectrum.

- The spectrum of a nearby star has an absorption line at a wavelength of 0.656 microns. That's the H-alpha line, which comes from electrons jumping between the second and third energy levels of hydrogen. We call 0.656 microns the rest wavelength because the star isn't moving very fast relative to us. It's the same wavelength we'd observe in a physics laboratory.



- But when we observe the starlight from a galaxy far, far away, the pattern of lines is shifted to longer wavelengths. Now, the H-alpha line is at, for example, 0.689 microns, 5% longer than usual.

$$\text{redshift} \quad z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_{\text{obs}} - \lambda_{\text{rest}}}{\lambda_{\text{rest}}}$$

$$z = \frac{0.689 - 0.656}{0.656} = 0.05 = \frac{v}{c}$$

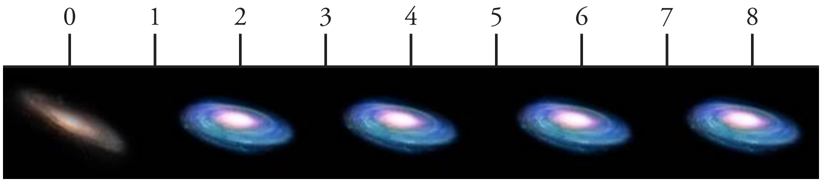
- The redshift, z, is defined as $\Delta\lambda/\lambda$, the observed wavelength minus the rest wavelength divided by the rest wavelength. In this case, z is 0.05. And if we interpret that shift as an ordinary Doppler shift, z is equal to v/c.

- An even more distant galaxy shows the H-alpha line at 1.3 microns, which is twice as large as usual! That would seem to imply that v is twice the speed of light, contradicting the fundamental principle that nothing can travel faster than light.

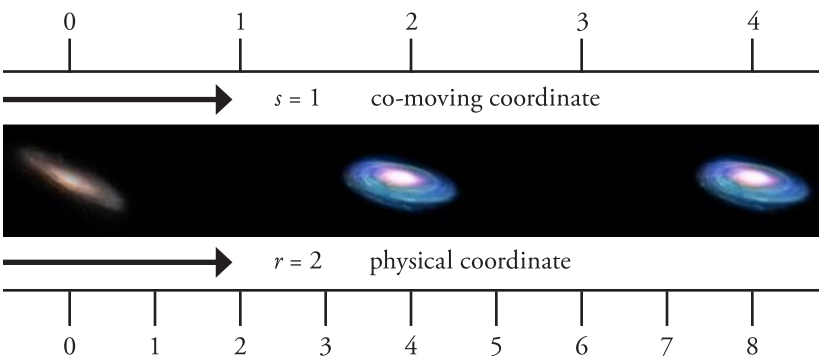
- The formal resolution of this apparent paradox is that the expanding universe needs to be described with general relativity, in which the "velocity" of a distant galaxy is not a meaningful concept. There's no unique way to define the relative velocity between 2 objects in widely separated locations when space is changing with time. So, the velocity cz that we would naively compute for a distant galaxy has no physical significance.

- Another way to think about it is that the redshift of a distant galaxy is not an inherent property of the galaxy; it doesn't depend on how fast the galaxy is moving with respect to anything else. Instead, the redshift is something that happens to photons as they travel from that galaxy to our telescopes. It's a consequence of expanding space.

- So, we need a mathematical model for expanding space. For simplicity, imagine that the universe is 1 dimensional, with all the galaxies strung out along a line. Let's use a ruler with some tick marks to keep track of their locations and put the Milky Way at the origin.



- Now imagine that our linear universe is expanding. The galaxies stay on the same tick marks, but the physical distance between the ticks is growing with time—the ruler is expanding. To emphasize the point, we put a stationary ruler beneath the expanding ruler.



- To describe the expansion, we introduce a function, $a(t)$, called the cosmological scale factor, defined as the factor by which the tick marks have expanded by the time t. It's proportional to the distance between tick marks.

- If the galaxies were all coasting away from each other at constant speeds, $a(t)$ would be proportional to t. But that's not necessarily true. For example, gravity acting alone to pull the galaxies back together would cause $a(t)$ to vary as $t^{2/3}$.

- But for now, let's not commit to any specific function; we'll just leave it as $a(t)$. And because we can use whatever units we want to measure distances, we'll choose to measure the distances between galaxies relative to their current distances—$a = 1$ at the current time, t_0. In the past, a was less than 1, and in the future, a will be greater than 1.

$$a(t_0) = 1$$

- We can track the position of a galaxy using either the stationary ruler or the expanding ruler. On the stationary ruler, the galaxy's coordinate changes with time; we'll use r for this physical coordinate.

- But on the expanding ruler, a galaxy's coordinate stays the same. Each galaxy stays close to whichever tick mark it started on. The tick marks on the expanding ruler are called comoving coordinates; it's a coordinate system that expands along with space. We'll use s for the comoving coordinate to distinguish it from the physical coordinate, r. The relationship between them is $r = a(t)\, s$.

- Now let's see what Hubble's law looks like in this new language. Consider a galaxy at a comoving distance of s. At any time t, the physical distance, r, is $a(t)\, s$. The recession velocity, v, is dr/dt, which is equal to da/dt times s. To put that purely in terms of physical distance, we'll replace s with r/a.

$$r = a(t)\, s$$

$$v = \frac{dr}{dt} = \frac{da}{dt}\, s = \frac{da}{dt}\left[\frac{r}{a(t)}\right]$$

- Then, we can rearrange the equation.

$$v = \underbrace{\left(\frac{1}{a}\frac{da}{dt}\right)}_{H} r$$

- That's Hubble's law: Velocity is proportional to distance. In our new language, the constant of proportionality—the Hubble constant—is equal to $(1/a)(da/dt)$.

- However, we see now that the Hubble constant need not be a constant in time; it could be changing, over billions of years, depending on $a(t)$. That's why purists refer to $(1/a)(da/dt)$ as the Hubble "parameter," H, and reserve the name Hubble "constant" and the symbol H_0 for the currently measured value of 70 kilometers per second per megaparsec.

INTERPRETING GALAXY REDSHIFTS

- Suppose at some time t in the past that a distant galaxy emits photons that travel for billions of years through an expanding universe and end up inside our telescope at time t_0, the present day.

- It helps to conceptually divide the journey into lots of tiny steps and pretend there are alien astronomers all along the way who are all observing the light from that same galaxy.

- The first alien is at a physical distance of dr from the source—and, crucially, we'll let dr be such a short distance that the subtleties of relativity can't possibly matter. Because Hubble's law applies to everybody, the alien observes the galaxy to be receding with a small velocity, $dv = H\,dr$.

- Because the velocity is small, we can rely on the familiar nonrelativistic formula for the Doppler shift, $d\lambda/\lambda = dv/c$, which we can write as

$$\frac{d\lambda}{\lambda} = \frac{dv}{c} = \frac{H\,dr}{c}.$$

- We learned that H equals $(1/a)(da/dt)$, so let's make that replacement.

$$\left(\frac{1}{a}\frac{da}{dt}\right)\frac{dr}{c}$$

- We can also replace dr/c by dt, the time it takes for the light to travel a physical distance dr.

$$= \left(\frac{1}{a}\frac{da}{dt}\right)dt$$

- Those replacements lead to a simple equation:

$$\frac{d\lambda}{\lambda} = \frac{da}{a}.$$

- The fractional change in wavelength equals the fractional change in the scale factor during the time interval dt.

- The same logic applies to the second alien down the line, who observes an additional fractional wavelength shift equal to the fractional change in the scale factor during the second time interval dt.

- To calculate the wavelength that we observe at the end of the journey, we need to integrate all the infinitesimal shifts the photons experience along the way. On the left side, we integrate from the original wavelength, λ_{rest}, to the observed wavelength. On the right side, we integrate from $a(t)$, the scale factor back when the light was emitted, to the current value, $a = 1$.

$$\int_{\lambda_{\text{rest}}}^{\lambda_{\text{obs}}} \frac{d\lambda}{\lambda} = \int_{a(t)}^{1} \frac{da}{a} \longrightarrow \ln\left[\frac{\lambda_{\text{obs}}}{\lambda_{\text{rest}}}\right] = \ln\left[\frac{1}{a(t)}\right]$$

$$\frac{\lambda_{\text{obs}}}{\lambda_{\text{rest}}} = \frac{1}{a(t)}$$

- The wavelength gets stretched by the same factor the universe has expanded throughout the photon's journey.

- This implies that the redshift, z, is equal to $1/a - 1$. Or, equivalently, $1/a = 1 + z$.

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{rest}}}{\lambda_{\text{rest}}} = \frac{1}{a(t)} - 1$$

$$\frac{1}{a(t)} = 1 + z$$

- This is the interpretation of the redshift we've been seeking. When we observe a galaxy to have a redshift of 2, instead of saying the galaxy is rushing away at twice the speed of light, it makes more sense to say the universe has expanded by a factor of 3 since the light was emitted.

THE COSMIC MICROWAVE BACKGROUND

- In the 1960s, Robert Wilson and Arno Penzias were radio astronomers employed by Bell Labs, where they had access to a microwave antenna that was unusually well shielded from terrestrial interference. But despite that shielding, they observed a persistent source of static and became frustrated. After painstakingly ruling out equipment problems, they concluded that space is awash with microwaves.

- Over time, observations showed that this ever-present radiation amounts to 400 photons per cubic centimeter flying in every direction with the spectrum of a nearly perfect blackbody.

- This cosmic microwave background (CMB) radiation is the most perfect blackbody known. No other light source in nature, or in any laboratory, matches the Planck function so closely over such a wide range of wavelengths.

- Interestingly, the implied temperature of the CMB that we obtain by fitting the data with a Planck function is 2.72548 Kelvin, hovering just a few degrees above absolute zero.

- What produced all those photons with such a perfect thermal spectrum? Why are they everywhere? And why is the temperature so cold?

- In our study of spectroscopy, we learned that blackbody radiation comes from optically thick materials in thermal equilibrium, the conditions in which the photons are constantly randomizing their energies through collisions, absorptions, and emissions by charged particles.

- But the universe isn't optically thick! The Earth isn't suspended in a fog; the night sky is black. The universe is transparent. Photons can travel for billions of years without hitting anything, straight from some distant galaxy to our telescopes. And the universe certainly isn't all at the same temperature; space is very cold, and stars are very hot.

- To make sense of the cosmic blackbody spectrum, we are led to the conclusion that the universe used to be optically thick—it used to be much denser. The existence of the CMB is another pillar of evidence supporting the big bang theory.

- At early times, the universe was like the interior of a star—just ions and electrons, everywhere, hot and dense enough to glow with blackbody radiation, like the inside of a kiln. But then, over time the universe expanded, the density dropped, and at some point the universe became transparent. All those photons were still there, but the chance of getting absorbed or scattered had become negligible.

- After that, the photons kept sailing along in straight lines, and they're still there, billions of years later. But, like all photons propagating across the expanding universe, their wavelengths have been stretched.

- A blackbody spectrum has an important mathematical property. If you start with a collection of photons whose wavelengths follow a blackbody spectrum with temperature T and then stretch the wavelengths of all the photons by the same factor, a, then the transformed collection of photons will still have a blackbody spectrum—but with a colder temperature: T/a.

- In other words, the expansion of the universe preserves the blackbody spectrum of the photons, even after there's no way to maintain thermal equilibrium—the photons aren't interacting with anything anymore. But the temperature of that blackbody spectrum drops in proportion to $1/a$.

- That's why the CMB has such a low temperature today. When the universe was like the inside of a star, it glowed at visible wavelengths. But since then, the universe has expanded, stretching the wavelengths from microns into millimeters—into the microwave region of the electromagnetic spectrum. Given that the CMB temperature is 2.7 Kelvin today and that it varies as $1/a$, in general the temperature is $2.7/a(t)$.

$$T(t) = \frac{2.7\,\mathrm{K}}{a(t)}$$

THE EPOCH OF RECOMBINATION

- If we were studying water vapor instead of the universe and allowed water vapor to expand and cool, we know that once it cools to 100° Celsius, it would condense into liquid, and if it kept cooling to 0°, it would freeze. The universe has 2 special temperatures, too, and when it cooled to those temperatures, important things happened.

- The most recent of these 2 events was the formation of the first atoms. At early times, the universe was too hot for atoms to exist. Electrons and ions existed separately, forming a plasma. If an electron did happen to combine with an ion to form an atom, it was quickly blasted apart by one of the countless high-energy photons flying around.

- But as the universe cooled, the photon energies dropped. At some point, the probability for an atom to get ionized by a photon dropped to nearly 0, which allowed the force of electrical attraction to bring electrons and ions together to combine permanently. This time period is called the epoch of recombination.

- This was also when the universe became transparent, because photons don't interact as strongly with neutral atoms as they do with electrons and ions. So, when we look at the CMB, we're seeing photons that have been traveling since the epoch of recombination.

The universe had to cool all the way down to 3000 Kelvin before atoms could form.

THE EPOCH OF NUCLEOSYNTHESIS

- When we turn back the clock starting from the epoch of recombination—with ions, electrons, and light everywhere—the scale factor gets smaller, and the density and temperature rise.

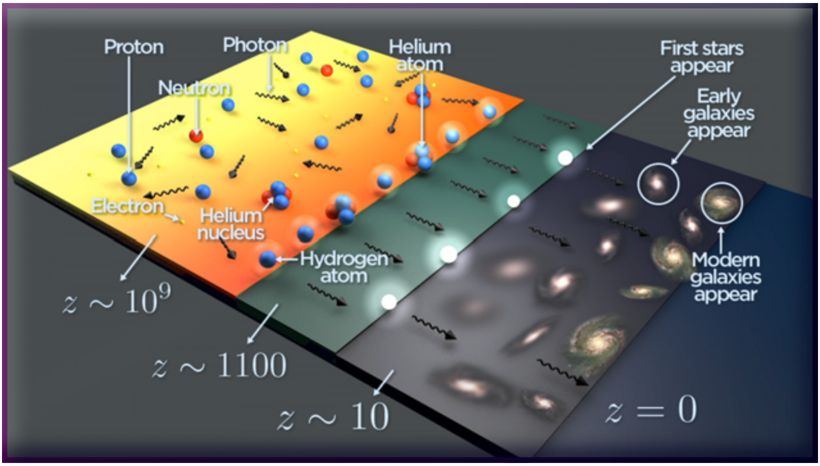
- Nuclei are no longer stable. If nucleons do come together to make helium or lithium, they get blasted apart by a gamma ray arriving soon afterward. The universe is a plasma of bare protons, neutrons, and electrons.

- This time period is called the epoch of nucleosynthesis. Only at the end of this epoch was it cool enough for nuclei heavier than hydrogen to form—and persist.

In 1925, Cecilia Payne (later Payne-Gaposchkin) correctly deduced that the stars are mostly made of hydrogen and helium. Today we know that this is because of the events that took place during the first half hour after the big bang: the epoch of nucleosynthesis.

- At the beginning of this epoch, the universe was a mixture of protons, neutrons, and electrons heated in a bath of gamma rays. As it cooled below a few billion Kelvin, the photons lost the power to disintegrate nuclei, freeing up the strong nuclear force to bind protons and neutrons together into heavier elements.

The temperature during the epoch of nucleosynthesis, when it was too hot for heavy elements, must have been billions of degrees.



- But as the universe expanded, the density of protons and neutrons decreased, making collisions less likely. And the neutrons started disappearing, because free neutrons spontaneously decay into protons, which repel each other, making it difficult to make nuclei. At this point, the abundances of different elements don't change any more—at least not until much later, when stars form and start fusing heavier elements at their cores.

- One of the triumphs of modern cosmology—a third pillar supporting the big bang theory—is that the results of these calculations match the measured abundances of the light elements hydrogen, helium, lithium, and their isotopes.

The theory of nucleosynthesis makes a bold prediction: The universe should be awash with neutrinos that were released by primordial nuclear reactions and have been sailing through the universe ever since.

In other words, just as there is a cosmic *microwave* background dating from the formation of the first atoms, there should be a cosmic *neutrino* background from the formation of the first nuclei.

The predicted energies of the cosmic neutrinos are so low that they seem nearly impossible to detect, but several experimental groups are working toward that goal.

READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chaps. 29 and 30.

Choudhuri, *Astrophysics for Physicists*, chap. 11.

Fleisch and Kregenow, *A Student's Guide to the Mathematics of Astronomy*, chap. 6.

Maoz, *Astrophysics in a Nutshell*, chaps. 8 and 9.

Ryden, *Introduction to Cosmology*.

Ryden and Peterson, *Foundations of Astrophysics*, chaps. 23 and 24.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

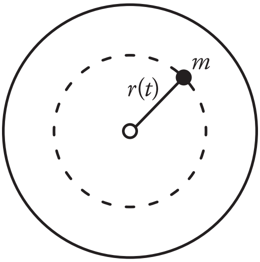
Lecture 24

THE HISTORY OF THE UNIVERSE

This lecture will attempt to provide an explanation of the entire history of the universe. History means what happened over time, so we need to add the dimension of time to our discussion. We've seen that a good way to describe the expanding universe is with the cosmological scale factor, $a(t)$. The distance between any pair of galaxies grows in proportion to a, but what equation determines how a varies with time? It's called the Friedmann equation, and this lecture will approach it in stages.

DESCRIBING THE UNIVERSE WITH CLASSICAL MECHANICS

- Let's assume that the entire universe is an enormous sphere of uniform density. It's only if we look closely that we see the specks of dust spread throughout its volume—those are galaxies. What determines how an arbitrary galaxy at a distance r from the center moves with time?



- The only force is gravity, and from Newton's theorem, the gravitational acceleration is directed inward with a magnitude of $\frac{GM_r}{r^2}$, where M_r is the total mass interior to r.

$$\text{accel.} = -\frac{GM_r}{r^2}$$

- This equation is the same one we solved during our study of planetary motion and black holes. This case is simpler, though, because there's just one variable, r, instead of r and θ. The trajectory is purely radial. So, in this model, the motion of the galaxy is the same as that of a spaceship near a black hole with no more fuel and no angular momentum.

- Even without solving the equation, we can guess what's going to happen. If the sphere starts from rest, the galaxy will fall inward. All the interior galaxies will fall, too, so M_r will remain constant as the sphere contracts.

- If the initial condition is a Hubble expansion—an expanding sphere with initial velocity proportional to distance—then gravity will slow it down. Whether a galaxy eventually gets pulled back or escapes to infinity depends on how initial speed compares to the escape velocity.

$$v_{\text{esc}} = \sqrt{\frac{2GM_r}{r^2}}$$

- We can visualize the possible outcomes with the graphical method, as we did for planets and black holes. We start by writing the total energy of the galaxy:

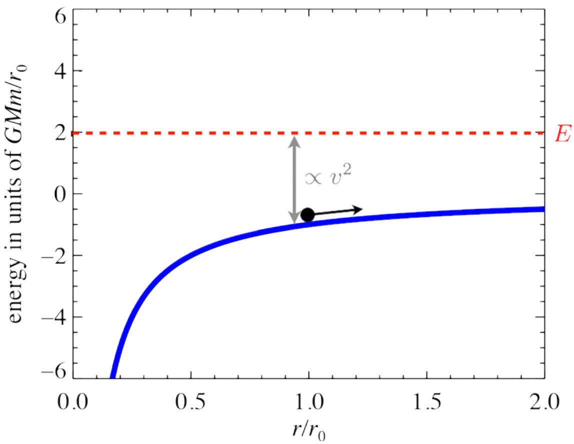
$$E = \frac{1}{2}mv^2 - \frac{GMm}{r}.$$

- We rearrange that to write the kinetic energy as the difference between the total energy and the potential energy:

$$\frac{1}{2}mv^2 = E - \left(-\frac{GMm}{r}\right).$$

- (Previously, we had another term, $L^2/2mr^2$, but in this case, $L = 0$.)

- Then, we sketch the potential energy as a function of r and make a horizontal line at the level E. The square of the speed at any location is proportional to the difference between the 2 lines: E minus the potential energy.

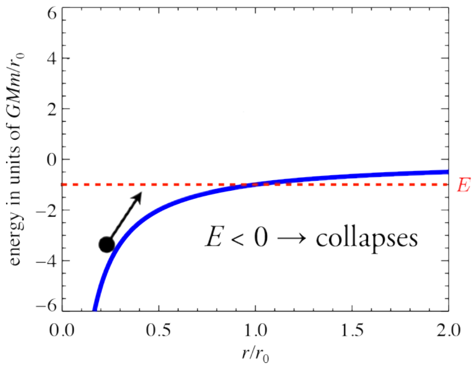


- Suppose E is positive and a galaxy starts with an initially outward speed. As the galaxy advances with time, the difference between E and the potential energy shrinks, so the galaxy slows down. As r goes to infinity, the potential energy becomes irrelevant and the speed approaches the square root of $2E/m$. That describes a universe that expands forever, coasting at a constant speed.

$$E > 0 \longrightarrow \lim_{t\to\infty} v(t) = \sqrt{\frac{2E}{m}}$$

expands forever

- If the energy is negative, then the lines cross at a certain point. In this case, the galaxy advances to that point, where it stops, turns around, and falls toward the origin. This would be a universe that expands for a while and then ends up collapsing into a black hole.



- In between these 2 cases is a special case when the total energy is exactly 0. That's like a spaceship with an initial speed exactly equal to the escape velocity. It corresponds to a universe that keeps expanding but at an ever-decreasing rate.

$$E = 0 \longrightarrow \lim_{t \to \infty} v(t) = 0$$

expands at ever-decreasing rate

- In this model, the fate of the universe depends on its total energy. If we could measure the total energy of the universe, we'd be able to determine its fate. There's a problem, though. On scales of gigaparsecs, we need to describe the universe with general relativity, not classical mechanics.

CALCULATING CRITICAL DENSITY AND THE AGE OF THE UNIVERSE

- To prepare for general relativity, let's dress up the energy equation in different clothing. First, let's divide through by m, converting everything to units of energy per unit mass. And because E/m is a constant, let's just call it k.

$$\frac{1}{2}mv^2 = E - \left(-\frac{GMm}{r}\right) \rightarrow \frac{1}{2}v^2 = \frac{E}{m} + \frac{GM}{r}$$

- Next, let's bring in the scale factor. Instead of $r(t)$, we'll write $a(t)\,r_0$, where r_0 is an arbitrarily chosen distance scale—for example, 100 megaparsecs, the distance from the Milky Way to the Coma galaxy cluster. With that, the velocity, dr/dt, becomes da/dt times r_0.

$$r(t) = a(t)\,r_0$$

$$v(t) = \frac{dr}{dt} = \frac{da}{dt}\,r_0$$

- Finally, instead of the enclosed mass, we'll write the equation in terms of the density of the universe, $\rho(t)$. We'll replace M_r with

$$\frac{4\pi(ar_0)^3}{3}\,\rho.$$

- We make all those substitutions and then tidy up by multiplying both sides by 2 and dividing by $(ar_0)^2$.

$$\frac{1}{2}\left(\frac{da}{dt}r_0\right)^2 = k + \frac{G}{ar_0}\,\frac{4\pi(ar_0)^3}{3}\rho$$

$$\left(\frac{1}{a}\frac{da}{dt}\right)^2 = \frac{2k}{a^2r_0^2} + \frac{8\pi G}{3}\rho$$

- The quantity on the left side, $(1/a)(da/dt)$, is the Hubble parameter, H. And with that change of notation, we have derived the classical Friedmann equation.

$$H^2 = \frac{2k}{a^2r_0^2} + \frac{8\pi G}{3}\rho$$

- In this new guise, the equation relates the cosmological scale factor and its time derivative to the overall density of the universe at any given time. This allows us to rephrase our statements about the fate of the universe in terms of its density.

- The critical case of 0 total energy corresponds to $k = 0$. In that case,

$$H^2 = \frac{8\pi G}{3}\rho.$$

- Another way to put it is that there's a critical density of

$$\rho_{\mathrm{c}} = \frac{3H^2}{8\pi G}.$$

- If the actual density equals the critical density, the universe expands forever at ever-decreasing speed. If the density is higher, the universe collapses. And if it's lower, the universe ends up coasting at constant speed.

- To figure out what's going to happen to our universe, we need to measure the density and compare it to the critical density. The current value of the critical density is

$$\frac{3H_0^2}{8\pi G},$$

  where H_0 is the Hubble constant, 70 kilometers per second per megaparsec. Plugging that in gives a critical density of 9×10^{-30} grams per cubic centimeter. A more interesting way to express that is 5.5 proton masses per cubic meter.

- That seems like a pretty low bar for the universe to jump over to achieve the critical density. But we need to remember that the universe is gigantic, and most of it is empty space. To measure the average density, we need to assess a representative volume of the universe that is large enough for entire galaxies to be like specks of dust.

- When astronomers did that throughout the 1980s and 1990s, they found that the universe does have an average density on the order of a few proton masses per cubic meter. Even the dark matter, it turns out, is very dilute. This remarkable result suggested that the universe is in that perfectly balanced state, with 0 total energy.

- But as the measurements got better, a density equal to the critical density was ruled out. The actual average density of matter is only 30% of the critical density.

$$\rho_0 = 0.30\,\rho_{c,0}$$

- Why should the density be of the same order of magnitude as the critical density but not quite equal?

- Many theorists suspected that the density really is equal to the critical density but the measurements were off—maybe astronomers were still missing a lot of the dark matter. Let's see where that logic leads.

- Let's solve the Friedmann equation and find $a(t)$ for the special case of $k = 0$.

$$H^2 = \left(\frac{1}{a}\frac{da}{dt}\right)^2 = \frac{8\pi G}{3}\rho$$

- Both a and ρ are functions of time, but they're linked by the fact that ρ is mass over volume, and because the total mass isn't changing, ρ must be proportional to $1/a^3$.

- This implies that $\left(\frac{1}{a}\frac{da}{dt}\right)^2 \propto \frac{1}{a^3}$, or, equivalently, $a\left(\frac{da}{dt}\right)^2$ is a constant.

- From there, we take the square root and then integrate to find that $a^{3/2} \propto t$, or $a \propto t^{2/3}$.

- We've just learned that the cosmological scale factor grows with time, but not at a constant rate; expansion at a constant rate would imply that a is proportional to t. Gravity decelerates the expansion, making it go as $t^{2/3}$ instead.

- We can write a as $\left(\frac{t}{t_0}\right)^{2/3}$.

- And we can calculate the value of t_0, the current age of the universe, based on the measured value of the Hubble constant.

- In general, $H = \frac{1}{a}\frac{da}{dt}$, which in this case is $\frac{1}{a} \cdot \frac{2}{3}\frac{t^{-1/3}}{t_0^{2/3}}$.
- When we take that derivative and evaluate it at time t_0, the left side is the Hubble constant, H_0.

$$H_0 = \frac{1}{1} \cdot \frac{2}{3}\frac{t_0^{-1/3}}{t_0^{2/3}} = \frac{2}{3t_0}$$

- This means that $t_0 = \frac{3}{2H_0} = 9.3 \times 10^9$ years.
- Plugging in 70 kilometers per second per megaparsec for H_0, the age of the universe comes out to be 9.3 billion years.
- But there's a problem: The Sun may be only 5 billion years old, but some other stars in our galaxy appear to be 13 billion years old. How could stars be older than the entire universe?
- These 2 issues—the density not quite equaling the critical density and getting the wrong age for the universe—are both artifacts of our oversimplified model. To model the universe correctly, we need to use general relativity. The relativistic version of the Friedmann equation solves these problems, but in a shocking and disturbing way.

RELATIVISTIC EFFECTS

In general relativity, when we work out the problem analogous to the expanding sphere of uniform density, we find 3 features that don't show up in the classical Friedmann equation.

1 The simplest new feature is an extra constant on the right side: $\Lambda/3$, where Λ is called the cosmological constant. It's a constant of integration that we get when deriving the Friedmann equation from Einstein's more general field equations.

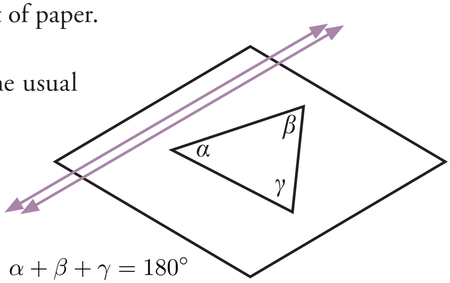
$$H^2 = \frac{2k}{a^2 r_0^2} + \frac{8\pi G}{3}\rho + \frac{\Lambda}{3}$$

2 A subtler change is that ρ isn't just mass over volume. Relativity teaches us that energy and mass are related, $E = mc^2$, so particles that are essentially pure energy, like photons and neutrinos, also affect the expansion of the universe. We have to understand ρ as the density of matter plus the energy density of all the radiation divided by c^2.

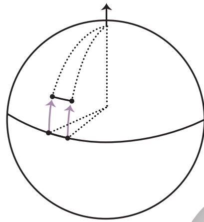
$$\rho \longrightarrow \rho + \frac{u}{c^2}$$

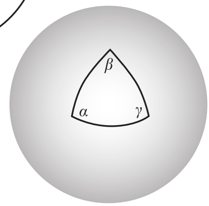
3 The subtlest change is that k, the constant representing energy per unit mass in our classical model, acquires a deeper interpretation. It specifies the curvature of space.

- If $k = 0$, space is flat, and geometry adheres to its normal principles (i.e., parallel lines never meet and the sum of the angles in a triangle is 180°). In a 2-dimensional universe, if $k = 0$, then the universe is like an endless flat sheet of paper.
- But if k isn't 0, space is curved, and the usual rules of geometry don't apply.

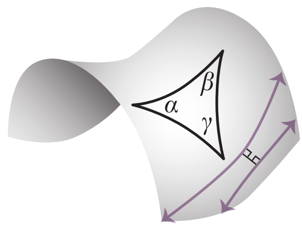


- If k is less than 0, the universe is curved like the surface of a sphere, where 2 lines that start out parallel at the equator eventually cross at the north pole and a triangle drawn by connecting 3 points with the shortest-possible paths has the sum of its angles as more than 180°.





- If k is greater than 0, the universe is curved like an infinite saddle, where parallel lines always diverge and triangles have angles that sum to less than 180°.



- If you were trapped on a giant, featureless surface, how could you tell if you were living on a flat sheet, a sphere, or a saddle? One way would be to draw a triangle and measure the angles to see if they add up to 180 or not.

- The real universe has 3 dimensions of space, making the curvature difficult or impossible to visualize, but the logic is the same. And this experiment has been done over the past few decades. Of course, nobody actually went around the universe in a rocket ship with a marker drawing triangles. The experiments are less direct; they're based on the cosmic microwave background radiation.

- From these experiments, the result is clear: k is equal to 0 within a few percent. Our universe is flat. This implies the universe has exactly the critical density.

- But this finding seems to contradict the fact that measurements of the total amount of matter in the universe on the largest scales imply that the density is only 30% of the critical density. From just those measurements, we would have expected k to be less than 1 and the universe to be curved like a saddle.

- Let's set aside that contradiction for now.

- If k really is 0, the Friedmann equation simplifies to $H^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3}$.

- Physically, that constant term, $\Lambda/3$, doesn't make sense. The best way to see that is to factor out $8\pi G$ so that the Λ term is being added directly to ρ.

$$H^2 = \frac{8\pi G}{3}\left(\rho + \frac{\Lambda}{8\pi G}\right)$$

- This makes clear that a constant value of Λ has the same effect in the equation as a density that is constant in time.

- This is crazy, because the density of anything should decrease as the universe expands. If a grows by a factor of 2, then the mass density decreases by a factor of 8. But the Λ term would stay the same, like a type of mass that can't be diluted—as if each new cubic centimeter of the expanding universe came into existence filled with new mass, or energy.

- It seems like Λ is a purely mathematical artifact. Any "reasonable" universe must have $\Lambda = 0$—in which case the Friedmann equation becomes the same as the classical equation.

$$\Lambda = 0 \longrightarrow H^2 = \frac{8\pi G}{3}\rho$$

- We're back to the same equation that we started with! It seems like none of the relativistic effects matter.

Einstein originally thought Λ was a negative number so that it would cancel out ρ and zero out the right side of the equation. But that's because he was working before Hubble. Einstein thought that the universe was stationary, so he wanted da/dt to equal 0.

Years later, after Einstein learned about the evidence for an expanding universe, he regretted this decision.

THE MYSTERY OF DARK ENERGY

- In the mid-1990s, astronomers started getting good at discovering Type Ia supernovas in distant galaxies. These act like standard candles: We can determine the distance based on the flux-luminosity relationship, and we can measure the redshift of the galaxy where the supernova took place. Each new supernova adds a new data point to the Hubble diagram—and astronomers were adding them at ever-greater distances and higher redshifts.

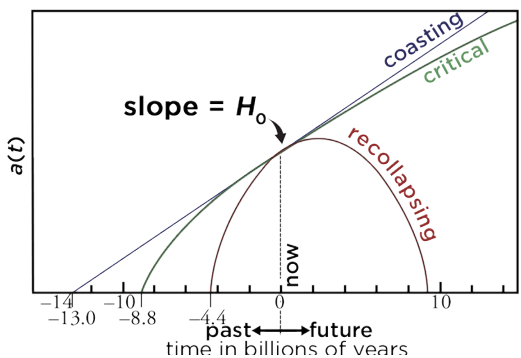
- The distance, d, divided by c tells us how much time has elapsed since the supernova went off. And the redshift, z, tells us the value of the cosmological scale factor at that time.

$$t = \frac{d}{c}$$

$$\frac{1}{a} = 1 + z$$

- So, we can convert the redshift-distance data into a chart of a versus t. Our solution of the Friedmann equation said that a should be growing like $t^{2/3}$. We can plot that curve.

- We can also plot some other cases. If gravity were irrelevant and the universe were just coasting along, we'd see a is proportional to t. And if the density were higher than the critical density, the universe would recollapse.



- However, none of these models turned out to fit supernova data; the data points for supernovas are higher than any of the curves.

- If we want connect the data points to the present day, when $a = 1$ and the slope is H_0, we need to draw a curve that bends upward. The scale factor is not just increasing; the rate of increase has grown with time. In other words, the expansion of the universe is accelerating.

- That seems absurd. It's easy to understand why the expansion rate might slow down, from the attraction of gravity. But to speed it up, you'd need some kind of antigravity! It doesn't make much sense, but that's where the data have led us.

- This is the top unsolved problem in astrophysics. Cosmologists, astronomers, and particle physicists have united in the effort to understand what's going on. The phenomenon—the force, or substance, that propels the expansion of the universe with ever-increasing speed—has become known as dark energy, in analogy with dark matter. But it's just a label; its true nature is unknown.

- What might have something to do with it is Λ, that integration constant we so casually discarded on the advice of Albert Einstein. So, Λ might have a physical meaning after all.

- Let's return to the Friedmann equation, but this time we'll retain Λ and see what happens.

- In an expanding universe, as time goes on, ρ decreases. Matter and radiation get diluted. But Λ, being constant, persists. Eventually, we reach a point at which we can neglect the ρ term altogether.

$$H^2 = \left(\frac{1}{a}\frac{da}{dt}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3}$$

$$\rho = 0 \longrightarrow \left(\frac{1}{a}\frac{da}{dt}\right)^2 = \frac{\Lambda}{3}$$

- This implies that $\frac{1}{a}\frac{da}{dt}$ is a constant—an equation for which the solution is exponential growth.

$$\frac{1}{a}\frac{da}{dt} = \sqrt{\frac{\Lambda}{3}}$$

$$a = e^{\sqrt{\Lambda/3}(t-t_0)}$$

- That could be what we're observing today: a transition in the history of the universe when the gravitational attraction of ordinary matter has been overcome by a universal repulsive force, represented by the cosmological constant.

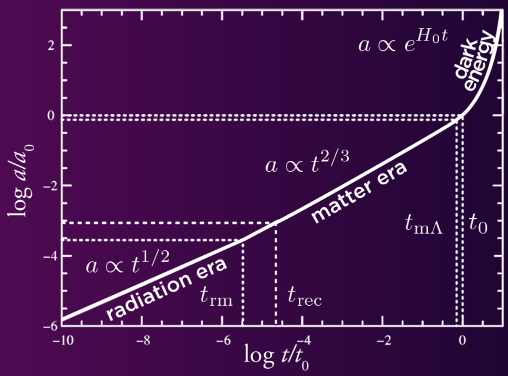
- Fitting the supernova data to an upwardly bending line also has the effect of increasing the age of the universe; the curve doesn't cross $a = 0$ until 14 billion years ago. So, the cosmological constant solves the problem of the stars that appeared to be older than the universe.

- It also explains how the universe can be flat even though the density of matter is less than the critical density. We previously found that in the Friedmann equation, $\Lambda/8\pi G$ acts like a density that gets added to ρ. According to the data, ρ is 30% of the critical density and the Λ term makes up the other 70%.

$$H^2 = \frac{8\pi G}{3}\left(\underset{\substack{\uparrow \\ \text{30\% of critical density}}}{\rho} + \underset{\substack{\uparrow \\ \text{70\% of critical density}}}{\frac{\Lambda}{8\pi G}}\right)$$

- Even though we don't understand dark energy, once we invoke it, everything fits together nicely.

This logarithmic chart of *a* versus *t* is what results when we solve the Friedmann equation, including the effects of matter, radiation, and dark energy—all at the numerical levels consistent with data from Type Ia supernovas, the cosmic microwave background, and numerous other sources.



READINGS

Carroll and Ostlie, *An Introduction to Modern Astrophysics*, chaps. 29 and 30.

Maoz, *Astrophysics in a Nutshell*, chaps. 8 and 9.

Ryden, *Introduction to Cosmology*.

Ryden and Peterson, *Foundations of Astrophysics*, chaps. 23 and 24.

Tyson, Strauss, and Gott, *Welcome to the Universe*.

QUIZ
LECTURES 19–24

1 What would be the effects on society if a nearby supernova occurred in modern times that was bright enough to be seen in the daytime? [LECTURE 19]

2 Suppose a carbon white dwarf accretes material from a companion until it reaches the Chandrasekhar mass, triggering a nuclear explosion in which all the carbon is converted to iron. How much energy is released? How does this compare with the gravitational binding energy of a typical white dwarf? Note: The atomic masses of carbon-12 and iron-56 are 12.0000 and 55.9349 times the mass of a proton. [LECTURE 19]

3 Does the detection of gravitational waves leave any room for doubt about the existence of black holes? What further proof would be needed? [LECTURE 20]

4 Check the latest news for discoveries from the LIGO and VIRGO projects. How many black hole mergers have been detected? How many neutron star mergers? Have there been any other types of detections, such as a merger between a black hole and a neutron star, or a supernova? [LECTURE 20]

5 Look up the following galaxies: Whirlpool, Triangulum, Messier 63, Messier 94, UGC 12591, the Cartwheel, the Sombrero, Centaurus A, NGC 3370, and NGC 4038. Which is your favorite? A good reference is the Astronomy Picture of the Day (https://apod.nasa.gov). [LECTURE 21]

6 The radius of influence of a black hole is defined as GM/σ^2, where M is the mass of the black hole and σ is the velocity dispersion of the bulge of stars within which the black hole resides. Why does this definition make sense? What is the radius of influence of a black hole with the mass of 10^8 suns in a bulge with σ = 160 km/s? [LECTURE 21]

7 Look up the Bullet cluster, which is often cited as evidence that dark matter is real and that the problem is not with the law of gravity. What is the basic argument, and is it convincing? [LECTURE 22]

8 For a galaxy with a flat rotation curve (i.e., a rotation velocity independent of distance), how does the mass density vary with radius? [LECTURE 22]

9 Try to think of some other explanation for the observation that the Earth is surrounded by a blackbody spectrum of photons with a temperature of 2.7 K. How might these other possibilities be tested or ruled out? [LECTURE 23]

10 Pair production refers to the process by which a photon can spontaneously transform into an electron and a positron. For this to occur, the photon must have an exceeding $2m_ec^2$, where m_e is the mass of an electron. By what factor has the universe expanded since it was hot enough for the average photon to undergo pair production? [LECTURE 23]

11 Prior to 1998, it seemed to many scientists that a flat, matter-dominated universe was the most beautiful possible cosmological model because of its simplicity. Can the modern cosmological model be considered beautiful? Should we expect human aesthetics to be a reliable guide to the fundamental properties of the universe? [LECTURE 24]

12 The energy density of radiation varies as $1/a^4$, rather than $1/a^3$. (The extra factor of $1/a$ is because the wavelengths are stretched in proportion to a.) Solve the Friedmann equation to find $a(t)$ for a flat universe dominated by radiation. What is the implied age of the universe, given the measured value of H_0? [LECTURE 24]

Go to page 339 for solutions.

QUIZ SOLUTIONS

LECTURES 1–6

1 44 km and 0.9 m/sec.

2 In general, your weight is proportional to M/R^2. Constant density implies R is proportional to $M^{1/3}$, and therefore your weight, W, is proportional to $M/(M^{2/3})$, or $M^{-1/3}$.

3 Answers will vary.

4 The final mass divided by the initial mass is 2^n, where n is the number of days. The ratio of the mass of the Milky Way to the mass of an electron is $r = 2.2 \times 10^{72}$. For this to equal 2^n, we need $n = \log_2(r) = \log(r)/\log(2) = 240$ days.

5 Professor Winn tried this and found his eyes to have an angular resolution of approximately 50 arc seconds, with corrective lenses.

6 The difference is that the star emits light while the asteroid reflects sunlight. The flux of sunlight reaching the asteroid varies as $1/r^2$, and the fraction of that light reaching the Earth varies approximately as $1/r^2$, giving a net dependence of $1/r^4$.

7 Answers will vary.

8 17.2 Earth masses.

9 The energy and angular momentum increase. The orbit becomes elliptical, with the location of the rocket burn becoming the distance of closest approach.

10 From Kepler's third law, the semimajor axis is 17.834 AU. The distance of closest approach is $a(1 - e) = 0.59$ AU.

11 Answers will vary.

12 P_{min} depends only on the density of the orbiting body. It is approximately 12.6 hours divided by the square root of density, expressed in g/cm^3.

Click here to go back to the quiz.

LECTURES 7–12

1 Answers will vary.

2 By setting the Roche radius equal to the Schwarzschild radius, the limiting black hole mass is found to be 320 million solar masses.

3 L is proportional to R^2T^4, giving about 25.

4 About 10 µm and 845 W for a typical human. Note, though, that your body also absorbs radiation from the ambient air.

5 Answers will vary.

6 Assuming the planet absorbs all the incident sunlight and radiates in all directions equally, the surface temperature is $(5777\text{ K})(R/2a)^{1/2}$. The corresponding habitable zone is from 0.56 to 1.04 AU. In reality, the atmosphere will lead to a higher surface temperature at a given orbital distance. The planet's reflectivity also plays a role.

7 Answers will vary.

8 The angular diameter of the stellar image is 1 arc second. The star's image is a circle of radius 0.80 square arc seconds, containing about 800 photons from the sky. Therefore, the signal is 300 and the noise is $\sqrt{1100}$, giving a signal-to-noise ratio of 9.

9 The atmosphere absorbs ultraviolet radiation. At 100 MHz, terrestrial interference from FM radio is severe.

10 Divide the shortest wavelength by twice the maximum baseline, giving 9.4 nanoradians, or 0.002 arc seconds.

11 The star would be limb-brightened, instead of limb-darkened, and would show an emission-line spectrum.

12 Venus, human body, incandescent lamp, oven, freezer, lava. Probably not the lightsaber, although Kylo Ren's lightsaber may be an exception.

Click here to go back to the quiz.

LECTURES 13–18

1 When we can track the motion of both stars, we learn their individual masses. When only one star is visible, we can only learn the total mass.

2 5.15 and 4.52 solar masses.

3 Answers will vary.

4 The transit probability is 0.65%. Transits would occur every 225 days and last a maximum of 11 hours. The Sun would appear to get fainter by 75 parts per million.

5 Answers will vary.

6 1.3×10^{19} kg/sec and 0.00021.

7 Lifetime is proportional to M/L. For a main-sequence star, L is proportional to M^3, so lifetime is proportional to $1/M^2$. Less massive stars live longer.

8 Evaluating the equation of hydrostatic balance for a sphere of uniform density gives an estimate of 2.7×10^{11} N/m^2. More sophisticated modeling gives a density of 3.6×10^{11} N/m^2.

9 Answers will vary.

10 0.47 solar masses, 0.013 solar radii, and 0.012 solar luminosities.

11 There are 2 reasons: The efficiency of burning decreases as the atomic mass of the fuel approaches that of iron, and the stellar luminosity rises during the later phases of evolution.

12 Answers will vary.

Click here to go back to the quiz.

LECTURES 19–24

1 Answers will vary.

2 The explosion releases 3×10^{44} joules, as compared to $3/5\ GM^2/R \sim 2 \times 10^{43}$ joules of binding energy. Hence, enough energy is released to blow the white dwarf into smithereens.

3 Answers will vary.

4 Answers will vary.

5 Answers will vary.

6 The radius of influence is where the velocity dispersion that would be produced solely by the black hole is comparable to the actual velocity dispersion of the surrounding stars. For the given parameters, it is 17 pc.

7 Answers will vary.

8 The density varies as $1/r^2$, assuming a spherical mass distribution.

9 Answers will vary.

10 In general, $T = 2.7$ K divided by a. We need $2.7kT$ to equal $2m_e c^2$, giving $T = 4.4 \times 10^9$ Kelvin, and $1/a = 1.6$ billion.

11 Answers will vary.

12 $a(t) = (t/t_0)^{1/2}$. The age of the universe is $t_0 = 1/2H_0 = 7$ billion years.

Click here to go back to the quiz.

IMPORTANT NUMERICAL VALUES

CONSTANTS, UNITS, AND LAWS

astronomical unit (AU)	1.496×10^8 km = 215.1 R_{sun}
Bohr radius (a_0)	5.29×10^{-11} m
Boltzmann constant (k)	1.381×10^{-23} J/kg
Coulomb's constant (η)	8.988×109 N m^2/C^2
electron mass (m_e)	9.11×10^{-31} kg
Hubble constant (H_0)	70 km/s/Mpc
Newton's gravitational constant (G)	6.673×10^{-11} m^3/kg/s^2
parsec (pc)	3.26 light-years = 206,265 AU
Planck's constant (h)	6.626×10^{-34} J s
proton mass (m_p)	1.67×10^{-27} kg
speed of light (c)	2.998×10^8 m/s
Stefan-Boltzmann constant (σ)	5.670×10^{-8} W/m^2/K^4
Wien's law	$\lambda_{max}T = 2.9$ mm K

OTHER VALUES

distance to the Andromeda Galaxy	778 kpc
distance to the center of the Milky Way Galaxy	8.0 kpc
distance to the Coma galaxy cluster	100 Mpc
effective temperature of the Sun	5777 K
luminosity of the Sun	3.826×10^{26} W
mass of the Earth	5.974×10^{24} kg
mass of the Sun	1.989×10^{30} kg
radius of a proton	0.88×10^{-15} m
radius of the Earth	6378 km
radius of the Sun	6.955×10^5 km

BIBLIOGRAPHY

Abbott, B., et al. "Observation of Gravitational Waves from a Binary Black Hole." *Physical Review Letters* 116 (2016): 061102. Reported the first detection of gravitational waves.

Bartusiak, M. *Einstein's Unfinished Symphony: The Story of a Gamble, Two Black Holes, and a New Age of Astronomy*. Yale University Press, 2017. A casual and friendly history of our understanding of gravitational waves and the effort to detect them.

Blundell, K. *Black Holes: A Very Short Introduction*. Oxford University Press, 2016. Concise account of general relativity and black holes for first-year physics students (i.e., expect some equations).

Carroll, B., and D. Ostlie. *An Introduction to Modern Astrophysics*. 2^{nd} ed. Addison-Wesley, 2006. A voluminous undergraduate textbook on astronomy and astrophysics.

Carter, J., et al. "Kepler-36: A Pair of Planets with Neighboring Orbits and Dissimilar Densities." *Science* 337 (2012): 556–559. The journal article reporting the discovery of an extraordinary transiting planet system with 2 planets located close to one another, leading to chaotic motions.

Choudhuri, A. *Astrophysics for Physicists*. Cambridge University Press, 2010. Pursues all the major topics of this course at the graduate level. Especially good on fluid physics and plasma physics, 2 very important topics that fell outside the scope of this course.

Chown, M. *Solar System: A Visual Exploration of All the Planets, Moons and Other Heavenly Bodies That Orbit Our Sun*. Black Dog & Leventhal Publishers, 2011. Beautiful photographs and technical illustrations are used to expose the basic properties of the solar system.

de Pater, I., and J. J. Lissauer. *Planetary Sciences*. Cambridge University Press, 2015. A standard text on planetary science, aimed at upper-level undergraduates or beginning graduate students.

Doyle, L., et al. “Kepler-16: A Transiting Circumbinary Planet.” *Science* 333, 6049 (2011): 1602–1606. The original journal article reporting the discovery of the first transiting circumbinary planet.

Fleisch, D., and J. Kregenow. *A Student’s Guide to the Mathematics of Astronomy*. Cambridge University Press, 2013. An elementary guide to some basic equations that occur throughout astronomy for students who are less comfortable with mathematics.

Levin, J. *Black Hole Blues and Other Songs from Outer Space*. Anchor, 2017. The story of the Laser Interferometer Gravitational-Wave Observatory (LIGO), based on interviews with the key personalities.

Maoz, D. *Astrophysics in a Nutshell*. 2nd ed. Princeton University Press, 2016. Excellent and concise textbook meant for physical science students with no prior exposure to astrophysics. Covers many of the same topics as this course, occasionally at a somewhat higher level of sophistication.

Perryman, M. *The Exoplanet Handbook*. Cambridge University Press, 2014. A concise and comprehensive overview of exoplanet-observing techniques and accomplishments. Written for the working astronomer, but some sections can be understood by anyone who has studied college physics.

Phillips, A. C. *The Physics of Stars*. 2nd ed. Wiley, 1999. Strongly physics-oriented treatment of stellar structure, stellar evolution, and stellar remnants. The author’s discussion of radiative diffusion was an inspiration for this course.

Prialnik, D. *Theory of Stellar Structure and Evolution*. 2nd ed. Cambridge University Press, 2009. Outstanding treatment of how stars work and how they change with time. The author’s treatment of the core’s-eye view of stellar evolution is especially insightful.

Ricker, G., et al. "Transiting Exoplanet Survey Satellite." *Journal of Astronomical Telescopes, Instruments, and Systems* 1 (2014): 014003. https://doi.org/10.1117/1.JATIS.1.1.014003. Overview of the NASA Transiting Exoplanet Survey Satellite (TESS) mission, launched in 2017, whose purpose is to study and discover thousands of new transiting exoplanets.

Rybicki, G. B., and A. P. Lightman. *Radiative Processes in Astrophysics*. Wiley-VCH, 2004. The standard graduate-level work on the subject.

Ryden, B. *Introduction to Cosmology*. 2nd ed. Cambridge University Press, 2016. Convivial and clear introduction to the subject.

Ryden, B., and B. Peterson. *Foundations of Astrophysics*. Addison-Wesley, 2010. A good college textbook on astrophysics as well as a source of inspiration in designing this course.

Scharf, C. *The Zoomable Universe: An Epic Tour through Cosmic Scale, from Almost Everything to Nearly Nothing*. Scientific American/Farrar, Straus and Giroux, 2017. A richly illustrated guide to the orders of magnitude of the universe.

Shapiro, S. L., and S. A. Teukolsky. *Black Holes, White Dwarfs, and Neutron Stars*. The standard graduate-level work on the subject.

Thorne, K. *Black Holes and Time Warps: Einstein's Outrageous Legacy*. Norton, 1995. A book for the general public about general relativity written by one of the winners of the Nobel Prize for the detection of gravitational waves. Though the field has changed quite a bit since 1995, this is still a great read.

Tyson, N., M. Strauss, and J. Gott. *Welcome to the Universe: An Astrophysical Tour*. Princeton University Press, 2016. Covers the same ground as this course, at a much less technical level and with many creative flourishes.

Winn, J., and D. Fabrycky. "The Occurrence and Architecture of Exoplanetary Systems." *Annual Reviews in Astronomy and Astrophysics*, 53 (2015): 409–447. https://doi.org/10.1146/annurev-astro-082214-122246. A review article written for astronomers. Discusses the shapes, spacings, and orientations of exoplanet orbits as well as planets in binary star systems.

INTERNET RESOURCES

Astronomy Picture of the Day. https://apod.nasa.gov/apod/astropix.html. Check this site daily and enjoy a continuous and casual astronomy education.

Eames Office. "Powers of Ten™ (1977)." Published on August 26, 2010. YouTube video, 9:00. https://www.youtube.com/watch?v=0fKBhvDjuy0. Classic video that inspired the first few lectures of this course.

Khan Academy. "Introduction to Logarithms." Khan Academy course. https://www.khanacademy.org/math/algebra2/exponential-and-logarithmic-functions/. Review of the key properties of logarithms and exponentials used throughout this course.

NASA Exoplanet Archive: A Service of NASA Exoplanet Science Institute. https://exoplanetarchive.ipac.caltech.edu/. The most comprehensive catalog of known exoplanets, including their sizes, orbits, distances from Earth, etc.

Obreschkow, Danail. "Cosmic Eye (Original HD Portrait Version 2011). Published on January 11, 2012. YouTube video, 3:09. Modern update of the classic "Powers of 10" video. https://itunes.apple.com/app/id519994935.

PhET. "Radiating Charge." https://phet.colorado.edu/en/simulation/radiating-charge. Interactive simulation of electromagnetic radiation from an accelerating charge.

Planetary Fact Sheets. http://nssdc.gsfc.nasa.gov/planetary/planetfact.html. This website presents lots of basic data for the planets in the solar system.

IMAGE CREDITS

main title	Vikrant Agarwal / EyeEm/Getty Images
lecture titles	©vi73777/iStock/Getty Images
5	©GeorgiosArt/iStock/Thinkstock
5	Google Earth
6	Google Earth
16	©Science Photo Library - CDC/ Brand X Pictures/Getty Images
16	©Steve Allen/Stockbyte/Getty Images
27	©Wavebreakmedia/iStock/Getty Images
28	And1mu/Wikimedia Commons/CC BY-SA 4.0
77	©OstapenkoOlena/iStock/Getty Images
78	©OstapenkoOlena/iStock/Getty Images
97	And1mu/Wikimedia Commons/CC BY-SA 4.0
108	Sch/Wikimedia Commons/CC BY-SA 3.0
112	NASA images courtesy Science Operations Center at the Johns Hopkins University Applied Physics Laboratory.
113	©MarcelC/iStock/Thinkstock
126	ESA/Hubble & NASA
143	David Broad/Wikimedia Commons/CC BY 3.0
146	©Glowimages/Getty Images
148	NASA/CXC
150	NASA
154	BélaBéla/Wikimedia Commons/CC bY-SA 4.0
157	Smithsonian Institution/flickr/Public Domain
158	NASA/STSci
158	NASA
159	NASA/STSci
163	NASA/SDO & the AIA, EVE, and HMI teams
203	©Astrobobo/iStock/Getty Images

223	Adam Block/Mount Lemmon SkyCenter/University of Arizona
248	ESA/Hubble, NASA
252	NASA/CXC/SAO
257	NASA/CXC/NCSU/M.Burkey et al
260	©MARK GARLICK/Science Photo Library/Getty Images
267	LIGO Lab/Caltech/MIT
272	LIGO Scientific Collaboration and Virgo Collaboration
275	ESA/Gaia/DPAC/CC BY-SA IGO
277	©blackphobos/iStock/Thinkstock
278	©Stocktrek Images/Thinkstock
279	©azstarman/iStock/Thinkstock
280	NASA, ESA, and the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration
288	NASA, ESA, and The Hubble Heritage Team (STSci/AURA)
289	NASA and The Hubble Heritage Team (STScI/AURA); Acknowledgment: Ray A. Lucas (STScI/AURA)
291	NASA, ESA and the Hubble Heritage Team (STScI/AURA)
292	NASA/CXC/M.Weiss
294	NASA/STScI
295	Charles Steidel (California Institute of Technology, Pasadena, CA) and NASA/ESA