Stellar and Galactic Astronomy Notes

Victor Andersen
University of Houston
vandersen@uh.edu

March 24, 2006

Contents

Chapter 1

Early Cosmological Ideas

Cosmology is the study of the universe as a whole. Questions that cosmologists try to answer about the universe include:

- How big is it?
- How old is it?
- What is its history?
- What is in it?
- What is our place in it?

Early Ideas about the Nature of the Universe

Mesoptomian and Babylonian Ideas

The ancient Mesopotamians and Babylonians believed that the divine and natural world were tied together. In particular they believed that many things that occurred in nature were due to the direct action of different gods. They also believed that the gods would attempt to communicate their desires to mankind through astronomical phenomena. This lead the Mesopotamians and Babylonians to make careful and regular observations of the sky, in order to be able to recognize any repeated patterns in the behavior of astronomical objects. Knowing what these repeated patterns were they could then search for unexpected occurrences (such as the appearance of a new comet for example), that might indicate something particular the gods were trying to communicate. In fact, these civilizations developed astrological systems to help them interpret such astronomical signs. Because they believed that the gods took a very active role in nature, the Mesopotamians and Babylonians never tried to develop natural models to explain the behavior of astronomical bodies.

The Ancient Greeks

The ancient Greeks believed phenomena in nature followed a set of divinely inspired natural laws. Thus, the Greeks took a two pronged approach to understanding nature; they looked to their religious beliefs, but also observed nature to determine what the underlying laws governing natural phenomena were. Because many ancient Greeks believed that mathematics was also divinely inspired, the laws to describe nature that they developed often took a mathematical form.

Cosmology is the study of the whole universe and its behavior. In their cosmology, the Greeks wanted to explain the motions of each type of object they could see in the universe, namely:

- The stars.
- The sun and moon.
- The planets = "wandering stars".

The Greeks believed that any successful cosmology should not only be able to explain the observed motions of the heavens, but also incorporate philosophical ideas, such as Plato and the Pythagoreans idea of spheres and circles being perfect shapes, etc. The ideas that formed the basis for the Greek picture that survived until the 16th century where put forth by Aristotle:

- Earth is at the center of the universe, and all celestial objects revolve around that center.
- The celestial objects are attached to crystalline spheres (or combinations of spheres), centered on the earth, that rotate at constant speed.

Galileo Galilei

Galileo Galilei was born in 1564 in Italy. Galileo made many important contributions to physics and astronomy, and can quite rightly be described as the first true modern scientist ("let data tell us the answer".) Galileo was a Copernican, but unlike Copernicus, he favored this model of the universe based solely on observational, and not philosophical grounds. Galileo was the first to use a telescope to make systematic observations of the heavens. With his telescope Galileo saw:

- Many faint stars (showing Galileo that there was more in the universe than accounted for by Aristotle.)
- Mountains and craters on the Moon (which to Galileo meant that the moon was another world, just like the earth, violating Aristotle's assumption that the earth and heavens where fundamentally different.)
- Sunspots, which appeared and disappeared with time (showing Galileo that celestial objects were neither perfect nor immutable.)
- Phases of Venus (which weren't expected if Venus revolves around the earth, but were naturally explained if both the earth and Venus were revolving around the sun.)
- Moons of Jupiter (showing Galileo that other centers of motion existed in the universe besides the earth, violating Aristotle's assertion that the earth lay at the one center of motion for all objects in the universe.)

Galileo published all this in *The Sidereal Messenger,* which caused the Catholic church to prohibited him from publicly discussing or holding Copernican ideas. After the death of the pope, Galileo approached the new pope, and convinced him to let him to publish "fair and balanced"–and hypothetical–look at the geocentric and heliocentric theories. This discussion was published in the *Dialog Concerning the Two Chief World Systems.* Three characters appeared in the dialogs, the important two being:

- Simplicio, who espoused the viewpoint of the pope, and
- Salviati (The hero) who demolished the viewpoints put forth by Simplicio, and instead argued persuasively for a Copernican explanation for the motions of the heavenly bodies.

Because of this, Galileo was tried by the inquisition, and forced to recant his Copernican beliefs.

Chapter 2

Modern Science

The practice of science is many things, but bloodless and coldly rational it is not. Nor, being an activity practiced by humans, can it be. –William Keel in *The Sky at Einstein's Feet*

Chapter 3

Matter and Radiation

Virtually all the information we have about distant stars and galaxies in the universe comes from the electromagnetic radiation that is emitted by these objects and that eventually reaches us here at Earth. Before you can understand how to use the information carried by this radiation, you must first understand the basic nature of matter, and how it emits and absorbs electromagnetic radiation.

The Nature of Light

In the seventeenth century, an important scientific question concerned the nature of light. There were two schools of thought as to what light might be. Some–including most famously Isaac Newton–believed that light was some sort of particle. Others believed that it was instead some sort of wave. In order to decide conclusively between these two viewpoints, it was necessary to devise an experiment, the outcome of which depended unambiguously on the different behavior of particles and waves. One complication in doing this is the high speed with which light travels, approximately 3×10^8 meters per second in a vacuum. In order to understand the experiment that was finally done to answer the question of the nature of light, you must first understand some of the basic properties of waves.

Waves

A wave is a traveling disturbance that transports energy, without transporting material. Waves have the following general properties:

- **Amplitude** A measure of the "height" of a wave. For a wave of a particular frequency, the amount of energy carried by the wave increases as the amplitude squared.
- **Wavelength** The distance separating two successive peaks or troughs in a wave. We usually denote the wavelength of a wave with the Greek letter lambda (λ).
- **Period** The time for two successive peaks or troughs (i.e. one whole wavelength) to pass a fixed position, usually denoted as T.
- **Frequency** The total number of waves to pass a fixed location per second, denoted as f. Notice that $f = 1/T$.
- **Speed** The speed of anything that travels is simply the distance traveled divided by the time it takes to travel that distance. Thus: for a wave $v = \lambda/T$, or, in terms of the frequency, $v = \lambda f$.

Young's Interference Experiment

Imagine having a room with a single window that is covered with some sort of opaque screen. If you made a small hole in the screen, and sunlight were to fall on it, you would get a small spot of light on the opposite wall of the room. Now, what would happen if you took a piece of card-stock, cut a thin strip from it, and used that to split the beam of light in two by placing the strip right in front of the hole? If light was a particle, you would expect to get a small spot of light on the opposite wall from each side of the hole. On the other hand, if light were a wave, each side of the hole would act like a separate light wave source, so you would expect a pattern on the far wall due to the interference of the light waves from each source; you should get alternating bands of light spots (where the interference was constructive) interspersed with dark bands (where the interference was destructive.)

This experiment (pretty much as I've described it here) was first performed by an English physician named Thomas Young. What he observed was a banded interference pattern. From this he (and other scientists as well) was forced to conclude that light was a wave phenomenon.

What Type of Wave is Light?

Although Young was able to show that light was a wave, his experiment couldn't determine what type of wave light was. It took about sixty more years before scientists were able to answer that question. Surprisingly, the answer to this question lies in the study of electricity and magnetism.

Electric Fields

Electric charges exert forces on other electric charges, with like charges repelling one another and unlike charges attracting. Notice that this is a rather mysterious effect, since the charges seem to be able to reach out through empty space to act on one another (this behavior is sometimes known as "action at a distance.") One way to explain how this "action at a distance" occurs is to imagine that each electrical charge produces a field around it, and that any charge within this electric field will experience a force, produced by the field. This electric field is then the way that one charge can act on another through empty space.

Magnetic Fields

Magnets have at least two poles, with two possible magnetic polarities, North and South seeking. Magnetic fields are created by moving electric charges (i.e. an electric current.) Electric charges that are stationary in a magnetic field experience no force due to that magnetic field. Moving charges may experience a force due to a magnetic field. Whether they do or don't depends on the direction of the particle's velocity vector relative to the direction of the magnetic field lines. If the velocity vector of the charged particle is parallel to the field lines, no force is experienced. If, on the other hand, the velocity vector cuts across the field lines, the particle will experience a force due to the field. The direction of the force that the particle experiences is perpendicular both to the direction of the magnetic field line and the velocity vector of the particle.

Therefore the magnetic force allows charged particles to move freely along magnetic field lines. Motion perpendicular to the field lines causes the particle to circle the field line instead. The combination of these two motions in general causes charged particles to spiral around magnetic field lines. Because of this, magnetic fields can serve as "traps" for charged particles, in which the particles spiral around the field lines, but cannot easily exit the region where the magnetic field is.

Electromagnetism

It turns out that there is another way besides those just mentioned to produce electric and magnetic fields. The British scientist Michael Faraday discovered if he produced a magnetic field that was changing with time, an electric field is produced by the changing magnetic field (most electric generators use this fact to produce electricity.) It turns out that an electric field that changes with time will produce a magnetic field. Therefore, although we often think of electricity and magnetism as two distinct things, they are in fact just two aspects of a single phenomenon. We refer to this single, unified phenomenon as electro-magnetism.

Electromagnetic Waves

Notice that electromagnetism gives rise to a possible feedback mechanism. A changing electric field will give rise to a magnetic field. But unless you have the electric field change with time in exactly the right way, the created magnetic field will be changing with time. This changing magnetic field will create an electric field. This electric field will probably be changing with time and so produce a magnetic field. This magnetic field can then produce an electric field, and so on. As I think you can see, it is possible under the correct conditions for this feedback to continue indefinitely.

The important point here is that if you somehow set up such oscillating fields they can be self supporting, and in fact will travel as an electromagnetic wave. Electromagnetic waves were first predicted by the British scientist named James Clerk Maxwell, who was the first person to write down all the formulas necessary to completely describe electromagnetism. After doing so, Maxwell was able to show that according to these formulas, electromagnetic waves should exist, and should travel through empty space at the speed of light. Maxwell realized that this result implied that light must be one example of an electromagnetic wave. It took about ten years from the time that Maxwell predicted electromagnetic waves for the German scientist Heinrich Hertz to produce and detect them in his laboratory.

Some Annoying Problems

Although Maxwell and Hertz had convincingly demonstrated that light was an electromagnetic wave, a few problems concerning the nature of these waves still existed.

1. **What is light a wave in?** Every other type of wave we know about in nature travels through something (a string, water, air, etc.) But even in the 1800's, scientists knew that the light from stars reached us from great distances, through a vacuum. To get around this problem, it was necessary to postulate that the universe was filled with a medium through which light traveled. This medium was call the luminiforous ether (or just the ether for short.) This was unsatisfactory to some physicists, because there was no other independent evidence for the existence of the ether.

2. **Why don't glowing bodies have the spectrum electromagnetism predicts?** Experiments showed that all hot solid bodies glowed with a similarly shaped spectrum, that depends only on the temperature of the body, but not what the body was composed of.[1] When theoretical physicists tried to calculate what the curve should be–based upon the known laws of thermodynamics and electromagnetism–they calculated a curve that matched the measured curve well at the long wavelength end of the spectrum. At the short wavelength end however, the calculated spectrum continued to rise, while the measured spectrum peaked and then dropped. The question then was whether the mis-match was due to something understandable in electromagnetism or thermodynamics that had simply been overlooked, or whether it was due to some yet to be discovered physical laws about how radiation and matter behaved.

Planck Finds an Answer to the Glowing Body Problem (That He Doesn't Like)

A number of physicists tried without success to derive the correct blackbody curve using just the laws of thermodynamics and electromagnetism. One of

[1] A body that is glowing under these conditions is known as a blackbody, and so the spectrum it produces is known as a blackbody spectrum.

these was a German physicist named Max Planck. Planck was able to find a formula that actually gave the correct shape for the blackbody curve, but was unable to derive it using the known laws of physics at the time, although he tried very hard to do so. In some sense, Planck was in the same position as a student who looks up the answer to a problem in the back of a textbook; he knew the answer, but couldn't explain *why* the answer was what it was (which after all is the most important thing in science!)

Finally, Planck was forced to conclude that some new law of physics must be invoked to explain why the blackbody curve was shaped as it was. As Planck describes it, "... the whole procedure was an act of despair because a theoretical interpretation had to be found at any price, no matter how high that might be." The new assumption Planck made that allowed him to actually derive the blackbody law was that on the microscopic scale, the material that was radiating could only have discrete amounts of energy (e.g. they could have energies of E, $2E$, $3E$, but not something like $1.356342E$), and that the energy and frequency of radiation emitted were related by the formula $E = hf$, where h is a constant. This property of having only allowed, discrete values is known as quantization. Initially, Planck was not at all happy about this quantization, and believed that the idea of such quantization was only a calculational trick that did not reflect reality. Planck worked very hard to find a way to calculate the blackbody curve without this assumption, but was unable to do.

Another Problem

About the same time as people were studying the nature of blackbody radiation, another phenomenon related to light was being explored, known as the photoelectric effect. When the surface of metals are illuminated, it is possible to observe electrons being ejected from the surface of the metal. This by itself is no surprise if light is a wave, since waves carry energy; the electron simply absorbs enough energy from the wave to break free from the surface. The problem was that electrons were only ejected for light above a certain frequency. If light below that frequency was used, no electrons would be ejected, even if the intensity (i.e. amplitude of the wave) was turned up to very high levels. Remember that the amount of energy carried by a wave depends both on the frequency of the wave as well as its amplitude, so that in the wave picture either a high amplitude or frequency for the wave should be enough to cause electrons to be ejected.

An explanation of this discrepancy was proposed by a young patent clerk–who was educated as a physicist–working in a Swiss patent office. The clerk proposed that the quanta that Planck used in his computation of the blackbody curve were not simply a calculational tool as Planck thought, but actually existed; a beam of electromagnetic radiation could be thought of as a beam of these quanta, each with an energy that related to the frequency of the beam as

$$E = hf$$

where h is a constant that today is known as Planck's constant. Notice that this allows a simple understating of the photoelectric effect; the ejection of electrons from the surface of the metal is due to the absorption of an individual quanta by an electron. If the quanta has high enough energy (i.e. frequency), then it will be able to eject an electron. If the frequency of the light (and thus the energy of the individual quanta) is to low, then no electrons will be ejected, just as observed.

The Swiss patent clerk in question was Albert Einstein, and this discovery–and not his theory of relativity–lead to him receiving the Nobel Prize in physics. As the telegram Einstein received notifying him that he had won the Nobel prize told him, he was awarded the prize "...for your photoelectric law and your work in the field of theoretical physics."

Light Behaves Like a Wave and a Particle

So, at this point you must be wondering, "is light a wave or a particle?" The answer turns out to be that the idea of a pure wave or pure particle is too simple to completely describe light's behavior, although there are some aspects of its behavior that mimic both waves and particles. In what follows, we will often treat light as if it is purely a wave or purely a particle, depending on which aspect gives us the best picture of light's behavior in the circumstance we are considering.

Matter on Small Scales

The Makeup of an Atom

There are three types of particles that make up an atom:

1. **Protons:** positively charged particles with a mass $m_p = 1.67 \times 10^{-27}\ kg$.

2. **Electrons:** particles with a negative charge exactly the same size as the proton's positive charge, with a mass approximately $(1/2000)m_p$.

3. **Neutrons:** particles with no electric charge, and with a mass approximately the same as that of a proton.

What distinguishes one chemical element from another is the number of protons in an atom of the element (this number is known as the atomic number of the element.) So, all hydrogen atoms have one proton, all helium atoms have two protons, all lithium atoms have three, and so on. Adding neutrons to an element does not change what element it is. For instance, most hydrogen atoms in the universe have no neutrons, although a small percentage have one neutron and one proton (hydrogen of this type is known as deuterium), while an even smaller percentage have two neutrons and one proton (known as tritium.)

The Structure of the Atom

The "Plum Pudding" Model

After the discovery of the electron (but before that of the proton and neutron), people realized that atoms must be made up of electrons (which have a small mass and a negative electric charge), and some part that contained most of the mass and had a positive electric charge. The question was, how do you put these two different parts together to make an atom? Because the positively charged bits had most of the mass, people naturally assumed that they were also physically much larger in size than an electron. This gave rise to what was referred to as the plum pudding model of the atom; most of the space being taken up by these large positively charged particles, with the small, negatively charged electrons embedded like 'plums' in this large 'pudding' of positive electric charge.

The Rutherford Gold Foil Experiment

In order to test whether the 'plum pudding' model was correct, and to gain more information about the structure of the atom, Ernest Rutherford (who was from New Zealand, although he was working in England) and his associates devised an experiment to shoot alpha particles[2] produced by radioactive decays at a very thin sheet of gold foil. The idea was to observe how much the beam of alpha particles was deflected by passing through the foil. From this pattern Rutherford and his co-workers could work backward to figure out how things are distributed in the atom, because the atoms in the foil are the things causing the deflections of the alpha particles.

What Rutherford and his collaborators expected to find (based upon the the fact that in the plum pudding model the mass in the atom was very spread out) was that the alpha particles would be deflected by a few degrees from the direction they started out in. Instead, the discovered something very shocking. Most alpha particles were deflected less than expected by the atoms in the foil, but a small number were actually deflected backward. The conclusion that Rutherford and his co-workers reached based on this was inescapable: the positive charge in the atom was not uniformly spread out like the plum pudding model assumed, but instead was concentrated in a very small, dense, positively charged nucleus.

The Bohr Atom

The results of the Rutherford experiment meant that the plum pudding model must be rejected. The large mass and small size of the positive part of the atom suggested a model analogous to what we see in the solar system, where the electrons would orbit the small positively charged nucleus, in the same way that the planets orbit the sun. The size of the atom is then determined by the size of the electron's orbit around the nucleus. There was only one problem with this model for the atom; electromagnetic theory predicted that such an atom would not be stable, but instead that the electron would lose energy by electromagnetic radiation, and in less than a second would spiral into the nucleus.

In order to rescue the 'solar system' type of model for the atom, a young Danish physicist named Niels Bohr suggested that unlike in the solar system where there is no restriction on the size and thus energy of an orbit, that in an atom only very restricted (and stable) orbits around the nucleus were allowed for the electron. Bohr worked out the consequences of such an atomic model, and found that it described very well the structure of the hydrogen

[2]We know today that an alpha particle is just a nucleus of helium; a combination of 2 protons and 2 neutrons.

atom. However, there were some serious drawbacks to Bohr's model for the atom:

1. Bohr had no justification (other than rescuing the 'solar system' model of course) for assuming that only certain orbits were allowed to electrons in atoms.

2. When people tried to apply Bohr's idea to atoms of anything heavier than hydrogen, it failed miserably in explaining the structure of those atoms.

DeBroglie Waves

A solution to the first of the problems with Bohr's atom was found by a French physicist, Louis-Victor de Broglie (pronounced *Duh-Broy*). De Broglie suggested that since light showed both wave and particle like behavior, that perhaps things we normally thought of as particles (like electrons for example) might also have wave-like properties. De Broglie worked out the consequences of this assumption, and discovered that the allowed energy levels for electrons in Bohr's picture of the hydrogen atom corresponded to orbits with total lengths that were a whole number of these 'electron waves' (i.e. the allowed orbits in the Bohr atom were those for which a whole number of waves could fit.) A few years after he proposed them, the reality of de Broglie waves was confirmed by experiments that showed under the right conditions, beams of electrons could be made to produce interference patterns.

This idea by de Broglie finished what Planck and Einstein started. For matter on the small scale it is too simplistic to think of objects as being purely a wave or a particle. All such phenomena in nature are a hybrid, that under the correct circumstances will behave in a way the is either wave-like or particle-like.

Wave-Particle Duality

What physicists discovered is that the naive conception of phenomenon being waves or particles breaks down on the small scale; *everything* on the subatomic scale acts like a wave in some circumstances, and a particle in others; this behavior is referred to as wave-particle duality. The meaningful question is thus not "is that a wave or a particle?" but instead "how do the wave-like properties of the thing we're talking about relate to the particle-like properties of that same thing?" The answer is that the amplitude of the wave at some point on the wave (electron, proton, photon, etc...) tells us about the probability that the particle is located at that point in space. We can no longer say with certainty that "the electron is at that point", but only describe the probability that the electron is found at different positions.

The Modern Picture of the Atom

With the knowledge that we can only know probabilities of finding particles at different locations, we must modify our conception of the atom a bit from that described by Bohr. Instead of exactly defined orbits for the electrons around the nucleus, we must instead think of some smeared out distribution of probable locations for the electron. This smeared out distribution is known as the electron cloud of the atom.

Chapter 4

Telescopes

Almost all objects astronomers study are distant and low in brightness. Telescopes allow astronomers to collect more light from distant objects, and also produce magnified images of those objects.

Reflection and Refraction

In some cases, it is possible to ignore both the wave-like and particle-like properties of light, and treat light as something that travels in a straight line. The straight line path that light follows is referred to as a light ray, and this approach to analyzing the behavior of light is known as geometric optics. In geometric optics, there are two processes that can redirect a light ray: reflection and refraction.

Reflection

At the surface separating one material from another, it is possible for light to be reflected. The direction of the reflected light ray bears a simple relationship to the direction of the incoming light ray; they both make equal angles to the perpendicular to the surface at the point the light ray is reflected from. This means that by shaping a reflecting surface in the proper way, it is possible to get all the light rays coming from a particular direction that strike the surface at different points to focus to a single point.

Refraction

When light passes from one material to another (from air to water for example), the direction of the light ray in the new material will not be the same as that in the old material. This bending at the interface is known as refraction. The size of the deflection depends on the speed that light travels through the two materials. By forming a piece of glass or plastic, it is possible to use the refraction of light within that material to focus light that strikes at different points on the surface at a single point. A piece of material that has been formed to act in this manner is known as a lens.

Reflecting and Refracting Telescopes

What distinguishes one lens or mirror from another is the distance from the lens or mirror at which parallel light rays are focused to a point. This distance is known as the **focal length** of the lens or mirror.

Refracting Telescopes

Telescopes that are made using lenses as the optical elements of the telescope are known as refracting telescopes. The simplest arrangement of lenses that makes a telescope is to have two lenses; a large lens on which the light falls known as the **objective lens**, and a smaller lens at the other end that you look into, known as the **eyepiece**.

Reflecting Telescopes

Arrangements of mirrors can also be used to construct a telescope; such telescopes are known as reflecting telescopes. Because mirrors reflect light back in the direction that the light was coming from, the place where light is focused by a refracting telescope tends to be back along the line from the telescope to the thing being observed. In order to not block the light you are trying to collect, various clever arrangements of mirrors have been devised to move the position of the image being formed out of the path of the incoming light rays–this means that there are several different designs for refracting telescopes. Despite this complication, essentially all telescopes used by professional astronomers for research are reflecting telescopes. This is because building large mirrors with good optical qualities is much easier than building large lenses with the same.

Reasons to Use a Telescope

Light Gathering Power

Imagine being assigned the task of catching rain water in a bucket, and being given several different buckets to choose from to do the job; some with larger openings on the top, others with smaller openings. If you wanted to catch the largest number of gallons of rain water, which bucket would you choose? Well of course, you'd choose the bucket with the largest area opening, since it would catch more rain drops than something with a smaller opening. The light gathering power of a telescope is analogous to the rain water gathering ability of a bucket; the largest area aperture (i.e. objective lens or primary mirror) will collect the most light. Therefore, in order to detect the faintest possible objects, you want to build a telescope with the largest possible diameter (since the area of circle is given by $A = \frac{\pi}{4}D^2$, where D is the diameter of a circle, the light gathering power goes up like the square of the diameter of the objective lens or primary mirror.)

Angular Resolution

Imagine being out in west Texas, at night, on a lonely road. Because you are in west Texas, the ground is flat, the road is long and straight, and there is little vegetation (or buildings for that matter) to block your line of sight. You spot the lights from a car in the distance. At first, the lights appear as a single dot, but as the car approaches, you are eventually able to see two distinct headlights. This effect, where you can discern two lights for nearby vehicles, but only one for more distant vehicles is due to **angular resolution**.

When your eye–or any optical device such as a telescope–has light enter it, the image that is formed is blurred by the interference of light waves from different parts of the aperture (this is known as diffraction.) The larger the size of the aperture, the smaller the blurred spot. Now think about having two small light sources near one another; each will be blurred by the diffraction of the aperture, and if close enough together, the blurred image of each light source will overlap. This is what occurs with the headlights of the car; only when the car gets close enough does the angle we observe between the two light sources get large enough that we can discern distinctly the two blurred images of the separate headlights.

How far apart do the sources have to be? The smallest angle between the sources that allows them to be easily resolved depends on both the diameter of the aperture (D) and the wavelength of light being used (λ), and and can be estimated using the formula

$$\theta \approx 250,000(\frac{\lambda}{D})(''),$$

where θ is the angle separating the two objects, in arc-seconds.[1] This angle is called the diffraction limit of the telescope. At a given wavelength, a larger diameter will give a smaller diffraction limit, and thus better angular resolution.

Magnification

Distant objects will appear to have very small angular sizes (i.e. the angle you measure from one side of the object to the other is small.) In order to be able to see more detail in such an object, it is desirable to produce an image with a larger angular size. This process is known as **magnification**. For a refracting telescope, the magnification depends on the ratio of the focal lengths of the objective and eyepiece. If the angular size of the object observed without the telescope is θ_O, and the angular size of the image produced by the telescope is θ_i, then the magnification M is given by

$$M = \frac{\theta_i}{\theta_O} = \frac{f_o}{f_e}$$

where f_o is the focal length of the objective, and f_e is that of the eyepiece. It is important to realize the ultimately, magnification cannot overcome the limitation imposed by resolution; it is not possible to reveal details in an object smaller than the resolution angle by using magnification.

[1] See the supplement on angular measures on the course website for the definition of an arc-second.

Chapter 5

The Laws of Motion and Gravitation

Physics Terms

If you haven't figured it out by now, scientists are picky about how we use terminology. We give specific definitions to terms, and then try to be careful to only use those terms in a way consistent with the definitions (i.e. we try not to be "sloppy" with our use of terminology.) Physicists are no different than other scientists in this regard; terms in physics have specific meanings, and in many cases where students get confused about things, the confusion is due to them not being careful when they read or use specific terms. To make matters worse, some terms that are interchangeable in every day language are not when you are using them in "physics speak". For example, the terms velocity and speed have very specific, somewhat different meanings in physics, so changing speed to velocity in a particular statement can change the meaning of that statement, possibly even changing the statement from a true statement to a false statement. The bottom line is, be careful with your terminology when talking to a physicist or astronomer (or taking one of their exams!)

To start off our discussion of physics, here are two terms that it is important to understand:

- **Vector:** a vector is a quantity that has both magnitude (size) and direction (note: don't confuse direction and position; they're different.) Examples of quantities that are vectors include velocity, force, acceleration, and momentum. When we write down formulas that include vectors, we denote all the vectors in the formula in bold face, in order to distinguish them from quantities that are just numbers (e.g. we write $\mathbf{F}$ for a force instead of F.)

- **Scalar:** A scalar is just a number. Examples of quantities that are scalars include mass, speed, and the amount of money in your bank account.

Quantities Related to Motion

In order to predict the motion of an object, we must start by defining quantities that help us keep track of an object's motion.

- **Velocity:** In physics, we define an object's velocity to be the change in position, divided by the time it takes for the position to change. We denote velocity as $\mathbf{v}$. Notice that the change in position tells us a direction, so that velocity is a vector.

- **Speed:** The size (i.e. the length) of the velocity vector is referred to as speed. We denote the speed as v. Notice that an object's speed is therefore a scalar, not a vector, since it doesn't include any direction information. When you are driving in a car, the speedometer tells you your speed, but not your velocity (because it doesn't give you information about the direction of travel.)

- **Acceleration:** Notice, if an object's velocity never changed, predicting its position in the future would be easy; just take the object's starting position, and use its velocity to project its new position after the amount of time you are interested in. In most cases, however, an object's velocity will change with time. We define the change in velocity divided by the time it took for the velocity to change the acceleration. We denote acceleration as $\mathbf{a}$.

Notice that there are two things that we can change about a vector, its length or its direction. This means that when something is accelerating, that either its speed, or its direction (or both) are changing.

Isaac Newton

Isaac Newton was born in England in 1642, and most certainly must be considered to be one of the few greatest physicists of all time. Newton had a large number of achievements in both mathematics and physics, including developing calculus, along with Leibniz. In addition, it is reasonable to consider him to be the father of "modern" physics, because of his statements of the laws motion and of gravity. Newton's Laws of Motion are:

- 1st Law (Law of inertia): Unless it is acted on by an unbalanced external force, an object in motion will stay in motion, traveling in a straight line at a constant speed, while an object at rest will stay at rest. (First stated by Galileo)
- 2nd Law (Law of force): $\mathbf{F} = m\mathbf{a}$.
- 3rd Law (Law of Action and Reaction): For every action there is an equal and opposite reaction.

If you know the forces acting on any body, you can use Newton's laws of motion to analyze the motion of that body.

Conservation Laws

- In physics when we say something is conserved, we mean that the total amount of that something doesn't change with time. This makes conservation laws particularly useful in physics, since we can analyze many problems simply by keeping track of how much stuff is there before and after something happens.
- An objects momentum is defined to be $\textbf{Momentum} = m\mathbf{v}$.
- Newton's 3rd law means that momentum is conserved for an object or group of objects any time there is no unbalanced external force acting the object(s). For example, two billiard balls will have the same amount of momentum right before and right after a collision. So, to analyze what happened as the result of the collision of the two balls, we add up how much momentum they had right before the collision, and set that equal to the total amount they must have right after the collision.

- Energy is also conserved, although it can be converted from one type of energy to another. Some types of energy are:
  - Kinetic energy, "energy of motion". If an object has a speed v and a mass m, then its kinetic energy will be given by $\frac{1}{2}mv^2$, as long as the objects speed is well below the speed of light..
  - Potential energy, which you can think of as the energy "stored in a force". For example, when an apple falls from a tree, it gains speed as it falls and thus kinetic energy. This energy is available because at the same time the apple is losing potential energy, in this case gravitational potential energy.
  - Heat, which is microscopic kinetic and potential energy.
  - Rest mass energy ($E = mc^2$.)

Angular Momentum

- Any body that is rotating or revolving has angular momentum. A body's angular momentum depends on two things, the speed with which the body is rotating or revolving, and how the mass of the body is distributed with respect to the axis of rotation or revolution. The further the mass is from the axis, the greater the angular momentum will be. Notice that angular momentum is *not the same as momentum!*

- As a consequence of Newton's second and third laws, angular momentum is a conserved quantity.

- Angular momentum is a vector, with the angular momentum vector pointing along the rotation axis. Notice because angular momentum is conserved, that both the size of a body's angular momentum and its direction will not change.

- A Torque, is a "twisting force, and relates to angular momentum in the same way force relates to momentum. So, in order to change a bodies angular momentum, you must apply a torque to that body.

Mass vs. Weight

In physics, mass and weight are not the same thing!

- Mass is an intrinsic property of a body that depends on how much inertia a body has (i.e. it is defined by how it shows up in $\mathbf{F} = m\mathbf{a}$.)
- Weight is the force of gravity acting on a body. On the moon, your weight would be less than it is on the earth (because the gravitational force at the surface of the moon due to the moon is less than the gravitational force at the surface of the earth due to the earth), but your mass would be the same (because your mass is an intrinsic property of you.)

Newton's Law of Universal Gravitation

Any object that has mass can produce a gravitational force, that can act on other masses. Newton realized that the law describing gravitation is a universal law (i.e. it applies no matter whether you are on Earth, or in the depths of space.) This was the final nail in Aristotle's idea that you *must* have separate rules for the earth and the heavens. For 2 masses, gravity:

- is attractive, directed from one mass to the other along the line separating the two masses.
- diminishes with square of the separation of the two masses.
- is proportional to the product of the two masses.
- As a formula we can write the strength of the force between masses m_1 and m_2, that are separated by a distance r as:

$$F = \frac{Gm_1m_2}{r^2}$$

  G in this formula is called the Gravitational Constant, and is "just a number".

Newton's Laws and Orbits

By using data about the positions of planets over time, Kepler was able to determine the laws that governed the orbits of planets around the sun. However, Kepler couldn't explain why the orbital laws were what they were. In addition, Kepler's laws applied only to things orbiting the sun, but couldn't be applied to things like the orbit of the moon around the earth, or Jupiter's moons around Jupiter. With his laws of motion and gravitation, Newton could determine the general laws for orbits, and show that Kepler's laws arose as a natural consequence of those laws.

1. **Orbits are ellipses** When he solved the equations arising from the laws of motion and the law of gravity, Newton discovered that any closed orbit of one object around another would be in the shape of an ellipse.

2. **Equal areas in equal times** The conservation of angular momentum means that things move faster on their orbits when close to the object they are orbiting, and slowly when they are further away. When worked out mathematically, this leads to equal areas in equal times.

3. $\mathbf{P^2 = a^3}$ To see how this arises, we'll solve the problem of an object in a circular orbit around another object (remember, all circles are ellipses, but most ellipses aren't circles; the answer we will get for the circle turns out to be the same as the one we would get if we worked the answer out for *any* ellipse, but it will take about 20 pages less mathematics to do the problem for a circle.) We must start with the laws of motion and the law of gravity.

   First, we have to figure out the direction of the force acting on the orbiting body. When we do that we will know the direction of the acceleration, since $\mathbf{F} = m\mathbf{a}$. The law of gravity says that the force between two objects always points from one to another, so the force on the orbiting body will always point to what is orbiting. Since the orbit is circular, the object being orbited must be at the center of the orbit[1], so, the gravitational force and thus the acceleration must point toward the center.

   Now that we know the direction of the acceleration, we can look at $F = ma$ to see how the size of F and a relate. In this case, the force is the force of gravity, so

$$F = \frac{Gm_1m_2}{r^2} = ma.$$

[1]Actually, both the planet and the Sun orbit the center of mass of the combination; because the Sun's mass is much greater than any of the planets in the solar system, the center of mass lies very close to the center of the Sun.

Now, lets go ahead and assign the object orbiting the mass m_1 (i.e. set $m = m_1$), and assign the thing being orbited m_2, which we'll rename as a capital M (so $m_2 = M$.) The formula above can then be written

$$\frac{GmM}{r^2} = ma.$$

Notice that there is an m on each side, so they will cancel out:

$$\frac{GM}{r^2} = a.$$

Next, we need to take care of the acceleration term. When an acceleration is pointed toward the center of a curved path, that acceleration is referred to as a centripetal (i.e. a "center seeking") acceleration. The size of a centripetal acceleration is $a = v^2/r$, so making the replacement for a in our formula makes it look like:

$$\frac{GM}{r^2} = \frac{v^2}{r}.$$

In order to make this look like Kepler's third law we need to do one more thing: get rid of the speed v and replace it with something to do with the orbital period P. This is actually pretty easy, since the period is the time to complete one orbit, a distance of $2\pi r$ (i.e. the circumference of a circle), so the speed–which is just distance divided by time–is $v = 2\pi r/P$. Replacing v with that in our formula turns it into:

$$\frac{GM}{r^2} = \frac{(\frac{2\pi r}{P})^2}{r} = \frac{4\pi^2 r}{P^2}.$$

OK, now we just solve for P^2, and the formula looks like:

$$P^2 = (\frac{4\pi^2}{GM})r^3.$$

Now for a circle, its semi-major axis a is just its radius, so that

$$P^2 = (\frac{4\pi^2}{GM})a^3.$$

Now, let's compare this to the way Kepler wrote the third law:

$$P^2 = (1\tfrac{AU^3}{yr^2})a^3$$

$$P^2 = (\tfrac{4\pi^2}{GM})a^3.$$

Apparently, $1AU^3/yr^2 = (4\pi^2/GM)$, (remember that here M is the mass of the sun.)

Chapter 6

More Useful Physics

Before we get on to talking about stars, there are a few more pieces of physics that you will need to know.

Temperature Scales

The temperature scales in everyday use are the Fahrenheit and Celsius scales. The Fahrenheit scale is set up so that the freezing point of water is $32^\circ\ F$ and the boiling point is $212^\circ\ \ F$; for the Celsius scale the freezing point is $0^\circ\ C$, and the boiling point is $100^\circ\ C$. Therefore to convert from Fahrenheit to Celsius

$$T(^\circ C) = \frac{5}{9}(T(^\circ F) - 32^\circ),$$

and from Celsius to Fahrenheit

$$T(^\circ F) = \frac{9}{5}T(^\circ C) + 32^\circ.$$

A body's temperature turns out to measure the average of energy of each molecule in that object. Because of this, a truly useful temperature scale for many uses in science is one where 0 on the scale corresponds to zero energy (note that neither of the Fahrenheit or Celsius scales fits the bill in this regard, since both have 0 arbitrarily defined). Such a temperature scale is known as an absolute scale, and zero on such a scale is known as **absolute zero**. The absolute scale we will use is the absolute scale corresponding to the Celsius scale, known as the Kelvin scale (so a 1 degree change on the

Kelvin scale equals a 1 degree change on the Celsius scale.) The conversion from Celsius to Kelvin is

$$T(K) = T(C) + 273.16.$$

Energy and Temperature

You probably have a pretty good intuitive idea about what energy is. The concept of energy is a particularly important one in physics, because energy is conserved in nature: that is, the total amount of energy in the universe remains constant, although the energy can (and does) change from one form to another. As you'll see as we go on in this class, energy can take a number of different forms.

Moving objects have energy associated with their motion, known as kinetic energy, the amount of which just depends on the speed of motion and mass of the object. Written as a formula, an object that is traveling well below the speed of light will have a kinetic energy is given by:

$$KE \;=\; \frac{1}{2}mv^2.$$

In a gas, the temperature of the gas is just related to the average kinetic energy of the molecules in the gas:

$$\frac{3}{2}kT = (\frac{1}{2}mv^2)_{average},$$

where $k = 1.38 \times 10^{-23}\ J/K$ is known as Boltzmann's constant. All the formulas we will discuss here are correct *only* if an absolute temperature i.e. Kelvin temperature) is used.

Pressure and the Ideal Gas Law

In a gas, there is a simple relationship between the pressure, density (in this case we define the density to be the number of gas molecules per unit volume), and temperature:

$$p = nkT,$$

where p is the pressure in the gas, and n is the density of the gas. Notice that this law says that if you increase the pressure in a gas, that either n must increase, T must increase, or perhaps both.

Temperature Differences and Energy Transport

If you take two objects at different temperatures, energy will be transported between them until they reach the same temperature. In nature there are three ways for this heat to flow between the objects:

- **Conduction**, where heat is transferred due to physical contact between bodies.
- **Radiation**, where warmer regions emit, and cooler regions absorb the energy in the form of electromagnetic radiation.[1]
- **Convection**, the mechanism that occurs when all else fails. Hot material rises, and mixes with cooler material raising its temperature. At the same time, cooler material sinks and mixes with hotter material at the level, reducing its temperature.

Radiation Laws for Blackbody Radiators

Before we go on, there are two useful radiation laws for objects that radiate a blackbody spectrum.

Wien's Law

Wien's law relates the temperature of the blackbody to the wavelength at which the blackbody curve peaks (i.e. the wavelength that the object is giving off the most radiation at.) It turns out that

$$\lambda_{peak}T = \ constant = 0.002898 \ meters \times Kelvin$$

Notice that this means that the temperature of anything that acts like a blackbody (like a star for example) can be determined just by finding the wavelength at which its spectrum peaks.

[1] Actually any object with a temperature above absolute zero will emit electromagnetic radiation; objects at lower temperature emit less than those at higher temperature. This means that if two regions are at a different temperature, the cooler will emit less than it absorbs from the warmer, and vice-versa.

The Stefan-Boltzmann Law

The other important radiation law for blackbodies relates the temperature of the body to the total amount of energy per second the body is losing from its surface by radiation. This law says that the amount of energy lost by every square meter of of the object's surface increases with temperature as temperature to the fourth power. Written as a formula, this means that the total amount of energy per second radiated by a black body (its luminosity) is given by

$$L = \sigma A T^4.$$

In this formula, L is the luminosity, A is the surface area of the body, and σ is a constant. Since stars are close to spherical, their surface is just that of a sphere, or $A = 4\pi R^2$.

The Doppler Effect

If a wave source is moving with respect to you, it is possible that you will observe a different wavelength for the waves that reach you than someone riding along on the wave source would see. This is known as the Doppler effect, or the Doppler shift.

- If the wave source is moving toward you, the source is constantly chasing after the waves it has already emitted. This means that you will see a smaller separation between wave crests than someone riding on the wave source, and thus a shorter wavelength. This is often referred to as a **blueshift** of the spectrum.
- If the wave source is moving away from you, the source is constantly moving away from the waves it has already emitted. This means that you will see a larger separation between wave crests than someone riding on the wave source, and thus a longer wavelength. This is often referred to as a **redshift** of the spectrum.
- Note that any motion toward or away from you will produce a Doppler shift, but any part of the sources motion that is not directly toward or away from you will not.[2]

[2] Actually this is true only if the emitting object has a speed well below the speed of light. For objects moving close to the speed of light there is also a transverse Doppler effect, but we won't worry about this effect in this class.

The size of the wavelength shift is just related to the speed of the source toward or away from you. If λ_e is the wavelength that the wave was emitted with, λ_o is the wavelength you observe the wave to have, c is the speed that the wave travels with (the speed of light in this case), and v_r is the velocity of the source toward or away from you (this is sometimes called the radial velocity, which is why it gets an r as a subscript), then (for objects moving well below the speed of light):

$$z = \frac{v_r}{c} = \frac{\lambda_o - \lambda_e}{\lambda_e}.$$

Chapter 7

The Basic Properties of Stars

Parallax and the Distance to Stars

Imagine observing a particular star from Earth, at times separated by 6 months. Because the earth moves approximately half way around the sun on its orbit in this time, you are looking the star from two slightly different directions, and thus its position will appear to shift slightly with respect to more distant objects. This shift is known as **parallax**. If we measure the size of the angle the star seems to have shifted by, and divide it in half,[1] we have what is known as the parallax angle of the star, which we'll denote as p. Notice that we can then construct a triangle with with a base that has a length that is the average radius of Earth's orbit around the Sun (1 astronomical unit), with the star at the apex of the triangle, and the angle between the legs of the triangle being the parallax angle. As long as the parallax angle is smaller than a few degrees, the length of the legs of the triangle (D) is related to the length of the base of the triangle (b) by $b \approx p \times D$,[2] which means that the distance to the star is given by $D = b/p$.

Astronomers normally measure parallax angles in arc-seconds. If we leave the angle in those units, and the Earth-Sun distance in AU, then because

[1] We divide by 2 so that the triangle we are constructing will have a base equal to 1 AU.

[2] When two things are approximately equal to one another we denote that with a $\approx$, when two numbers are roughly the same size (within a factor of 2, say) we denote that with a $\sim$.

$b = 1\ AU$,

$$D = \frac{1}{p}.$$

This choice of units for b and p defines a new unit of distance for D called a parsec (pc)–short for a parallax-arc-second. 1 parsec is a little over 3 light years; the distance to the nearest star to the sun–Proxima Centauri–is about 0.7 pc. A parsec is thus not a bad estimate for the average separation of stars in our galaxy.

The Magnitude System and Stellar Brightness and Luminosity

The total amount of energy per second radiated from the surface of a star is known as the luminosity of the star. The apparent brightness of a star depends on both its luminosity and distance; in general the brightness of a star–or any other source of radiation for that matter–will drop off as the distance to the star squared. Astronomers use a particular system to measure the luminosity and brightness of stars known as the magnitude system. The magnitude system has a number of peculiar features that take a bit of getting used to:

- The system runs backward, so that bright objects have smaller magnitudes (note that negative and fractional magnitudes are quite possible; -3.92 is a perfectly good magnitude for a star for example.)

- One step on the magnitude scale corresponds to 2.5 times in actual brightness (i.e. in energy per second.)

In astronomy we actually define two types of magnitude, one that corresponds to the luminosity of the star (known as the absolute magnitude, and denoted with a capital M), and one that corresponds to the brightness that the star appears to have at the distance that the star is actually at (known as the apparent magnitude of the star and denoted with a lower case m.) The relation between a stars absolute magnitude, apparent magnitude, and distance is given by the formula [3]

$$m - M = 5\log(\frac{D}{10\,pc}).$$

If we know what the apparent and absolute magnitude of a star is we can solve the above formula for D, and calculate its distance:

$$D = 10^{\frac{m-M+5}{5}}\ (pc).$$

Stellar Spectra

The spectrum of a star–and most other astronomical objects as well–can be used to determine many of the physical properties of the star, including its surface temperature, composition, and motion.

Types of Spectra in Nature

In nature there are three types of spectrum that occur:

- **Continuous Spectrum** In a continuous spectrum, all wavelengths in the electromagnetic spectrum are present, although not with equal intensities. One example of a continuous spectrum is the blackbody spectrum we have already discussed. A continuous spectrum is produced by a hot solid or hot, dense gas.
- **Emission Line Spectrum** Recall that in an atom, the electrons can only orbit the nucleus on orbits with certain energies; orbits with other energies are not allowed. Electrons can move between orbits of different energy, but must gain or loose exactly the energy difference between the two orbits, and thus must emit or absorb just that much energy in the form of a photon. This gives rise to a spectrum that is a set of lines at discrete wavelengths. A hot, low density gas will exhibit an emission line spectrum.
- **Absorption Line Spectrum** Imagine a situation where an object producing a continuous spectrum shines through a low density gas. The

[3] 10 pc appears in this formula because the definition of an object's absolute magnitude is the apparent magnitude the object would have if seen from a distance of 10 pc.

only way a photon from the continuous spectrum can be absorbed in the gas is if it has just the right energy to cause an electron in an atom of the gas to jump from its current energy level to a higher level. Absorption will not occur at all wavelengths in the continuous spectrum, but only at wavelengths that correspond to possible energy jumps for electrons in atoms of the gas. In this situation, you will observe the continuous spectrum from the continuous source, with light absorbed from it at the same wavelengths as the emission lines that the low density gas would produce. A typical star has an absorption line spectrum; an underlying blackbody spectrum, with overlying absorption lines.

Temperature from Wein's Law

Because stars have an absorption line spectrum, one way to figure out a star's surface temperature is to find the peak in its underlying blackbody spectrum, and use Wein's law to determine the temperature using the wavelength at which the peak occurs.

Spectral Types and Temperature

The strength of different absorption lines in a star's spectrum depends on both the surface temperature of the star and the amount of the element in the star that produces that line. Astronomers have developed a system of spectral types based on the relative strength of hydrogen lines in a star's spectrum. Since hydrogen is common in all stars, these spectral types give another way to determine the surface temperature of a star.

Spectral types are denoted by letters, the original system had the following types (arranged from hottest to coolest):

OBAFGKM.

With modern infrared detectors, we have discovered stars cooler than type M stars, and so the system has been recently extended:

OBAFGKM(LT).

Each spectral type is further divided into 10 subtypes denoted by the digits 0-9. For example, the sun has a spectral type G2, which is slightly cooler than stars of type G1, and slightly hotter than stars of type G3.

Composition from Spectra

Because different elements produce different patterns of emission or absorption lines, the relative amounts of different elements in the outer layers of a star can be determined by carefully analyzing the relative strengths of absorption lines in the stars spectrum. Analysis of the sun's absorption line spectrum by Cecelia Payne in the 1930's showed that the Sun is composed largely of hydrogen and helium (by number of atoms, 92 % and 7.8 % respectively), meaning that only 1.2 % of the material is heavier elements. This is a general result for all stars; the majority of the gas in their outer parts is hydrogen and helium , with in general only 0-2 % of heavy elements.

Motion and Rotation from the Doppler Effect

The absorption lines in the spectra of stars occur at known wavelengths. Often it is observed that in the spectra of actual stars that the lines occur at different wavelengths than those same lines in the Sun. This difference in wavelengths is due to the motion of a star producing a Doppler shift in the observed wavelengths of the lines.

Binary Stars and Stellar Masses and Diameters

Because mass is defined by the way it appears in the laws of motion and gravitation, we can use the orbits of stars in binary star systems to determine the mass of those stars.

Types of Binary Systems

Astronomers classify binary star systems based upon the way the system appears as seen from Earth.

- **Optical Doubles** Not really binary systems at all, optical doubles are stars that appear near one another as seen on the sky, but that are significantly different distances from the earth.
- **Visual Binaries** A binary system where you can separate the two stars of the system visually.

- **Eclipsing Binaries** Binary systems where the orbits of the two stars allow one to pass in front of another as seen from the earth, allowing one star to eclipse another.
- **Spectrum Binaries** Binary stars where the stars are to close together too separate them visually, but where a spectrum of the system shows two distinct sets of spectral lines, one for each star.
- **Single Line Binaries (Spectroscopic Binaries)** Sometimes one member of a binary system is too dim to be seen. Because both stars in the system orbit their common center of mass, the spectrum of the visible member of the pair shows a Doppler shift due to this orbital motion.

Getting Masses from Visual Binaries

The stars in a binary system orbit around their common center of mass, so for the two stars, the relationship between their masses and distances from the center of mass is given by:

$$\frac{M_1}{M_2} = \frac{D_2}{D_1},$$

where M_1 and M_2 are the masses of the two stars, and D_1 and D_2 are the distances of the stars from from the center of mass (note in this case that the M's are *not* being used to denote magnitudes.) Because they are orbiting one another, the stars must also obey Kepler's laws for orbits, so that

$$M_1 + M_2 = \frac{D^3}{P^2}$$

where D is the total separation of the stars (so $D = D_1 + D_2$), and P is the period of the orbit (i.e. the time for the stars to orbit each other once.) By measuring D_1, D_2, and P, we can then solve the above two equations for the masses of the two stars.

The Most Important Diagram in Astronomy: The Hertzsprung-Russell Diagram

A Hertzsprung-Russell diagram (or H–R diagram for short), is a graph showing the surface temperatures of stars on the horizontal axis, versus the their luminosities on the vertical axis. Notice that because we have multiple ways of measuring a star's temperature and luminosity, there are multiple ways such a diagram can be constructed:

- Instead of surface temperature, we can plot the color or spectral type of the star on the horizontal axis.
- Instead of luminosity, we can also plot absolute magnitude on the vertical axis.

Chapter 8

The Sun as a Star

The Sun's Energy Production

The Sun radiates a tremendous amount of energy from its surface every second, and has been doing so at roughly the same rate for the past 4.5 billion years. The source of this energy production remained mysterious for a very long time; it wasn't until the discovery of the nature of the atomic nucleus in the 1920's and 1930's that astrophysicists realized the the Sun's energy was being produced by the fusion of light nuclei to heavier nuclei in the core of the Sun.

Nuclear Fusion

The nuclei of all elements are made up of protons and neutrons, where the number of protons in a particular nucleus determines which element that nucleus is (e.g. Hydrogen contains one proton, Helium two protons, and so on.) Notice this means that there must be processes in nature that can stick together two nuclei to make a heavier nuclei (fusion), and to split a large nuclei into smaller nuclei (fission). Depending on the nucleus undergoing the fusion or fission process, energy may be released or absorbed during these processes.

Fundamental Forces

When boiled down to its essence, there are only 4 different mechanisms by which a particle in the universe can act on another. Physicists refer to these as the fundamental forces in nature:

- **Gravity** We've already talked a bit about gravitation. Any particle with mass can act by and be acted upon by the gravitational force. [1]
- **Electromagnetism** Any particle with an electric charge can act by and be acted upon through the electromagnetic force.
- **Strong Nuclear** This is the force that works between nuclear particles (i.e. protons and neutrons), and allows them to bind together to form nuclei.
- **Weak Nuclear** This force is responsible for certain types of radioactive decays.

These forces can be grouped together based upon the range of distances over which they act. Gravity and electromagnetism are examples of **long range forces**; although the strength of these forces gets weaker the further you get from the object producing them, they never drop to zero. On the other hand, the weak and strong nuclear forces are **short range forces**. They act over dimensions comparable to that of an atomic nucleus, and over larger distances quickly drop to zero.

The Proton-Proton Chain

In the core of the Sun, energy is produced through a multi-step process in which 4 hydrogen nuclei (i.e. 4 protons) are fused to form a single helium nucleus (which contains 2 protons and 2 neutrons). The mass of the neutron is slightly larger than that of the proton, so 4 free protons will have slightly less mass than 2 free protons and 2 free neutrons. Surprisingly then, the mass of a single helium nucleus is 0.7 % less than that of the 4 protons. This difference in mass is just the amount of energy it would take to break the helium nucleus back into free neutrons and protons (this conversion of mass to energy–or vice-versa–is possible because $E = mc^2$), and is referred to as the binding energy of the nucleus. That 0.7 % of mass is released as energy in the form of photons during the fusion reactions. This basic process is the ultimate power source for the Sun.

[1] Because mass is convertable to energy–and vice-versa–even massless particles such as photons produce and respond to gravitation. In the current day universe however, the mass in massive particles greatly exceeds that in massless particles, so we usually neglect the effect of massless particles.

Subatomic Particles

The atoms that makeup us, the earth, the sun, etc. are themselves made up of more fundamental particles. These particles have a number of basic properties that control how they behave. The particles that compose atoms (i.e. protons, neutrons, and electrons) are known as Fermions (named after the Italian physicist Enrico Fermi.) These particles can further broken into two sub-classes, baryons and leptons, which will discuss in a bit.

Particles and Anti-Particles

Each type of particle has a corresponding anti-particle that has:

- the same mass as the particle
- opposite electric charge as the particle (if the particle is neutral, so is the anti-particle.)

If a charged particle encounters its anti-particle, they will annihilate one another, creating 2 photons (2 photons are necessary in order to conserve mass/energy and momentum in the annihilation.)

Baryons

The particles that makeup the nucleus of the atom (i.e. protons and neutrons) are baryons, although there are many other more exotic baryons as well. All baryons are actually composed of three even more fundamental building blocks, known as quarks. Baryons respond both to the strong and the weak nuclear force, as well as gravitation (although gravitation is to weak on the scale of atoms or nuclei to be important), and if they have an electric charge to the electromagnetic force (which is strong enough to be important for atoms and nuclei.)

Leptons

The electron is an example of a lepton. There are only 12 leptons that are known to occur in nature. Leptons respond to the weak force but not the strong force. Leptons may be divided up into three families. Each family contains a charged lepton and its anti-particle, as well as a uncharged lepton known as a neutrino, and the anti-particle of that neutrino. Unlike baryons, leptons are not made up of quarks, or any other smaller units for that matter. Thus, the ultimate indivisible building blocks of the universe seem to be quarks and leptons. The table below summarizes the properties of the different leptons. The charge of each lepton is given in terms how many times the proton charge that lepton has.

Family	Electron	Muon	Tau lepton
Symbol	e^-	μ^-	τ^-
Charge	-1	-1	-1
Anti-Particle	positron	anti-muon	Anti-Tau
Symbol	e^+	μ^+	τ^+
Charge	$+1$	$+1$	$+1$
Neutrinos	electron neutrino	Mu neutrino	Tau neutrino
Symbol	ν_e	ν_μ	ν_τ
Charge	0	0	0
Anti-Neutrinos	electron neutrino	Mu neutrino	Tau neutrino
Symbol	$\overline{\nu_e}$	$\overline{\nu_\mu}$	$\overline{\nu_\tau}$
Charge	0	0	0

Gauge Bosons

In modern quantum theory, forces between all particles in the universe are viewed as being due to the exchange of particles between the objects acting on one another (notice that this eliminates the action at a distance problem that we encountered with the electromagnetic and gravitational forces...) In this picture, each of the four fundamental forces is carried by a particular type of particle that goes with that force. These force carrying particles are known as gauge bosons. As an example, the gauge boson that transmits the electromagnetic force is the photon (the symbol for a photon is the Greek letter gamma, γ.)

Conservation Laws for Subatomic Particles

In any interaction between two subatomic particles (collision, nuclear reaction, particle decay, etc...) there are certain physical quantities that must be conserved:

- **Electric charge** the total amount of electric charge before and after the interaction must be the same.

- **Baryon number** The baryon number is a way of adding up all the baryons and anti-baryons (i.e. baryon anti-particles) that participate in an interaction. To figure out the baryon number, add up all the baryons that participate as +1, and all the anti-baryons as −1. For example, if you had a reaction between two baryons and one anti-baryon, the total baryon number would be $+1 + 1 - 1 = +1$. Whatever particles that emerged from this interaction would therefore have a total baryon number of +1.
- **Lepton number** Leptons obey a conservation law similar to baryons, including using the same rule for adding in leptons versus anti-leptons.

The Steps in the Proton-Proton Chain

The most common way in which four protons can be fused to form one helium nucleus in the sun is a three step process known as the proton-proton (p-p) chain:

$$^1H + \ ^1H \longrightarrow \ ^2H + e^+ + \gamma + \nu_e$$

$$^2H + \ ^1H \longrightarrow \ ^3He + \gamma$$

$$^3He + \ ^3He \longrightarrow \ ^4He + \ ^1H + \ ^1H$$

Energy Transport in the Sun

Recall that there are three ways that nature transports energy due to temperature differences; conduction, radiation, and convection. In the interior of the sun and most other stars, only the last two are important. In the sun, radiation is efficient enough to transport energy in the central regions, while in the outer regions the energy transport is largely by convection.

Learning About the Solar Interior

The photons escaping from the surface of the sun tell us about conditions at the Sun's surface, however in order to completely understand the sun we need to know details about its interior. In order to get this information we must be clever, and use means other than just collecting photons from the Sun's outer layers.

Helioseismology

Waves travel through the interior of the sun, in the same way that seismic waves travel through the interior of the earth. The speed that these waves travel with depends on the temperature and density of the gas at the location within the sun that they are at. By observing the standing waves that exist at the surface of the sun, we can learn about the speeds they had as they traveled through different layers of the sun, and thus about the solar interior.

Solar Neutrinos

Because neutrinos have no electrical charge, and are leptons, the only fundamental force that gives them a chance to interact with normal matter (remember, for subatomic particles, gravitation is normally too weak to be important) is the weak force. Because the weak force is both weak and short range, it is particularly difficult to get neutrinos to interact with normal matter. This has two consequences: first, neutrinos are hard to detect (because we detect them when they interact with whatever we are using as a detector), and second, any neutrino produced in the sun's interior is likely to escape the sun without interacting with anything on the way out. Remember that in the proton-proton chain that neutrinos are produced. In fact there many other nuclear reactions that occur in the sun's interior that produce neutrinos as well. This means that we can directly get information about the nuclear reactions in the interior of the sun by studying the neutrinos that come from the core of the sun (provided we can build a detector that can detect enough of them of course!)

Over the past 30 or more years, measurements of the number of neutrinos being produced in the core of the sun have been made, with a surprising result: the number of neutrinos observed has been only about 1/3 the number predicted by the calculations of the nuclear reactions in the core. How could this be? One possibility is that our knowledge of nuclear physics is flawed, and thus the calculations are predicting the wrong number of neutrinos. Another possibility is that under certain circumstances a neutrino of one family could change into a neutrino of another family (for example, a ν_e could turn into a ν_μ or a ν_τ.) This possibility is referred to as neutrino oscillation. If neutrinos oscillate, then we would expect only 1/3 the number of electron neutrinos produced to be detected at the earth, because the other 2/3 would have oscillated to be ν_μ's or ν_τ's.

New neutrino detectors have been built, that have detected signs that neutrino oscillation does indeed occur. One consequence of neutrino oscillation is that neutrinos can't be massless like photons are, but instead must have at least a small mass.

Chapter 9

Stellar Evolution

Star Formation

Stars form due to the collapse of diffuse, low temperature clouds of interstellar gas that are composed largely of hydrogen and helium. This collapse normally proceeds through several distinct phases:

- The most dense parts of the cloud will collapse the fastest. Because the highest density portions of a cloud tend to be at the centers, clouds usually collapse from the inside out.

- Any rotation on the part of the cloud will cause it to rotate more rapidly as it collapses, due to the conservation of angular momentum. Since most clouds have at least a little rotation (if nothing else, the orbital motion of the cloud around the center of the galaxy will give the cloud a bit of rotation.) In addition to rotating more rapidly, the cloud will be able to easily collapse along its rotation axis (since the important thing in angular momentum is the distance of the mass from the rotation axis, collapsing parallel to the axis isn't impeded by angular momentum conservation), meaning that clouds will tend to collapse into disks.

- The increase in density in the core of the cloud will eventually cause the core to become opaque to radiation. After this occurs, the core will begin to rise in temperature significantly, since it can no longer cool itself efficiently by radiation.

- The increased temperature of the collapsing cloud core leads to a strong wind blowing outward in the cloud, which will stop the further collapse of the cloud.

Evolution Before the Main Sequence

The rapid gravitational collapse and opaque nature of the cloud core will greatly increase the temperature of the proto-star, due to the conversion of gravitational potential energy into heat. This will give the proto-star a high luminosity, although a fairly low (for a star) surface temperature, meaning that stars enter the H-R diagram on the upper right portion. As the star radiates away this energy, it will slowly contract and the core will heat up. Eventually the core of the star will reach high enough densities and temperatures for hydrogen fusion to become efficient enough to start producing energy, and the star will evolve downward and to the left on the H-R diagram until it reaches the main sequence. The line where stars of different mass join the main sequence is known as the zero age main sequence (ZAMS for short.) The time between when a proto-star first enters the H-R diagram and reaches the zero age main sequence depends on the mass of the proto-star; higher mass proto-stars evolve more quickly than less massive proto-stars.

Mass Limits for Stars

In nature there turns out to be a limit to how large or small the mass of a star can be. The low mass limit comes about because the lower the mass of a star, the less compressed its core will be, and thus the lower the density and temperature. Because the ability to achieve self-sustaining nuclear fusion requires a certain density and temperature, a gas cloud that has to low a mass will never be able to start steady fusion, because its core will never reach the necessary temperature or density. The lowest mass star turns out to have a mass of about $0.08 M_\odot$ (the symbol $M_\odot$ stands for the mass of the sun.) An object that has a mass below this limit is known as a **brown dwarf**.[1]

[1] In the system of spectral types for stars that we have already discussed, type L stars probably represent the least massive stars, while type T are brown dwarfs.

On the other hand, as stars have higher and higher masses, the fusion reactions in their cores become more and more efficient. As the mass of a star increases therefore, the outward flow of energy exerts a greater and greater pressure on the outer layers of the star, and if this flow of energy becomes to great, the outer layers of the star will be blown outward. Thus, there is a limit to the highest mass a star can have, because any star that exceeds that limit will produce energy too efficiently and blow away its outer layers, until its mass drops to the value where this no longer happens. The size of this mass limit for stars is less well known than for the lower mass limit, but is somewhere in the range of $100 - 200M_{\odot}$.

Energy Production in Main Sequence Stars

When a star is on the main sequence, it is processing hydrogen to helium in its core to produce energy; the majority of a star's energy producing lifetime is spent on the main sequence. Stars approximately the mass of the sun or less generate most of the energy via the proton-proton chain. In more massive stars, another chain of reactions is more efficient and generates the bulk of the energy. This chain is known as the CNO cycle, short for carbon, nitrogen, and oxygen because they participate in different steps in the chain. It is important to realize that overall carbon, nitrogen, and oxygen are neither created nor destroyed in the chain; the overall effect of the chain is to fuse produce 1 helium nucleus from 4 hydrogen nuclei. The steps in the most common form of the CNO cycle are:

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

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

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

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

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

$$^{15}N + \ ^{1}H \longrightarrow ^{12}C + \ ^{4}He$$

Evolution on the Main Sequence

As the hydrogen in the core of the star is consumed, helium builds up in the core of the star, requiring the core of the star to contract slightly to maintain its high pressure. This increases the core temperature slightly, thus speeding up the nuclear reactions slightly[2]. Because of this, a star on the main sequence leaves its starting position on the main sequence–the so called zero age main sequence or ZAMs for short–and slowly evolves on the main sequence to higher surface temperatures and luminosities. The total amount of time a star spends on the main sequence is known as the **main sequence lifetime** of the star, and can vary from millions of years for the most massive star, to 60 billion years or more for the least massive. Since the age of the universe is about 15 billion years, essentially all stars ever formed with masses $0.8M_{\odot}$ and less are still on the main sequence.

Evolution After the Main Sequence

As long as it is on the main sequence, a star uses the energy produced by hydrogen fusion in its core to counter act the effects if gravity and maintain a stable structure, neither contracting or expanding. Once the hydrogen in the core of the star is exhausted, it must either find a new source of energy, or change its structure. In fact, both processes occur. The change in the star's energy production rate and structure means that its observational properties (that is, its surface temperature and luminosity) must change, and so the star must move off the main sequence to different regions of the H-R diagram.

Shell Hydrogen Burning

With its hydrogen exhausted, the core of the star will begin to collapse and consequently to heat up. Eventually this temperature rise will heat the regions outside the core that are still hydrogen rich enough so that hydrogen fusion begins in a shell around the core (this process is, not surprisingly,

[2]Since the fusion process leaves the core of the star with less particles present, and because the density n in the core of the star is the number of particles per volume, the density in the core will drop. Because the pressure in the core of the star is given by the ideal gas law $P = nkT$, the drop in density means that the temperature must increase in order for the pressure not to drop

referred to as **shell hydrogen burning**.) During this phase, the star will actually produce more energy every second than it did on the main sequence, which will cause the outer parts of the star to expand significantly in radius and cool somewhat. The star will thus move to lower temperatures and higher luminosities on the H-R diagram and enter the red giant branch (or if it is high enough in mass the super-giant branch) on the H-R diagram.

As long as there is sufficient hydrogen in the shell, the outer parts of the star will be able to maintain their new equilibrium distribution as a red giant. Since the energy is being produced outside the core however, it will not keep the core of from collapsing further. In fact, a rather strange piece of physics having to do with the behavior of sub-atomic particles normally comes into play to halt the collapse of the star's core.

The Pauli Exclusion Principle and Degeneracy

Recall that the basic particles that makeup matter (i.e. protons, neutrons, and electrons) are all fermions. Beside their other basic properties such as mass, electric charge, etc, all subatomic particles have a property that is known as spin. For fermions, there are two values that a particle's spin can have, referred to as up or down. All fermions obey a law known as the **Pauli exclusion principle** (named after its discoverer, the Austrian physicist Wolfgang Pauli):

> In a group of fermions, No two fermions can share the same set of values for their basic properties (i.e. position, spin, energy, etc.)

As a consequence of this, it is difficult to pack fermions too closely together, since by doing so we are attempting to force the fermions to share positions, spins and energies with each other, which the exclusion principle forbids. Packing fermions close together therefore forces some of the fermions into states with very high energies; energy states with much higher energy than the fermion actually possesses. This effect is known as **degeneracy**. In essence, tightly packed fermions will behave has if they have much higher energies than they actually do, including having a pressure that greatly exceeds what you would expect based on their density and temperature.

In the collapsing core of a star, the densely packed electrons will provide this degeneracy pressure, keeping the core from collapsing further.

The Helium Flash

As shell hydrogen burning goes on, the helium produced in the shell settles onto the already helium rich degenerate core of the star. If this core gets massive and hot enough, helium fusion can begin, in what is known as the helium flash. Most of the energy released during the flash goes into causing the star to expand out of its degenerate state, and so the star does not significantly brighten due to the helium flash.

Stars more massive than about $M = 2M_\odot$ never develop a degenerate core. In these stars core helium burning starts in a more gentle way, and the star never experiences a helium flash.

Core Helium Burning

In this phase of the star's life, it is converting helium to carbon using a multiple step process known as the triple alpha process (because it takes 3 helium nuclei–also known as alpha particles–to create one carbon nucleus):

$$^{4}He + \,^{4}He \longrightarrow \,^{8}Be$$

$$^{4}He + \,^{8}Be \longrightarrow \,^{12}C + \gamma$$

During the core helium burning phase, the outer parts of the star shrink somewhat and at the same time heat up. The combined effect of lower radius and higher temperature tend to cancel each other out, meaning that the luminosity stays close to constant as the star traverses a nearly horizontal path on the H-R diagram known as the **horizontal branch**.

During this phase, most stars will cross a region of the H-R diagram known as the instability strip; when the star is in this strip, it is unstable to oscillation and becomes what is known as a pulsating variable star. The pulsations cause the star to vary in luminosity because the radius of the star is oscillating during this time. There are various classes of these pulsating variable stars, including two types that we will see later; Cepheid variable stars and RR Lyrae stars.

Shell Helium Burning

Eventually, in the same way as for hydrogen, the helium in the core of the star will be exhausted. For stars about the mass of the sun or greater, helium burning will now begin in a shell around the stellar core. Unlike in the case of hydrogen shell burning, the energy release in this phase is not steady, but instead goes through a number of thermal pulses. This leads to rapid mass loss from the surface of the star in the form of a stellar wind; the rate at which the mass is lost can be very high, up to 1/10,000 of a solar mass per year. During this phase of its evolution, the star's outer parts will cool just slightly, but the radius will increase significantly, causing the star to move almost straight upward on the H-R diagram. This phase of evolution is known as the **Asymptotic Giant Branch**, or AGB for short.

Final Stages of Evolution

The path a star will follow after this point depends on its mass.

Stars with $0.8M_\odot \leq M \leq 8M_\odot$

Beyond the helium burning phase, if a star's mass is less than about 8 $M_\odot$, its core will become degenerate, and will never reach temperatures high enough to initiate reactions to fuse heavier nuclei than carbon. Shell hydrogen and helium burning progress in shells closer and closer to the surface of the star, causing its outer envelope to swell even more. Eventually the the burning becomes unstable, and the outer part of the star will go through several pulses, that will lead to the outer envelope of the star being ejected. During this short lived phase–lasting only 50,000 years or so–the glowing envelope (heated by the ultraviolet and X ray emission from the still hot degenerate core) will be visible as a **planetary nebula**. Once this expanding shell has faded from view, the degenerate core remains as a **white dwarf**, which we will discuss more in the following chapter.

Stars with $M \geq 8M_\odot$

For stars with masses $\geq 8M_\odot$, their cores will eventually get hot enough to fuse elements beyond carbon. This can continue up until iron is created. Iron is the most strongly bound nucleus in nature; this means that although fusing iron with other nuclei to form heavier nuclei is possible, the reaction must *consume* energy and not release it. This is the end of the road for stars of high mass; as we will discuss in the next chapter, the ultimate fate of these stars is a titanic explosion known as a **supernova**.

Chapter 10

The Endpoints of Stellar Evolution

A star's evolution well after the main sequence, as well as the final endpoint in a its life depends upon the mass it had when it was on the main sequence.

Stars with $0.8M_{\odot} \leq M \leq 8M_{\odot}$

White Dwarfs

The degenerate core left over at the end of a planetary nebula is known as a white dwarf. White dwarfs have small radii but high surface temperature, and so appear in the lower left segment of the H-R diagram. No nuclear reactions are occurring in the white dwarf; the high surface temperature is a result of heat left over from earlier stages. The star will thus cool slowly as it evolves, since it is no longer producing energy to shine. The degeneracy of the white dwarf means that its average density is 3,000,000 times that of liquid water–300,000 times that of lead.

Once matter is degenerate, compressing it further causes the pressure to rise only slowly (since the electrons in the material are already forced into unnaturally high energy states, squeezing them together further affects their energy little.) This means that adding mass to a white dwarf has a peculiar effect; the white dwarf actually shrinks in radius. The smallest radius white dwarfs have the largest masses!

If enough mass is added to a white dwarf, the degeneracy pressure will no longer be able to support the over-lying mass, and it will collapse until some other physical effect acts to stop the collapse. The mass at which the white dwarf can no longer support itself is known as the Chandrasekhar limit (after the Indian astrophysicist who first calculated its value), and occurs at a mass of about $1.4M_{\odot}$.

Stars with $M \geq 8M_{\odot}$

In stars more massive than $8M_{\odot}$, the core will exceed the Chandrasekhar limit, and will collapse further. The collapse continues until the matter in the core is sufficiently compacted that a reaction known as **neutronization** occurs:

$$^{1}H + e^{-} \longrightarrow n + \nu_{e}.$$

This reaction releases an energy equal to the total produced by the star by nuclear fusion over its entire lifetime, within a few seconds. When the density of the now neutron rich core is high enough, the neutrons (which are fermions) become degenerate in the same way that the electrons in a white dwarf are, which brings the core collapse to a sudden halt. This creates a shock-wave that travels outward into the in-falling layers, halting the collapse. The large amount of mass in the layers of the star outside the degenerate core eventually cause the shock to lose enough energy that the shock stalls. The burst of neutrinos produced by the neutronization process deposit energy in the shock, causing the shock to blast away the outer layers of the star in a titanic explosion. This explosion is known as a type II supernova, and in the days and weeks after the explosion occurs will briefly become as luminous as an entire galaxy full of stars.

Neutron Stars

If conditions are right, after the explosion of a type II supernova, the degenerate neutron core of the star will be left behind; this remnant is known as a neutron star. Because of their higher mass, neutrons can be packed much more densely than electrons, so that the matter in the neutron star has a density approximately 100,000,000 times that of a white dwarf. In the same way as for white dwarfs, there is an upper limit to the mass of a neutron star of about 2.7 $M_{\odot}$, sometimes referred to as the Oppenheimer-Volkov limit. What happens if a neutron star exceeds this mass? It will collapse, and we know of no physical process that is sufficiently strong to stop the collapse; the neutron star must collapse to form a black hole in this case.

General Relativity

Black holes are objects that are predicted by Einstein's theory of general relativity; before you can understand the properties of black holes, you must understand a bit about this theory.

Black Holes

If matter becomes too dense, it must form a black hole. How dense is too dense? For a given amount of mass (M), it will be too dense if it is packed within a radius

$$r_s = \frac{2GM}{c^2},$$

where G is the gravitational constant, c is the speed of light, and r_s is known as the **Schwarzschild radius** of the black hole. Once matter is more dense than this, the laws of general relativity *require* that the matter collapse down to a single point, known as a **singularity**. The Schwarzschild radius defines how close you can get to the black hole; if you stay outside that radius, it is possible to escape the hole. If you enter within the Schwarzschild radius however, you can never escape; this applies to matter and light as well. How big is the Schwarzschild radius? For the Sun, the answer is about 3 kilometers (so something with a mass of 10 $M_\odot$ would have a Schwarzschild radius of 30 kilometers, and so on.) In general in terms of star masses measured in solar masses,

$$r_s = (3\ km)(\frac{M}{M_\odot}).$$

Chapter 11

Our Galaxy: The Milky Way

Early Attempts to Determine the Size and Shape of the Galaxy

Up until the renaissance, virtually all astronomical inquiry concerned the motions of objects on the sky. Only after the idea that stars might represent objects similar in properties to the Sun at great distance from the earth did serious exploration of the distribution of stars within space begin. It is important to note that the people performing these early observations did not realize that they were studying one of many galaxies in the universe, but the extent of the knowable universe itself.

- In 1610, Galileo observed the band of the Milky Way with a telescope, and found that that it was made up of many dim stars.

- Around 1800, William and Caroline Herschel used telescopic observations to count stars in different directions. From these counts, they concluded that the galaxy was a disk shaped structure with the Sun near the center.

- Around 1910, Jacobus Kapteyn tried estimating the distances to stars in different directions. Using these estimates, he concluded that the galaxy was a disk about 2 kilo-parsec[1] (kpc) thick and 10 kpc in diameter, with the Sun lying about 2 kpc from the center.

[1]kilo = 1,000, so 1 kpc = 1,000 pc.

- Today our best estimates are that the disk of the galaxy is somewhere between 25-50 kpc in diameter, with the Sun 8 kpc from the center. The early estimates were erroneous because they did not correctly account for the obscuring effects of dust in the disk of the galaxy.

Dust in the Galaxy

- Dust absorbs starlight, absorbing better at the blue end of the spectrum than at the red end of the spectrum.
- Dust also scatters starlight (i.e. deflects the light off in a different direction.) Again, this happens more efficiently at the blue end of the spectrum than at the red end.
- The presence of dust between the earth and a particular star thus makes that star appear dimmer and redder than it would in the absence of dust.
- The absorption of starlight heats the dust in our galaxy to temperatures in the range 20-40 K. Using Wein's law, you can calculate that the wavelength where most of the radiation is emitted by the dust is in the far infrared part of the electromagnetic spectrum.

Shapley Uses Globular Clusters

In order to determine the true size and extent of the galaxy, it was necessary to find some way of avoiding the obscuring effects of the dust within the plane of the galaxy. An American astronomer name Harlow Shapley noticed that a particular type of star cluster, known as a globular cluster, was distributed widely across the sky and not confined to the galactic plane, and thus did not suffer the effects of dust like stars and clusters in the plane. Further, Shapley noticed that the number of globular clusters was not the same in all directions on the sky, but that there were more globular clusters in the direction of Sagittarius than in other directions. Shapley reasoned that the globular clusters most be a part of the galaxy, and thus that the grouping of clusters in Sagittarius must be in the direction of the galactic center. Shapley therefore figured that if he could find the distances to the globular clusters, he could use those distances to find the distance of the sun from the center of the galaxy.

In order to find the distances to the globular clusters, Shapley used a type of variable star that occurs in globular cluster known as an RR Lyrae star (discussed more below.) When he did this, Shapley found that the Sun was not near the center of the galaxy, and that the galaxy was much larger than the determinations made using stars in the galactic disk. In fact, Shapley's estimate for the position of and distance to the center of the galaxy was not terribly different from modern estimates.

RR Lyrae Stars

RR Lyrae stars are stars with masses somwhat less than that of the Sun that have left the main sequence and are on the horizontal branch, passing through the instability strip on the H-R diagram. All RR Lyrae stars have about the same absolute magnitude of 0.6. Furthermore, RR Lyrae stars can be easily identified because their variations in brightness follow a distinctive pattern. Because of this, you can get the distance to an RR Lyrae by measuring its apparent magnitude, and using the formula you get by solving the equation for apparent and absolute magnitudes for the distance:

$$D = 10^{\frac{m-M+5}{5}}\ (pc).$$

Star Clusters

Star clusters are tight groupings of stars that are bound together by their mutual gravitation, meaning that the stars in the cluster are born, live, and die together. There are two basic types of star clusters within our galaxy; globular clusters and open clusters (also sometimes called galactic star clusters, because they are found only along the band of the Milky Way.) Globular star clusters:

- Have stars with relatively low contents of elements heavier than hydrogen or helium (astronomers often refer to all the elements heavier than hydrogen and helium as metals, so globular clusters stars are often called metal poor stars.)
- Have main sequences that end at fairly low mass stars on their upper ends, indicating that the clusters are fairly old.

- are not confined to the disk of the galaxy as open clusters are.

On the other hand, open star clusters:

- Have stars that are metal rich, compared to those in the globular clusters.
- Have main sequences that end at mid to high mass stars at their upper end, indicating that the open star clusters are middle aged to young clusters.

The Components of the Milky Way

Our galaxy can be divided into several distinct components, that have different structures, and different stellar and gaseous contents.

The Disk

The most obvious structure of our galaxy is a flat disk of stars, with a diameter somewhere in the range 25-50 kpc. The clouds of cool gas and dust within our galaxy are generally confined to this disk. The stars in the disk range in age from young to old, and in metal content from low to high. Star formation is an ongoing process in the galaxy's disk.

The Bulge

The other obvious component of the galaxy beside the disk is the central bulge. Unlike the stars of the disk which follow circular or close to circular orbits around the center of the galaxy, the stars of the bulge follow elongated orbits that rise well above the plane of the galaxy's disk. These orbits give the bulge of the galaxy its spheroidal shape. The stars of the bulge tend to be of moderate to old age.

The Nucleus

The inner few parsecs of our galaxy is a very dense aggregation of stars and gas. The high extinction due to dust in the galactic disk between the sun and the galactic center means that it is unobservable in the visual part of the spectrum, and we must instead use wavelength bands that suffer less from dust extinction to observe the nucleus, such as the radio, x-ray, and infra-red. The density of stars in the galactic nucleus is extremely high; the average separation of stars in the galactic nucleus is about 100 AU (i.e. the size of our solar system.) The orbits of stars and gas around the very center of the nucleus suggest that there is a mass of at least 2 million times the mass of the sun within the central 4000 AU or so. This large concentration of mass isn't stars or gas (since we would observe their electromagnetic radiation); in fact the most likely candidate for what the mass is is a giant black hole. There is also a bright and compact radio source at the galactic center, which is presumably emission due to material spiraling into the black hole.

The Halo

The globular clusters are one part of a low density spherical distribution of stars that extend well beyond the disk of the galaxy. The stars of the halo are old, and generally have very low contents of metals.

Chapter 12

Other Galaxies

It turns out that our own Milky Way galaxy is only one of many billions of galaxies within the observable universe. Galaxies span a wide range of shapes, masses, and other properties.

The Nature of the Extra-galactic Nebulae

In the early 1900's, observations of the sky showed many fuzzy blobs (i.e. nebulae) that seemed to avoid the plane of the Milky Way. There were two ideas about what these extra-galactic[1] nebulae might be:

1. Nebulae similar to HII regions or reflection nebulae that are part of the Milky Way.

2. Collections of stars, gas, and dust similar to the Milky Way itself.

Notice if the first were true, the distance to the nebulae would be in the parsec to kilo-parsec range. If they were truly similar to the Milky Way, their small apparent size and brightness meant that they must be at substantially greater distances. In order to resolve the question of the nature of the extra-galactic nebulae, it was necessary to be able to measure distances to them.

[1]extra-galactic in this case referred to the tendency of these objects be found above and below the disk of the Milky Way, and was not meant to imply that the objects weren't part of the Milky Way itself.

Cepheid Variable Stars

If the extra-galactic nebulae were truly at great distances, how could those distances be measured? RR Lyrae stars could have been used in principle, but in practice were not luminous enough to be detected at distances of Mega-parsecs or greater[2]. A much more luminous cousin to the RR Lyrae star are the **Cepheid variable stars**. Unlike RR Lyrae's, Cepheids have a wide range of luminosities and thus absolute magnitudes. Fortunately, the period for the luminosity variations of a Cepheid depend on its luminosity. So, to measure a distance using a Cepheid, we first measure its period and use that to compute its luminosity (or absolute magnitude), and then use that along with its apparent magnitude to compute its distance.

Edwin Hubble and his collaborators discovered Cepheid variable stars in several extra-galactic nebulae, and measured distances to these Cepheids that were much greater than the size of the Milky Way, showing that the extra-galactic nebulae were indeed other galaxies.

The Hubble Law

Before the nature of the extra-galactic nebulae had been resolved, spectra of several of them had been measured by Vesto Slipher of the Lowell Observatory. Slipher found that the great majority of the extra-galactic nebulae showed red-shifted spectra, and thus must be receding from the Milky Way. After measuring the distances to a number of galaxies using Cepheid variable stars, Hubble compared the distances to these galaxies to their radial speeds (obtained from the Doppler shift of emission or absorption lines in the spectrum of the galaxies). When he graphed their distance versus their radial speeds, there turned out to be a linear relationship between the two quantities; the more distant galaxies were receding at higher speed. This relationship is today known as **the Hubble Law**.

Recall that for a straight-line graph of x vs. y, the graph can be represented by the formula

$$y = mx + b$$

where m is the slope of the line, and b is the intercept (i.e. the y value that occurs when $x = 0$.) In the case of the Hubble Law, we are graphing D

[2] Mega = 1,000,000.

versus v, so that the straight-line formula will look like

$$v = mD + b.$$

Notice that the Hubble Law gives us a way to determine the distance to galaxies. We just need to determine the slope and intercept of the law using data from galaxies of known distance; for other galaxies of measured recession speed, we can then compute the distance by substituting into the Hubble Law.

Finding the intercept in the Hubble Law is actually very easy, since something at zero distance shows no recession; for the Hubble Law, $b = 0$. The slope astronomers normally write as H_0 (pronounced H-not, which is known as the **Hubble Constant**) instead of m, so the Hubble Law is normally written as

$$v = H_0 D.$$

The best current measurement of the Hubble Constant–determined from Hubble Space Telescope observations of Cepheids in nearby galaxies–is $H_0 = 71\ km/s/Mpc$.

The Types of Galaxies

After showing that extra-galactic nebulae were indeed other galaxies, Hubble began to study their basic properties. He obtained high resolution photographs of a number of the largest and brightest appearing galaxies, in order to learn the different shapes galaxies could take.

The Hubble Sequence

Hubble discovered that there were two basic shapes for galaxies, disk shaped (remember that the most prominent feature of the Milky Way is its disk), and spheroidal. Hubble thus devised a system of classification for galaxies based on the properties he observed for these two basic types of galaxies.

Spiral Galaxies

The disk shaped galaxies Hubble referred to as spiral galaxies, because most of the examples he first studied had prominent spiral arms. Hubble denoted spiral galaxies with an S, and denoted different subtypes for spiral galaxies based on the prominence and tightness of winding of the galaxy's spiral arms; galaxies with the least prominent and most tightly wound spiral arms he called Sa galaxies, those with the most prominent and least tightly wound arms he called Sc galaxies. Galaxies with intermediate arm prominence and winding he called Sb galaxies. Beside a disk, spiral galaxies also have a bright central bulge. Hubble noticed that in the galaxies he initially studied, the ratio of bulge to disk decreased as he went from types Sa to Sc. Hubble thus concluded that the relative size of spiral galaxy's bulge to its disk could also be used to assign its Hubble type. Finally, Hubble noticed that some spiral galaxies had central bars. To denote these galaxies, he added a B to their designation. So, for example, a spiral galaxy with a bar and moderately wound spiral arms would be classified as SBb.

Spiral galaxies tend to contain large amounts of cool and cold gas as well as significant amounts of dust. Stars form from the collapse of cold gas clouds, and spiral galaxies are observed to have significant rates of star formation. Sc galaxies have the greatest amount of star formation, with the amount decreasing as we go to types Sb and Sa. Having a bar doesn't affect the star formation properties of galaxies. So, for example, Sc and SBc galaxies have on average the same amount of star formation.

Elliptical Galaxies

Elliptical galaxies are spheroidal in shape. Hubble designated them with an E, followed by a number describing how flattened the galaxy appeared, with 0 for round galaxies, through 7 for the flattest ellipticals (for example, a round galaxy would be designated as E0.) It turns out that elliptical galaxies are actually triaxial in shape; that is, the have three distinct axes, each which has a different flattening (imagine something like a potato.) This means that the shape designation in an elliptical galaxy's designation tells us nothing interesting about the actual shape of the galaxy, since we would assign a different value if viewing the galaxy from a different direction.

In general, elliptical galaxies have little or no cool gas and dust, and also not surprisingly no signs of current star formation. In fact, it appears that all the stars in elliptical galaxies formed very rapidly when the galaxy first formed, and essentially no stars have formed since.

S0 Galaxies

After studying many more galaxies, Hubble found that there were additional types of galaxies that didn't fit his original system, requiring him to expand the scheme with additional classes of galaxies. One set of galaxies appeared to have both a disk and a central bulge, similar to spiral galaxies, but showed no signs of spiral arms or young stars in their disks. Hubble denoted these galaxies as S0 (that's a zero, not the letter O.) Like spiral galaxies, S0 galaxies can have a central bar, so both S0 and SB0 galaxies are possible. Hubble interpreted S0 galaxies as a transition type between elliptical and spiral galaxies (although, as we'll see, the situation turns out to be more complicated than that.)

Like elliptical galaxies, the great majority of S0 galaxies show no signs of current star formation, and appear to have formed almost all their stars early in the galaxy's history.

Irregular Galaxies

Another class of galaxy that didn't fit neatly into the original Hubble scheme were smaller disk shaped galaxies. Hubble these called irregular galaxies, which he denoted with an I. These systems have no central bulge and evidence of large numbers of young stars. Unlike spiral galaxies, irregular galaxies show no well defined spiral arms, but instead have an irregular, patchy structure (thus the name.) Irregular galaxies are generally smaller than spiral galaxies, but like spiral galaxies have large reservoirs of cool and cold gas.

Dwarf Galaxies

Irregular galaxies can be much smaller than the elliptical, spiral, and S0 galaxies we have discussed so far. For this reason, the larger types of galaxies are sometimes referred to as **giant galaxies**, and the smaller galaxies like the irregular galaxies are referred to as **dwarf galaxies**. It turns out that dwarf galaxies are much more common in the universe than giant galaxies.[3]

Many dwarf galaxies seem to be spheroidal in shape, and so are referred to as **dwarf elliptical galaxies**. This designation is somewhat unfortunate,

[3] Recall that the same is true for stars; high mass stars are rare, while low mass stars are plentiful. This seems to be a general trend in nature; low mass objects of a particular type are much more plentiful than high mass objects of that same type.

since it appears that most dwarf ellipticals aren't a low mass extension of giant elliptical galaxies.

The Distribution of Galaxies in Space

Galaxies are not randomly distributed in space; instead they tend to congregate in gravitationally bound groupings that are typically about 1 Mpc in size, and can contain anywhere from a few to thousands of galaxies. If the grouping is small, it is referred to as a galaxy group, while larger groupings are referred to as galaxy clusters. The predominant type of giant galaxy changes as you go from a typical small group of galaxies to the richest clusters. The majority of giant galaxies in small groups are spirals, while in the richest clusters, giant E and S0 galaxies predominate. The basic properties of different galaxy groups and clusters are summarized in the table below.

Name of Grouping	Number of Galaxies	% Spiral	% S0	% E
Galaxy Group	few-50's	80	10	10
Poor Cluster	50's-100's	50	40	10
Rich Cluster	100's-1,000's	20	60	20

Groups and clusters of galaxies are also not randomly distributed, but are clustered into larger structures known as super clusters. Viewed on a large scale, galaxies in the universe follow a filamentary structure, surrounding large regions that are mostly devoid of galaxies.

Chapter 13

Quasars and Active Galaxies

A few percent of galaxies in the local universe have signs of intense activity in their nuclei that is not powered by stars. These signs of activity take different forms; therefore there are several distinct types of active galactic nuclei (AGN) that have been identified. We believe that all AGN are powered by the same thing, accretion of gas into a massive black hole in the center of the galaxies.

Quasars

In the early 1960's, radio astronomers performed surveys of the sky, creating catalogs of hundreds of bright radio sources. Identification of optical sources that corresponded to these objects showed a wide range of objects, from supernova remnants in our own galaxy to external galaxies. One particularly puzzling type of object associated with these radio sources were star-like in appearance, and thus called Quasi-Stellar Radio Sources (or quasars for short.) These objects were puzzling because their spectra appeared like that of no known star, and in fact contained strong emission lines (remember, the great majority of stars exhibit absorption line spectra), which people were initially unable to identify.

The emission lines were finally identified as the highly red-shifted emission lines of hydrogen (it was this large shift that had caused the initial problem in identification.) Applying the Hubble law to the redshift measured indicated that the quasars must be at a large distance, and thus must be extremely luminous.

Surveys in other wavebands have shown that only about 10 % of quasars are strong radio sources. The host galaxies for several quasars have been detected by the Hubble Space Telescope. The number of quasars rises with redshift (i.e. quasars were more common in the past), with the number peaking at a redshift around 2, and the falling again. Many quasars exhibit brightness variations that occur over times of days, indicating that the emitting region are at the most a few light days across.

In addition to high levels of emission in the optical, quasars also emit large amounts of electro-magnetic radiation in the ultra-violet through the x-ray portions of the spectrum.

Other Types of Active Galaxies

Beside quasars, there are several other classes of galaxies that demonstrate activity that originates from a non-stellar power source in their nuclei.

Seyfert Galaxies

One to two percent of spiral and S0 galaxies have a nucleus that appears bright and unresolved (and thus star-like) in photographs, and show strong emission lines not dissimilar to those observed in quasars (although of much lower luminosity.) These Seyfert galaxies (named after Carl Seyfert, the first person to study them systematically) can be divided into distinct classes based upon the width of their emission lines. Those with the more narrow lines–indicative of gas rotating at around 500 km/s or so–are called type 2 Seyfert galaxies. Those with broader emission lines–indicative of of gas rotating at 5,000 km/s or so–are referred to as type 1 Seyferts.

Like quasars, Seyfert galaxies emit significant amounts of ultraviolet radiation and X rays. Although Seyfert galaxies show some level of radio emission, they are no where near as luminous in the radio as either radio loud quasars or radio galaxies (which we discuss next.)

Radio Galaxies

Another type of object that was common in the radio surveys from which the first quasars were identified were double lobed radio sources that turn out to be associated with some giant elliptical galaxies. the radio lobes were often found to extend well outside the optical galaxy, in the most extreme cases spanning around 1 Mpc from end to end. The core of the galaxy would also appear as a bright radio source, and in some cases faint jets could be discerned that stretched from the core of the galaxy out to the radio lobes. Optical spectra of the cores of these galaxies usually show strong emission lines.

Like the other types of AGN we have already discussed, the nuclei of radio galaxies tend to emit significant amounts of ultra-violet and x-ray radiation.

Blazars

Another class of bright radio sources associated with some elliptical galaxies are known as blazars (also sometimes referred to as BL Lacertae objects after the first one discovered.) The radio sources in blazars are compact, and can vary significantly in brightness over time. Optical spectra of blazars are different than those of the other AGN we have discussed, since they show only very weak emission or absorption lines.

Like the other types of AGN we have already discussed, blazars tend to emit significant amounts of ultra-violet and x-ray radiation, and are also strong sources of γ rays.

Chapter 14

The Expansion of the Universe and Cosmology

General Relativity and the Expansion of the Universe

Recall that general relativity is a theory of gravitation, the central idea of which is that mass and energy produce curvature in space-time. Any credible theory of the structure and evolution of the universe must at the very least be based upon general relativity (as we'll see in fact, there are other important pieces of physics that must be considered as well.) The trick for any would-be cosmology theorist is therefore to determine what the distribution of mass and energy within the universe is at some time; use that to write down the metric (remember the metric is the thing in general relativity that converts the mass distribution to how space is curved), and then use that to calculate the evolution of that universe with time. In principle this could be horribly complicated, since we know that matter is clumped together into concentrations of different size, shape, and distribution in space (e.g. planets, stars, galaxies, people, etc.) In practice what cosmologists do to make this problem tractable is to make an assumption about the behavior of the universe on large scales that is called the **cosmological principle**:

1. The universe is homogeneous (i.e. the same at all points.)

2. The universe is isotropic (i.e. the same in all directions.)

If the cosmological principle is true,[1] life gets much easier for you as a cosmologist, since the only thing you need to know to determine the metric is the *average* density of mass/energy in the universe.

Early Models

Soon after the publication of Einstein's paper describing general relativity in 1915, other physicists used it, along with the cosmological principle to construct model universes of different densities. They discovered that low density universes should continually expand, while higher density universes should first expand, and then collapse under the influence of their own gravitation. In these models therefore there was a **critical density**, below which expansion occurred and above which the universe was fated to eventually collapse. In fact, these results predated Hubble's discovery of the Hubble law by several years, but gave a natural explanation for what Hubble observed; all the galaxies in the universe are receding from us not due to motion *through* space, but instead are receding from us because they are *embedded in spacetime*, and are thus carried along with this universal expansion of spacetime predicted by general relativity.

The Big Bang

One of the early people who explored the implications of general relativity for cosmology was a Jesuit priest named Georges Lemaître. Lemaître reasoned that if you tracked the expansion of the universe back that the universe must have at earlier times been very dense, and so proposed the the universe began as an "explosion" from an extremely dense (perhaps infinitely dense) state. This proposed explosion is known today as the **big bang**[2]. The term is also used to denote the theory describing the subsequent evolution of the universe.

One consequence of the big bang idea was that the early universe must have been very dense and hot. Theoretical cosmologists concluded that there

[1] We know that this isn't true on small scales, but on larger scales it seems that it might be true.

[2] The term big bang was coined as a derisive one in 1949 by the British cosmologist Fred Hoyle, who was championing an alternative to big bang cosmology know as the **steady state theory**.

should be several consequences of this early hot, dense state.

1. In the early stages of the big bang, conditions should have been right for nucleosynthesis to occur. they calculated that at the end of this era of primordial nucleosynthesis, the matter in the universe should be approximately 75 % Hydrogen and 25 % Helium (percentages by mass), with only small traces of anything else.

2. For the first few 100,000 years after the big bang, the temperature would have been high enough that all the hydrogen would be ionized. Because electromagnetic radiation interacts strongly with free electrons, as long as the hydrogen was ionized the matter and radiation would be tied together, having exactly the same temperature. Over a very short time–due to the expansion and cooling of the universe–the hydrogen would capture all the free electrons (this process is sometimes referred to as **recombination**). From this point on, this early radiation would no longer interact with matter, and would simply cool with the expansion of the universe. Calculations showed that this radiation should have a blackbody spectrum, and at the current age of the universe ought to have a temperature of about 3 Kelvin.

The Cosmic Microwave Background

The cosmic microwave background radiation (the CMBR for short) was discovered accidentally by two radio astronomers–Arno Penzias and Robert Wilson–who were working for Bell Laboratories, studying the propagation of radio waves in the earth's ionosphere. This discovery was a great boost for the big bang picture, since it was a natural prediction of the big bang picture, but not of competing theories such as the steady state theory. Subsequent observations of the cosmic microwave background show that it:

1. Has a very precise blackbody shape, with a temperature of 2.725 Kelvin.

2. Is extremely smooth.

3. is extremely uniform in all directions on the sky.

The History of Universal Expansion

If the universe is expanding, then if we know the distribution of mass in the universe it should be possible in principle to use the mass distribution in the universe along with equations of general relativity to compute the expansion history of the universe.

As already noted, in the simplest universe, there are three possible fates for the universe; the mass density is low enough that the universe will expand forever (an **open universe**), the mass density is high enough to stop the expansion and cause it to collapse back to a dense state (a **closed universe**), or the mass density is on the dividing line (the so called **critical density**) so that as time goes on the expansion will slow, but never re-collapse (a **flat universe**.) The mass density is normally written as Ω_0 (pronounced omega-not), which is the ratio of the actual to the critical density as measured at the current cosmological time. These possibilities, along with the curvature that they would produce in the space of that type of universe (recall that in general relativity, mass produces curvature in space) are summarized in the following table.

Type	Ω_0	Space Curvature
Open	< 1	Negative
Flat	$= 1$	Zero
Closed	> 1	Positive

What is the value for Ω_0?

When people began to study the evolution of these different universes, they discovered a difficulty. If the universe were closed, and Ω_0 wasn't extremely close to 1 (the difference occurring only at a great number of decimal places), the universe would have already re-collapsed. On the other hand in an open universe, the expansion would proceed much too rapidly for the structure we see today (galaxies, stars, people, etc...) to have grown, unless Ω_0 was extremely close to 1 (again, to many decimal places.) This lead some cosmologists to conclude that something special about the universe guaranteed $\Omega_0 = 1$ *exactly*, because such a fine-tuned value seemed unlikely by chance[3].

[3]A complementary viewpoint is that only in a universe that was suitable for life to arise (i.e. one with Ω_0 extremely close to one) would we be here to ask questions about what its

By adding up all the mass we can see–detected either by its luminosity or its gravitational action on luminous matter–it should in principle be possible to determine the value of Ω_0. The table below summarizes the values obtained using various methods.

Source	Ω_0
Add up Luminous Matter	0.03
Rotation Curves of Galaxies	0.06
Big Bang Nulceosynthesis[4]	0.05
Orbits of Binary Galaxies	0.1-0.2
Clusters of Galaxies[5]	0.2-0.3

Problems with the Simplest Big Bang Picture

By the early to mid 1980's the big bang picture had racked up a number of successes in describing the universe we live in today (CMBR, big bang nucleosynthesis, etc...) but also had at least two troubling problems.

1. **The Flatness Problem** Tests that measured the large scale curvature of space, as well as the arguments about why Ω_0 should be very close to one mentioned earlier suggested that we live in a flat universe, On the other hand, attempts to measure the mass density by other means gave a value of Ω_0 significantly less than one.

2. **The Horizon Problem** In order for the CMBR to be as uniform as observed, all regions in the early universe must have been at the same temperature to high precision. However, in the simplest models, the different regions were separated by a distance larger than light could have traveled at that early time in the universe[6]. The only way for

value was, and therefore, we cannot use likely-hood arguments to tell us anything about *why* the values of certain cosmological parameters are what they are.

[4]Gives limit on baryonic matter only.

[5]We can separately use the orbits of galaxies within the cluster, the confinement of hot gas within the cluster, and gravitational lensing by the cluster to derive this value.

[6]This is where the term horizon comes from in this context. The distance light could travel if it was traveling for the entire age of the universe defines our **cosmic horizon**; in the simplest cosmologies, as the universe ages, the distance to this horizon increases. This means that for a particular observer in these cosmologies, the size of the observable universes grows as the universe ages.

every place in the universe to establish the same temperature was to be able to communicate with one another. The horizon problem boils down to the question, how did different regions of the early universe know how to be at the same temperature?

Inflation

A possible solution to both of these problems was suggested by a young particle physicist named Alan Guth. Guth pointed out that under certain circumstances in the very early universe (how early? somewhere between 10^{-40} and 10^{-30} seconds after the beginning) the universe could go through a brief period where it grew by 10^{50}-10^{100} times. This process–where the expansion rate greatly outstripped the expansion rate before and after–is referred to as inflation. If inflation occurred, two of the problems with the simplest big bang model could be solved.

1. **Flatness** By expanding by the immense amount indicated by inflation, the original curvature would have been immaterial; *any* curvature would appear flat when magnified by 10^{50} or more times.

2. **Horizon** During inflation, very small portions of the universe will be stretched out to immense size; all of the observable universe today would have come from a small region of the early universe, a region whose horizon before expansion was much larger than today's horizon.

While Guth's inflation picture could neatly solve these conundrums within the simple big bang picture, it did not explain whether there is any reason to think that such an event would actually occur.

Why Might Inflation Occur?

Maxwell showed that electricity and magnetism–two seemingly different forces–were actually two aspects of a combined electromagnetic force. What if it were possible that all the fundamental forces are different aspects of one underlying force in nature? This idea is known as **unification**, and in the late 1970's and early 1980's particle physicists showed that the electromagnetic and weak nuclear forces were two aspects of a single force known as the electroweak force. One consequence of this unification is that at sufficiently high energies, the action of the electromagnetic and weak force are indistinguishable. The forces are said to be symmetric, and at lower energies–where the actions of the two forces are different–the symmetry is broken. Many physicists have speculated that at high enough energies, all the fundamental forces might be unified into a single, symmetric, force.

If this is true, then in the very early universe, only this symmetric unified force operated, because the high temperatures at those early times corresponded to very high energies. As the universe expanded and cooled, the temperatures and thus energies would drop below that necessary for symmetry, and the symmetry would disappear as each of the four fundamental forces we know today became distinguishable. This symmetry breaking would take the form of a phase change in the early universe, which would release a tremendous amount of energy, possibly driving inflation.

Accelerating Expansion and Dark Energy

Cepheids are perfectly good for measuring distances to relatively nearby galaxies, but are not luminous enough to be detectable in distant galaxies. In order to extend the Hubble diagram to large distances, several groups of astronomers began to use supernovae as standard candles. What they discovered is quite surprising; the expansion of the universe is not slowing down, but instead appears to be accelerating. The currently unknown source of this acceleration is sometimes referred to as **dark energy**[7]. How might this acceleration arise in general relativity, and what is the source of this dark energy?

To answer this question recall that general relativity predicts either a contracting or expanding universe (but even if the expansion goes on for ever, it will slow with time.) In the 1910's and early 1920's however, most astronomers believed that the universe was neither expanding or contracting (i.e. was static.) In order to get a static solution to the equations of general relativity, Einstein introduced a "cosmological constant" (denoted as a capital Greek lambda (Λ)). Once it was shown that the universe was actually expanding, Einstein abandon the idea of the cosmological constant, declaring it to be "...the biggest mistake of my life."

[7]The term dark energy is somewhat unfortunate, since all they types of energy we have talked about so far would act to slow the expansion, just as mass does. Dark energy is *not* simply another form of energy similar to those we have already talked about.

In fact, it turned that the cosmological constant would not have solved the static universe problem, because although you could setup a static universe with a cosmological constant, that universe would not be stable; any small perturbation or fluctuation within the universe would set it expanding or contracting. In fact, if the value of the cosmological constant were set correctly, it could create an accelerating expansion.

What Might Produce $\Lambda \neq 0$?

To see what in the universe might create a non-zero cosmological constant, quantum physics again must enter the picture. Recall that one of the key results of quantum theory was that we cannot exactly predict the position of a particle, but only the probability that it is found at different locations. In essence, the thing we think of as a particle is smeared out, or stated another way, there is an uncertainty to the exact position of the particle (and in fact of many other properties of the particle, such as its momentum and energy.) In addition, the uncertainties of the different quantities associated with the particle are linked by what is known as the **Heisenberg uncertainty principle**. For example, the uncertainty in the energy of a particle (ΔE) and its time (Δt) are related to one another as

$$\Delta E \Delta t \geq h/2\pi.$$

Notice, the more certain the energy, the less certain the time, and vice-versa.

One implication of the uncertainty principle is that particle-anti-particle pairs should constantly coming to existence and disappearing again. How long can a pair like this exist? If we replace ΔE with the rest mass energies of the particles (E), they can exist for a time less than $h/(2\pi E)$, without violating any physical law of the universe. These particle-anti-particle pairs are called **virtual particles**. It is possible that because seemingly empty space should be filled with these virtual particles that they could create a non-zero cosmological constant.

A few problems exist with the cosmological constant being created by virtual particles, however.

1. It turns out to be quite a difficult trick to actually calculate how large a cosmological constant should be created by virtual particles, because in order to do the calculation you must make certain approximations. Attempts to do so by different people have variously determined that

its value should be zero exactly, or that its value should be much larger than indicated by the supernovae measured acceleration (by many powers of 10). Either the current estimates are incorrect, or another source for the cosmological constant other than virtual particles must exist.

2. If Λ is created by virtual particles, it must have a value fine-tuned to 54 decimal places.

Other Possibilities

Because of these difficulties, some cosmologists have suggested a different source for dark energy, known as **quintessence**.[8] Quintessence is a hypothesized energy field that has a strength that depends on the cosmological epoch, and thus avoids the fine tuning problem that exists with a cosmological constant. The presumed source(s) of quintessence are even more esoteric than virtual particles.

[8]The term quintessence comes from the ancient Greeks, who used it to describe the fifth element in nature from which all celestial objects were made.

Chapter 15

Galaxy Formation and Evolution

- How does structure we see today (galaxies, clusters, etc...) form from smooth universe we see in cosmic background radiation?
- Why are there both disk and spheroidal galaxies?
- Why do the properties of different disk galaxies vary by so much?

The Structure Problem

- In the simplest cosmology (i.e. one without non-baryonic dark matter), fluctuations in the cosmic microwave background are too small to grow to be the structure we see today.
- Since indications are that mass in the universe *is* dominated by non-baryonic dark matter, growth of structure due to gravitational collapse will be dominated by the behavior of the dark matter.

Hot vs. Cold Dark Matter

- If dark matter particles have high mass (WIMPs), then dark matter will be moving slowly (cold dark matter), and will clump up due to its self gravitation.

- If dark matter particles have low mass (neutrinos?), will be more smoothly distributed.

so:

- If dark matter is cold, small structures (sub-galactic) will form first, and galaxies will grow due to the merger of these sub-galactic clumps.
- If dark matter is cold, large structures (group or cluster sized) will form first, and galaxies will form from the fragmentation of these larger structures.

Observations of Young Galaxies

- Many galaxies at high redshift ($z \approx 5 - 6$) are small (kiloparsec or smaller.)

but...

- Also find fully formed ellipticals at redshifts ($z \approx 2 - 3$)

so:

- Need mix of cold and hot dark matter to explain? (Also needed to get correct amount of structure in large scale galaxy distribution.)
- Need to actually know what dark matter is to know if we're on the right track!

Why are their both Disks and Spheroids?

- Why do disk galaxies rotate?
  - Neighbors acting on neighbors?
  - Off-center collisions during buildup?
- Why don't spheroidal galaxies rotate?
  - Rapid merging of clumps during formation?
  - Made from merger of 2 disk galaxies?

Why are Disks so Different from One Another?

- How do bulges form?
  - Form rapidly, before disk formation? (Evidence in Milky Way supports this idea.)
  - Bulges form from mergers of dwarf galaxies with larger disks?
- Why are gas contents (and thus star formation) so different as we go from S0's to Sc's?
  - Environment can remove gas/inhibit addition of gas.
  - Gas poor spirals observed in centers of nearby clusters.
  - Percentage of of S0's much smaller in higher redshift clusters.

Chapter 16

Intelligent Life in the Universe

Sometimes I think the surest sign that intelligent life exists elsewhere in the universe is that none of it has tried to contact us. –Bill Watterson