Introduction to Astronomy 112

David Burstein

I. INTRODUCTION

1. Distribute course description and syllabus

a. The syllabus is to serve only as a general guide to the material covered by this course. Since this is the first time this course is being taught, the specific topics discussed on specific days will probably deviate from this schedule.

b. Textbook: Exploration of the Universe
Abell, Morrison and Wolff, 5th Edition

c. Parameters of the course: Exams
Grading system

d. The students will be responsible for the material covered in the book, the supplemental material given in the class lectures, and the material given in these notes.

2. An overview of the universe

a. A Sense of Scale: Relative sizes of

Humans: About 6 feet tall
A Mile: about 1000 Humans
Earth Radius: 4000 miles
Distance from Earth to Sun (= 1 Astronomical Unit): 28000 Earth radii
Radius of Solar System: 40 Earth-Sun distances
Distance to nearest star: about 250,000 Earth-Sun distance
Size of Galaxy: about 50,000 distances to nearest star
Distance to nearest Galaxy: about 10 times size of our Galaxy
Clusters of Galaxies: about 100 times distance to nearest galaxy
Size of Universe: about 100 times size of large clusters of galaxies

b. A Sense of Time: Relative units of time:

Second: 1
Day: 86,000 seconds
Year: 365 1/4 days
Age of Earth: 4,600,000,000 years (approximately)
Estimated age of Sun: 5,000,000,000 years
Esitmated age of Universe: 15,000,000,000 years

3. Scientific Notation: In science, and especially in astronomy, one often wants to compare objects that are relatively small (like the size of an atom) to objects that are relatively large (like the size of the Sun). For such comparisons, taking the difference between the two sizes does not make much sense; for example:

Sun has a radius of 69,600,000,000 cm

and an atom typically has a radius of .0000001 cm

However, taking the ratio of these two sizes is informative:

The Sun is 696,000,000,000,000,000 times bigger than the typical atom.

Rather than continually write all these big numbers, which is both tiresome and clumsy, scientists have adopted a way of expressing large numbers in terms of powers of ten. For example, from the above example, the Sun is

6.96 x 1017 times the size of a typical atom,

where the symbol 1017 means multiply 10 by itself 17 times, and the notation 6.96 x 1017 means multiply 10 by itself 17 times, and multiply this by 6.96. This form of writing big numbers is generally termed scientific notation. Your familiarity with this means of writing large numbers is assumed by the text, and scientific notation will be commonly used in this course. If you are unfamiliar with this kind of notation, Appendices 3, 4 and 5 of the text contain additional information and explanations of the notation and units of measure commmonly used in astronomy.

4. Astronomical Precision: The careful student will note that the numbers that I used for the scale comparisons above do not precisely match the numbers given in the book for similar comparisons. Why?

First, the only objects whose size, mass and absolute brightness we can measure precisely are those objects in our solar system, for which we have very accurate ’yardsticks’. All other objects in the Universe are so far away that we must use much less precise means of measuring distances. Without precise distances, we cannot measure precise masses or brightnesses. For example, our estimate of the visible size of the universe is nearly 10 times larger than the best estimate made by astronomers in the 1930s. The universe hasn’t changed during that time, just our means of measuring distances!

As a result, one of the most important objectives for astronomers is to know the relative scale of things: Namely, that the Sun is nearly 1018 times the size of a typical atom, not 102 the size, or 1016 the size. In a universe where ratios of sizes can be 1020 or more, factors of 1.2 or 1.5 in size can often be viewed as being very similar in size. (For example, Jupiter is about 1.2 times the size of Saturn, but over 11 times the size of the Earth. Therefore, Jupiter and Saturn are of similar size, and both are much bigger than the Earth.)

Thus, I will teach this course in such a way that you get used to the idea that both the actual sizes and the relative sizes of objects are important.

II. EARTH, SKY AND MOTIONS

1. How do we know the Earth is rotating on its axis?

a. Foucault Pendulum

The large pendulum in the front of the Physical Science F wing is an example of a Foucault pendulum, so-named after the Frenchman who invented it. The motion of a pendulum gives it angular momentum (remember: since a pendulum is constantly changing its direction of motion, a force (gravity) must also be continuously acting on it), and its motion is in a plane (that is, back and forth). That motion is not affected by twisting the cord holding the pendulum (since this cord moves with the pendulum). Therefore, any pendulum permitted to swing in an open area will appear to rotate around the ground, due to the fact that the Earth is rotating underneath the pendulum.

How fast a Foucault pendulum rotates with respect to the ground depends upon where you are on the surface of the Earth. At the North and South poles, all of the apparent motion of the Earth is in rotation, so the pendulum takes 24 hours to rotate. In Phoenix (or any other place not at the Equator), part of the apparent motion of the Earth is in rotation, part in horizontal movement, so that the pendulum takes longer to rotate with respect to the ground (about 44 hours). At the equator, the apparent motion of the Earth is completely in horizonal movement, so the pendulum will not rotate at all.

b. Coriolis Effect

Since the Earth rotates on its axis as a solid body, the actual speed of rotation must depend on your distance from that axis (which we call the North and South Poles). For an analogy, think of a 'whip' in ice skating: a line of skaters moves in a circle at approximately a constant rate; the skater at the end of the line picks up the largest velocity and, when released from the line, moves away with a velocity much higher than he/she could have generated on his/her own.

In order for you not to be moving with respect to the ground, you must also have the speed of rotation of the Earth corresponding to that point on the Earth where you are (here in Phoenix, that speed is about 600 mi/hr!). If you go up in the air here in Phoenix (that is, you are no longer on the ground) and move north, your velocity horizontal to the Earth will be greater than the velocity the Earth is rotating, so you will move eastward. Conversely, if you move south, you will have a lower velocity than points on the Earth south of you, and you will move westward. If you lived in the Southern Hemisphere, moving north and south would produce the opposite effects. This kind of motion is called a 'Coriolis' motion, and, as explained in the text, is the cause of cyclonic motions in our atmosphere (that is, the direction of rotation of storms).

2. Coordinate Systems

The need for coordinate systems is nearly universal: If I want you to go the corner of Indian School Rd. and 7th Ave, either you would have to know the road system in Phoenix, or I would have to explain it to you. In a much more general way, this is true for coordinates on the Earth, or in the sky: if I want to communicate with someone else, it is very helpful if we are using the same coordinate system.

Basically, there are two kinds of 'natural' coordinate systems in Astronomy, as used by us here on the Earth, spherical general and spherical local:

a. Spherical general: Our Earth is a sphere, and the sky around us looks spherical. The rotation of the Earth makes the sky appear to rotate in the opposite direction. In addition, the rotation of the Earth is about a well-defined axis. Thus, a coordinate system that is spherical is a natural coordinate system for measuring positions both on the Earth and in the sky. On the surface of the Earth, we call the coordinates of this spherical coordinate system longitude and latitude: latitude is measured between the two poles of the axis of rotation, with the convention being that the North Pole has a latitude of +90o, the South Pole a latitude of -90o, and the circle on the Earth that is equi-distant between the two poles (the 'equator) has a latitude of 0o. Longitude is measured around a circle at a constant latitude; where we place the zero of longitude is arbitrary. In the 1700's sea-faring nations agreed to place the zero of longitude at Greenwich, England.

The celestial spherical coordinate system is very analagous to the Earth's spherical coordinate system: Celestial North Pole, Celestial South Pole, Celestial Equator. The analague of latitude on the Earth is called 'declination'; the analogue of longitude on the Earth is called 'Right Ascension'.

b. Spherical Local: At any given place on the Earth, there are 'up-down' and 'all-around' directions. The 'up-down' direction is called altitude; the 'all around' direction is called azimuth. The highest point in the sky is called the 'zenith', and has an altitude of 90o (the lowest point is called the 'nadir', and is below our feet); the horizon has an altitude of 0o. The place were azimuth is zero is again arbitrary, and is taken to the be direction pointing north.

At any given point on the Earth, most stars will not pass through the zenith, but all stars will reach a highest altitude. This highest altitude lies on an imaginary line drawn between between the two celestial poles that passes through your zenith, and is called the 'celestial meridian.'

c. Meauring conventions:

i. Angular measure: Most coordinates are measured in terms of angles: there are 360o in a circle, 60 arc-minutes in a degree, and 60 arc-seconds in an arc-minute. Since you can always draw an imaginary circle around yourself, you can always measure angles with this system. Altitude, azimuth and declination are measured in terms of degrees, arc-minutes and arc-seconds.

ii. Measures of time: It takes the Earth 24 hours to 'move' the sky around a full circle of 360o, or 1 hour to 'move' the sky 15o, or 1 minute to 'move' the sky 15 arcminutes, or 1 second to 'move' the sky 15 arcseconds. Thus, we have the option of measuring the Right Ascension coordinate of the celestial sphere in terms of either angles or times. Astronomers find times more useful, since that can directly tell us (with a little additional information), where the star is on the sky tonight.

d. Angular Measure and Linear Measure

On a Circle: If you look on a sphere of the Earth, there are 360o of longitude at every latitude, but the actual circumference of the circle decreases in size as you move away from the equator (in either direction): that is what makes a sphere a sphere! The same is true in the celestial coordinate system: A difference of 1 hour in Right Ascension for two stars on the Celestial Equator places them much further apart on the sky than a difference of 1 hour in Right Ascension does for two stars near the North Celestial Pole.

3. The Seasons

Simply and directly put, we have seasons on the Earth because the rotation axis of the Earth is tilted 23.5o relative to the plane of the orbit of the Earth around the Sun.

The Terminology of the Seasons:

Equator, North Pole, South Pole

Tropic of Cancer - the latitude of +23.5o
Tropic of Capricorn - the latitude of -23.5o
Arctic Circle - the latitude of +90o - 23.5o = +66.5o
Antarctic Circle - the latitude of -90o + 23.5o = -66.5o

Solstice - (sun standing still): furthest travel of Sun north is called the summer solstice; furthest travel of Sun south is called winter solstice. Note that these names are biased towards people who live in the northern hemispheres. The start of summer in the southern hemisphere is marked by the winter solstice, and the start of winter there is marked by the summer solstice.

Equinox - (equal lengths of nighttime and daytime): when the Sun is directly over the Equator; Vernal equinox is when Sun is going from winter solstice to summer solstice; autumnal equinox is when Sun is going from summer solstice to winter solstice.

4. Proofs of the Earth's Motion around the Sun

The ancient Greeks understood that, if the Earth revolved around the Sun, then the stars should appear to move (parallax). Not seeing parallax with your eyes meant one of two things: either the Earth did not move, or the stars were too far away. We know now that the second reason is correct.

4. Proofs of the Earth's Motion around the Sun (continued)

The largest parallax known of any star is 0.75 arcseconds. However, there is another effect of motion that is much larger as seen here on the Earth: aberration. Aberration is the term used to descibe the apparent motion of an object towards you when you are moving horizontally with respect to that object (the book uses raindrops and a pipe; also think of a quarterback trying to throw a pass - he throws it to where the receiver will be, not where the receive is when he releases the ball).

The effect of abberation is relatively large: 20.5 arcseconds. Although parallax was known to the ancient Greeks, it was not until the late 1700's that abberation was discovered (by astronomers looking for the parallax of stars!).

5. The Many Motions of the Earth

As summarized in the book, the Earth has many different kinds of motions in the Universe. We do not feel these motions because we are traveling with the Earth. However, if we want to travel outside the Earth, we have to take these motions into account.

6. Time

a. Hour angle: The angle in Right Ascension between an object in the sky and the position it would have on the meridian. This angle is usually measured in terms of time, since that tells either how long it will take a star to reach the local celestial meridian, or how long it has been since a star was at the local celestial meridian. Stars that have a declination of zero (0o, or on the Celestial Equator) will spend 12 hours in the sky at Phoenix; stars with declinations greater than zero will spend longer than 12 hours, stars with declinations less than zero, shorter than 12 hours. All stars pass through an upper during their apparent journey across the sky (either at night or in daytime); stars that are circumpolar pass through both an upper and a lower meridian.

b. Different Systems of Measuring Time

Solar: with respect to the Sun

Sidereal: with respect to the Stars

Because we revolve around the Sun, our solar day is always longer than our sidereal day by 1/365 of a day, or about 3 min 56 sec. Over the period of a full year, this means that the there will be 366.242... sidereal days in a year, compared to 365.242... solar days.

Apparent Solar Time: Solar time measured where you are; namely, from the position of the Sun in the sky, relative to the meridian.

As explained in the book, the apparent solar day is not constant over a year, but varies due to 1) the velocity of the Earth in its orbit is not constant, but varies due to the elliptical shape of the orbit; and 2) the Sun is constantly changing its declination, which means the actual angle it travels along the ecliptic will correspond to different angles in Right Ascension.

Mean Solar Time: Average out the apparent solar times to an ’mean’ solar time that is approrpriate for everyone, at all latitudes on the Earth.

Standard Time: The Earth is divided into 24 time zones, each corresponding to 15o of longitude (= 1 Hour of Right Ascension), but some are modified to avoid dividing a city or an island into two time zones. Within each time zone mean solar time is the same. This system of 24 time zones measured in mean solar time is termed ’Standard Time’.

Daylight Time: Setting the mean solar time ahead one hour, to increase the number of daylight hours during the mean solar hours of 6AM to 9PM. Arizona does not use Daylight Time, so it is on a different mean solar time from its neighboring states from April to October.

International Date Line: As with longitude, there must be some place where the ’zero point’ of the day begins. Since days run continuously, there must be some point where you go from today to tommorow (if you travel east) or from today to yesterday (if you travel west). To minimize the confusion this causes travellers, the international data line is located in the middle of the Pacific.

c. Calendars: I have little to add to the exhaustive discussion of calendars given in the text.

7. Picturing Orbits in Three Dimensions

In order to understand the apparent motions of double stars, not to mention the motions of the Moon, Sun and the planets in the sky, it is helpful if you can picture their motions as someone not in orbit would see them, and view them in three dimensions. We realize that this is difficult to do (even for us astronomers!), but there are some aids:

The orbit of one object around another object (such as the Moon around the Earth, or the Earth around the Sun) is in a plane; that is, if you viewed the orbit from the edge, it would look flatter than you can imagine.

However, these flat orbits exist in a three-dimensional world. Most of you know what a Hula hoop is: a large plastic ring. Imagine the orbit of the Moon around the Earth as a Hula hoop. Imagine the orbit of Earth around the Sun as another Hula hoop. Now imagine that one of these hoops goes around a point on the other hoop. We will call this 'point' the position of the Earth; the hoop that goes around this position the orbit of the Moon; and the hoop the Earth is on, the Earth's orbit.

If you look at the Earth's orbit edge-on, the Moon's orbit can, in principle, be at any angle to the Earth's orbit. But the Moon's orbit must intersect the Earth's orbit at two points, since the Moon's orbit is around the Earth. Well, the Moon's orbit is not tilted at 'any' angle relative to the Earth's orbit, but at an angle of about 5o.

If you hold the two hoops up long enough, you might see that, even if we require the Moon's orbit and the Earth's orbit to be at an angle of 5o, we have not required that Moon's orbit tilt towards you, away from you, or in any other direction. In other words, the Moon's orbit can stay tilted at 5o with respect to the Earth's orbit around the Sun, but the two points where they intersect can make any angle relative to a line drawn betwen the Earth and the Sun. These two points of intersection are given a special name, the 'nodes'. When the nodes are lined up with the line joining the Sun and the Earth, we have eclipses.

III. NEWTON, MECHANICS AND GRAVITY

Isaac Newton was able to take the varied observations and empirical laws of orbits that existed and give them a fundamental framework - the concept of gravity. He was able to define 'laws' of motion that describe how objects move. In order to accomplish these tasks, Newton had to invent the branch of mathematics we now call calculus. Newton also experimented with telescopes and optics, providing the foundations of the science of Optics as well.

1. Newton's 3 Laws of Motion:

   a. The concept of inertia - a body at rest stays at rest; a body in motion stays in motion.

   b. The strength of a force felt on a body is equal to the amount of mass in the body times the acceleration the body feels (F = ma).

   c. For every action there is an equal and opposite reaction.

2. The basic concepts of mechanics:

   a. length, area (= length x length), volume (= length x length x length (measured in units of a centimeter)

   b. time (any measure of time will do, but the standard measure is the second)

   c. speed = length/time - how fast you are going

   d. velocity = length/time in a given direction

   e. acceleration = the change of velocity with time (length/time/time)

The three laws of Newton introduce additional concepts:

f. mass = the amount of matter in an object, technically defined in terms of Newton’s second law: a known force F, applied to a mass m, gives the mass an acceleration a. Two masses are considered to be the same if the same force produces the same acceleration (in the absence of friction, of course).

g. momentum = defined by Newton as (mass x velocity)

h. Force = since an acceleration is the change of velocity with time, Newton’s equation F = ma also can be interpreted as Force is equal to the change of momentum with time.

i. Volume = the ’space’ occupied by a mass

j. density = mass/volume

k. angular momentum = Any object in orbit around another object is constantly changing direction and, hence, changing its momentum as defined in the standard way. A separate definition of momentum is used for objects in orbit, defined by Newton as (mass x velocity x radius), where radius refers to the distance between the object and the center of its orbit. In general, the angular momentum of an object is the same as long as a constant force affects the object, directed along the line between the object and the center of motion. If the orbit of an object decreases by 2x, the velocity of that object must increase by 2x; if the mass of an object decreases (such as a satellite using fuel), its orbit will increase in size.

l. Centripetal acceleration = acceleration of an object produced by a force directed along a line connecting the center of orbit to the object.

3. Law of Universal Gravitation

Newton hypothesized that a universal attraction exists between all masses everywhere in the universe. Using Kepler’s laws, Newton deduced that the attraction between two objects is due to a force that is proportional to the masses of the two objects (lets call them m1 and m2), and inversely proportional to the square of the distance (called d) between their centers:

F = (G x m1 x m2)/d2

The letter G in this equation gives the conversion between how much mass in involved in a gravitational interaction, and how much gravitational force results from that interaction.

A gravitational attraction between two masses is true for any object having mass (a pencil, the desk, you, the building, the Earth, the Sun, the Galaxy, ...), and moving at any speed relative to each other (for example, a bullet shot from a rifle and a baseball fall to the ground at the same rate).

Weight is the term most commonly confused with the term mass, but our weight is really a force, not a mass: the result of the mutual attraction of our mass with the mass of the Earth. We are constantly being accelerated towards the center of the Earth, and are constantly stopped by the solid surface of the Earth (or floor). Hence your weight is really the force exerted on you by the gravitational field of the Earth.

If we look at Newton's law of gravitation in terms of Newton's second law, then

Gravitational Force(on m1 due to m2) = (Gm2/d2) x m1

and Gravitational Acceleration(on m1 due to m2) = Gm2/d2 ,

Gravitational Force(on m2 due to m1) = (Gm1/d2) x m2

and Gravitational Acceleration(on m2 due to m1) = Gm1/d2 .

Suppose the symbol m1 represents the mass of the Earth, and the symbol m2 represents your mass (or that of a pencil, the desk, the building, ...). The distance d is the radius of the surface of the Earth. Then the gravitational acceleration produced by the Earth on you is (mass of Earth)/(mass of you) times larger than the gravitational acceleration you produce on the Earth:

(mass of Earth = 6 x 1027 gm)/(mass of typical you = 8 x 104 gm)

or, the mass of the Earth is about 1023 times your mass. Therefore, the gravitational acceleration produced by the Earth on you is about 1023 times more than the acceleration you produce on the Earth.

Since the gravitational acceleration between two masses depends on the distances between their centers, you actually 'weigh' a little less on top of a tall mountain than at sea level, since you are futher from the center of the Earth.

4. Orbits

a. Down-to-Earth Background

Since gravity is a force, and forces produce accelerations (Newton's second law), then any two objects with mass will cause each other to be in continuous acceleration. Thus, the Earth going around the Sun, the Moon going around the Earth, a spacecraft going around the Earth, Mars going around the Sun, etc., etc. are all being continuously accelerated by gravity.

From Newton's third law, the gravitational attraction of two masses must conserve the total momentum of the two masses. Thus, every system of masses has a 'center of mass', or barycenter, whose motion is unaffected by the gravitational interaction of these masses. (However, note that the motion of the barycenter can be affected by forces external to the masses. For example, the relative momentum between two persons in a car, relative to the car (which is conserved), compared to the motion of the car along the road (which can speed up or slow down).

Two masses in gravitational attraction must therefore orbit around their barycenter. The position of the barycenter between two masses must lie along a line joining the centers of the two masses. The distance of the barycenter from the center of each mass is inversely proportional to the relative masses: in the example of you and the Earth, the barycenter is 1023 times closer to the center of the Earth than it is to you (or smaller than an atom's diameter from the center of the Earth!). When you jump up and down, techically you and the Earth are in mutual orbit around each other - it is just that the Earth moves very little, and your orbit has little height above the Earth!

b. Orbits of Planets and Stars

Objects in mutual gravitational attraction will orbit about their barycenter. Two examples: 1) The planets are much less massive than the Sun, so the Sun moves very little about the barycenter of the Solar System, and the planets will move around the Sun. 2) Two stars of equal mass that orbit each other. The barycenter will be inbetween the two stars, so they will both orbit around the barycenter. Neither star will appear stationary - both will be constantly in motion.

c. Kepler's Third Law

Kepler's three laws were derived by Kepler before Newton, at the beginnings of modern science. Kepler defined three 'laws' that describe the motions of planets around the Sun:

1. Planets move in orbits that are ellipses, with the Sun at one focus of the ellipse

2. Orbits sweep out 'equal area in equal time'

3. The period of an orbit is related to the size of its orbit by an equation given the name 'Harmonic Law' -

$$P2 = K a3 ,$$

where P is the period, a is the semimajor axis of the orbit, and K is a number that converts (size)3 to (time)2.

As explained in the book, Newton's laws of motion and gravitational law can be used to derive Kepler's three laws in terms of mathematical equations. Thus, Newton did not replace Kepler's three laws, but explained why they existed, and incorporated them into a broader picture of planetary orbits.

Of Newton's interpretation of Kepler's three laws, the one of most interest to most astronomer is the third law:

$$(m1 + m2) \times P2 = 4 \quad 2/G \times a3 \quad .$$

Newton found that K is not really a constant value, but

$$K = 4 \quad 2/(G \times (m1 + m2)) \quad .$$

K only appeared to be constant for planetary orbits because the Sun's mass is so much larger than the mass of most of the planets. For stars in orbit about one another, K has to be defined explicitly in terms of the masses of the two stars, since the stars are of comparable mass.

If we are very careful in how we choose to measure mass, distance and time, Newton's version of Kepler's third law can be simplified (see Table 4.1 of the book).

Kepler's third law is so useful in astronomy because it relates the masses of objects to their orbital periods and their physical separations. This is the principle and most reliable means of directly measuring the masses of objects in the universe.

NOTE: Gravity is a force that accompanies the presence of mass. In order to not collapse completely under the force of gravity, all objects with mass must possess a means of generating internal pressure. Our bodies do it by chemically generating energy. The Earth does it both chemically and by collisions of particles in the Earth (which heats the interior of the Earth, and creates enough internal pressure to support the outer layers of the Earth against gravity).

IV. ELECTROMAGNETIC RADIATION

Essentially all of the information that astronomers can collect about objects outside the Solar System comes to us via electromagnetic radiation: we see the stars; we can use light to measure the temperatures of stars; how big the stars are; how far stars are away from us; how much mass stars have.

The subject of this chapter is to explain to you how we can do all these things. The basic reason is that all objects are made of atoms and molecules; all objects have heat (some more, some less); heat is the way we measure the amount of motions of atoms and molecules; the motions of atoms and molecules in an object produces electromagnetic radiation; hotter objects produce more radiation and 'hotter' radiation; cooler objects produce less radiation and 'cooler' radiation. Now for the details:

1. Light as a form of Electromagnetic Radiation

What we call 'light' is that part of electromagnetic radiation that our eyes can see. Light can be described as a wave; the light we see as 'blue' has a range of wavelength from about .000042 cm to .000047 cm, light we see as red from about .000060 cm to .000065 cm.

However, electromagnetic radiation comes in all wavelengths; we just call it different names: In order of increasing wavelength: gamma rays; X-rays; ultraviolet radiation; optical radiation (that is, what we see); infrared radiation; microwaves (or millimeter waves); radio waves.

Waves can interact with each other to produce interference and diffraction.

2. Inverse-square law

The intensity of light decreases as the distance squared from the source of light: this is generally known as the inverse-square law for the propagation of light, and is simply a geometrical effect:

Draw a sphere around a light bulb of any size. Now draw a second sphere that is twice the size of the first; a third sphere 3x the size of the first, and so on and so on. From geometry, the

surface area of a sphere = 4 r2.

The amount of light hitting each sphere is the same, irregardless of how big the sphere is; therefore the amount of light hitting the same area, measure in cm, must decrease with distance in precisely the same manner as the area of a sphere increases with the increasing size of the sphere. Since the sphere increases in size as distance x distance (= distance squared), the intensity of light measured in one square cm must decrease with distance as 1/d2. The only way to modify this law is to either a) add additional light along the way or b) block or obscure part of the light.

IVa. ATOMS, MATTER AND LIGHT

The text presents the relationships of electromagnetic radiation to atoms, and the relationships of atoms to matter, in a particular fashion. The following discussion presents these same concepts in a different way, not because the manner in which the book treats this subject is bad, but because these are difficult concepts to comprehend, and seeing it done in two different ways should be helpful.

1. Atoms and Matter - Terminology:

a. element: cannot be broken dwon by chemical reactions to a simpler substance.

b. atom: the smallest unit of an element that possesses its chemical properties.

c. molecule: compound, or group of atoms chemically bonded together.

d. Principal parts of an atom:

i. Nucleus - small dense core of the atom, it is made up of

protons which have positive electric charge.

and neutrons which have no electric charge.

Protons and neutrons have essentially the same mass (1.67 x 10-24 gm)

ii. Electrons - 1/2000 the mass of a proton or neutron, they have negative charge, and orbit around the nucleus.

2. The Relationship betwen Color and Temperature

a. Scientists measure temperature in units of either degrees Centigrade or degrees Kelvin. The difference between the two systems is in the meaning of 0o: for the Centigrade system, 0o corresponds to the melting point of water at sea level; for the Kelvin system, it corresponds to the temperature at which there is absolutely no motion, a temperature called 'absolute zero'. Astronomers use the Kelvin system (oK).

b. Every object in the universe comprised of more than a few atoms, that has a temperature above absolute zero (0o K) radiates electromagnetic radiation. This radiation is called thermal radiation, since it is related to the temperature of the object. This thermal radiation is continuous radiation, that is, it emits electromagnetic radiation at every wavelength.

The color of an object, in astronomical terms, describes the region of the electromagnetic spectrum in which the object radiates most ot its radiation. It is given the name 'color' since we name the different wavelengths in the visible part of the spetrum by their color (e.g., Red, Orange, Yellow, Green, Blue, Indigo, Violet, or ROYGBIV).

For thermal radiation, the 'color' of an object is related in a strict sense to the 'temperature' of the object, be that object a star, a planet, a person or an ice cube.

In the latter part of the 19th century, three empirical 'laws' of thermal radiation were discoverd:

i. Wien's displacement law:

The peak of the thermal radiation emitted by an object moves to shorter wavelengths as the object gets hotter. The relationship between temperature and the peak wavelength is one-to-one, and is expressed by the simple equation:

peak = 0.29/T(oK) cm,

meaning that if you express temperature in degrees Kelvin, their is a constant number (0.29) that converts inverse temperature to wavelength in units of centimeters.

ii. Stefan-Boltzmann Law:

The total energy emitted by an object (covering all wavlengths) per square centimeter of area (that is, per a standard area) increases rapidly with increasing temperature. The energy increases as the fourth power of the temperature, or

Etotal =   T4 ,

where (again), the Greek symbol   is a constant number that relates T4 to the total energy released.

iii. Planck Law:

This is the description of how thermal radiation is distributed in the electromagntic spectrum for an object of a given temperature. This is, how does the radiation of a rock at a temperature of 400 oK compare to the radiation of a rock of the same size, but at a temperature of 2000 oK? Planck's law states that, per unit area, an object at a hotter temperature radiates more energy at all wavelengths that an object at a cooler temperature.

Moreover, Planck was the first scientist to make the fundamental assumption that all electromagnetic radiation is emitted in discrete intervals, called 'quanta' of energy (singular term is quantum).

To summarize: If one compares a cool object to a hotter object, the hotter object will:

radiate its peak energy at shorter wavelengths and, hence, have the characteristic color corresponding to those wavelengths (Wien's law)
radiate much more total energy (Stefan-Boltzmann law)
and radiate more energy at all wavelengths per unit surface area (Planck's law)

The careful reader will note that the term per unit surface area is underlined in the above text. That is because it is always important to keep in mind that how bright an object appears to us, due to emission of electromagnetic radiation, is dependent not only on its temperature, but also on how big it is and how far away it is (remember the inverse-square law of light). For example, the average human being has a temperature of about 315 oK, while the average filament in a 100-watt incandescent light bulb has a temperature of about 2800 oK. Although the average human being is about 8,000 times the size of the filament, the hotter filament puts out about as much energy as the average human being. Both objects have the same absolute brightness, but the filament is much hotter.

In addition, Planck's law also describes the manner in which thermal radiation is emitted with wavelength. As the book gives in footnote, the Planck law is a precise mathematical formula that relates the temperature of an object with the 'shape' of its spectrum. The light emitted from an object following a Planck law has a very distinctive shape (see the figure in the text).

c. Blackbody Radiation and the Planck Curve

The above description of thermal radiation is strictly true only for objects that emit completely continuous radiation. Any object that can emit only continuous radiation at all wavelengths, must also be able to absorb radiation at all wavelengths. This latter property must exist, since any light that hits an object is input energy, and that input energy must be redistributed according to the temperature of the object. Thus, an object that is a perfect emittor of radiation is also a perfect absorber of radiation. Such a perfect absorber of radiation would appear 'black' to the human eye when its temperature is so low that its thermal radiation has its peak in the infrared, where the human eye cannot see. A true blackbody will emit radiation that follows Planck's law exactly. The distribution of radiation following a Planck law is usually referred to as the Planck 'curve'. The radiation coming from stars, while not true 'blackbody' radiation, do follow a Planck curve reasonably well.

3. Spectral Lines and Continuous Radiation

a. If one spreads out (disperses) visible light from the Sun into its rainbow of colors, one observed both a continuous variation of color (the continuum) and dark lines in the spectrum (termed absorption lines).

b. If one disperses the light from a hot gas into its rainbow of colors, one observed both continuous radiation of bright lines (termed emission lines).

c. In Kirchoff's classic experiment, a hot gas and a continuum lamp can produce absorption lines + continuum, only emission lines, or only continuum, depending on how one views the gas with respect to the lamp:

i. If one looks at the continuum lamp through the cloud of hot gas, one will see a continuous spectrum with absorption lines in it at specific wavelengths.

ii. If one looks just at the hot gas, but not in the direction of the continuum lamp, one will see only emission lines at the same wavelengths as the absorption lines in (i), but no continuous radiation.

iii. If one only looks at the continuum source, one of course will see only continuum radiation.

d. Each atomic element has its own characteristic set of spectral lines, which can been seen either as an emission line or an absorption line). Molecules also have charateristic sets of spectral lines.

e. This property of atoms and molecules allows us to determine what a star is made of (it chemical composition) and to find out how fast the star is moving relative to us (how we do this is detailed later).

4. The Formation of Spectral Lines

Spectral lines are formed when there is a change in the amount of energy of a given atom or molecule. If the atom or molecule gains energy from the electromagnetic spectrum, it must remove that energy from the spectrum and, hence, form an absorption line. If the atom or molecule loses energy in the form of electromagnetic radiation, it must add energy to the spectrum and, hence, form an emission line.

In simplest terms, atoms are made of the three types of particles defined above: protons, neutrons and electrons. Protons and neutrons form the center of the atom, or nucleus. Electrons orbit this nucleus. An atoms that has as many protons as electrons is electrically neutral as a whole. Such atoms are called neutral atoms, and are the usual way atoms are found on our Earth. Note, however, that all neutral atoms are composed partly of charged particles.

Each element is determined by the number of protons and neutrons in its nucleus: A neutral hydrogen atom has one proton in the nucleus, and one electron orbiting the nucleus; a neutral helium atom has two protons and two neutrons in its nucleus, and two electrons orbiting; and so on up the table of elements (see, for example, the chart on the wall in PSF 101). Remember, there are only about 100 elements known - all matter as we know it in our everyday life is formed from these elements.

The absorption lines and emission lines we see in Kirchoff's experiment come from changes in the energy states of the electrons in the atoms of the gas, due to changes in the energies of their oribts. One of the fundamental discoveries of physics (made in the 1910's-1920's) was that electrons (and protons and neutrons and ...) can only have certain specified energies if they are part of an atom. These specified energies are usually called 'allowed' energy states. Electrons in atoms are simply not permiteed to have energies other than these allowed energy states. Each kind of atoms has its variation of a standard set of rules. The science that has developed to describe the atom is called quantum mechanics, since particles like electrons in atoms can only have energies in specific intervals, or quanta (the German term for interval).

The quantized energy states that can be occupied by an electron is an atoms are usually referred to as 'energy levels'. In every atom, the electron could occupy a wide range of specific energy levels. These energy levels are measured in units of energy corresponding to the energy of an electron, called an 'electron volt'. In a neutral hydrogen atom, the single electron can occupy energy levels of 10.15, 12.04, 12.69, 13.00, 13.16, ... electron volts, but not the energies in-between. Of course, the electron in the hydrogen atom can only be in one energy level at a time. The energy level occupied by the electron is what is meant by the term 'energy state' of an atom. Even in more complex atoms (that is, every other kind of atom) electrons can be in only one energy state at a given time.

Nature's rules governing how and when an electron can change from one energy level to another are lengthy and complex. However, one rule is simple: If an electron goes from a higher energy level to a lower level, it must emit an particle of electromagnetic radiation with an energy equal to the difference in energy to the levels. This particle of electromagnetic energy we commonly call a 'photon', and there is a direct relationship between the energy of a photon and the wavelength of this photon (the photon is a 'packet' of energy in the electromagnetic spectrum).

Remember the wave/particle duality of light: electromagnetic radiation can be described equally well as a particle (photon) or a wave (with wavelength and frequency). The photon possesses the energy of the wave. The higher the energy of the photon (or wave), the shorter the wavelength of the wave, and the higher the frequency of the wave. In terms of an equation,

Ephoton = hc/ = h ,

where h is another constant number called Planck's constant, which relates the wavelength of light to its energy.

For every change in energy of each electron in an atom, a photon of light will either be emitted (if the electron goes to a lower energy state) or absorbed (if the electron goes to a higher energy state). Obviously, the larger the number of electrons in an atom, the more the number of emission (or absorption) lines from that atom. Thus, an atom with 26 electrons (neutral iron) will have a much more complex line spectrum than neutral hydrogen, which has only one electron, or neutral helium, which has only two.

In an atom with many electrons, some electrons orbit close to the nucleus and, hence, are tightly bound. Some electrons orbit far from the nucleus and are loosely bound. An atom that has one or more of its electrons removed from it is called an ionized atom, or an ion. The amount of energy needed to remove an electron completely from an atom is also quantized, different for each electron in every atom, and termed the electron's ionization energy.

5. The Hydrogen Atom

The spectrum of the hydrogen atom is especially important in physics and astronomy both because it is the simplest atom that exists, and because hydrogen is the most abundant atom in the universe. Spectral lines that come from the transitions from the lowest energy state of hydrogen are called the Lyman series of lines (remember: since neutral hydrogen has only one electron, the lowest energy state of the atom is the lowest energy state of that one electron). Lines that come from the second-lowest energy state of hydrogen are called Balmer lines (the names refer to the people that discovered these lines in laboratory experiments in the 19th century). The text gives the names of other series of hydrogen lines, named after other people.

The appearance of the Lyman, Balmer or ... series of lines in a spectrum is relatively similar: There is a progression of lines from longer wavelength to shorter wavelength, with the spacing of the lines becoming closer as the wavelength of the lines becomes shorter. The first few lines of each series (which are the lines with the longer wavelengths) are given Greek designations, with 'alpha' given to the line with the longest wavelength, 'beta' to the next longest, etc. NOTE: The different series of hydrogen lines do differ in the wavelengths of their lines: the Lyman 'alpha' line has a wavelength of 1216 Ao; the Balmer alpha line has a wavelength of 6546 Ao, etc.

6. Collisional excitation, thermal radiation and continuous spectra

The temperature of an object measures how fast the atoms in that object are moving with respect to one another. If the atoms are close enough to each other, such as in a solid or a dense liquid, there will be many, many collisions among the atoms. Moreover, the interactions between atoms that make a solid greatly increase the number of energy levels that can be occupied by the electrons in the solid. The motions of atoms in a solid excite the electrons to higher energy levels (a process known as collisional excitation, which also occurs in gases). Electrons in the higher energy levels will fall to lower energy levels, and therefore release this energy in terms of electromagnetic radiations (photons). Each photon emitted by each atoms spans a small, but finite range in wavelength. The rate of collisions among the atoms depends strictly on the temperature (motion) of the atoms, which means the kind, and amount of radiation emitted from an object is dependent on its temperature - thermal radiation. Since even a tiny piece of sand contains over 1024 atoms, this thermal radiation will contain so many individual, overlapping lines that all regions of the spectrum will produce emission, and the strength of that emission will vary smoothly with wavelength - continuous radiation.

7. The Doppler Effect

The term Doppler effect, or Doppler shift is given to the physical effect that occurs when an object emitting electromagnetic radiation is moving with respect to the observer. A wave emitted by an object moving towards you will pass by more quickly, shifting the wave to shorter wavelengths, as seen by you, or towards the blue part of the spectrum. A wave emitted by an object moving away from you will pass by more slowly, shifting the wave to longer wavelengths, or to the red part of the spectrum. The speed of the object relative to you is directly proportional to the size of the shift in wavelength (as long as the speed of the object is much less than the speed of light). If the speed of the object is close to that of the speed of light, the relationship between observed wavelength and velocity is more complicated.

The Doppler effect is valid for any wave-like phenomena, including sound. You can hear the Doppler effect on campus anytime one of the little electric service vehicles with a beeper approaches you on the Malls: As the cart approaches, the beeps sound relatively high-pitched; as it passes you the sound chages to a lower pitch. Indeed, I suspect that the tone emitted by the beeper is so pure to permit the Doppler effect to let you know whether the cart is coming towards us or away from us. Many of you probably use this fact subconciously.

For planets, stars and the rest of the universe, the fact that emission lines and absorption lines are produced by atoms and molecules at very specific wavelengths permits astronomers to reliably measure the velocities that objects have either radially away or radially towards us.

8. Lasers and Masers

If you could get the outermost electrons in every atom in a gas or solid to be excited to the same energy level, and emit that radiation in the same direction, you can get all of the energy of that gas or solid into one extremely narrow emission line in a very narrow beam of light. Such an emission line would be tremendously bright. This is the concept behind a laser (light amplification by stimulated emission) and masers (microwave amplification by stimulated emission). Getting the electrons to move in such a coherent manner is called stimulated emission.

9. Cosmic Rays

Cosmic Rays are not really rays, but the nuclei of atoms that are moving at extremely high velocities (very almost at the speed of light, but not quite). What we see on the surface of the Earth is not the original particle, but the billions of secondary particles caused by the collision of the primary particle with atoms in the atmosphere (called cosmic ray 'showers').

We believe most cosmic rays come from extremely energetic events that occur in our galaxy, such as the violent deaths of stars. In such deaths, stars will emit particles of the very high energy needed to become cosmic rays. Cosmic rays of extremely high energy most likely come from some of the most energetic events in the universe!

In addition to the discussion in the book about cosmic rays, several years ago some atmospheric physicists suggested a connection between the shape of lightning and cosmic rays: They suggested that the path of charged particles emitted in a cosmic ray 'shower' (the 'muons') create a charged path in the atmosphere, which lightning discharges follow. If so, it could account for the 'tree-like' structure of lightning.

V. TELESCOPES

1. The principle purpose of a telescope is to collect as much radiation (be it electromagnetic radiation, cosmic rays, neutrinos or whatever) as possible. This is done by making the collecting area of a telescope as large as possible. The collecting area of a telescope is called its aperture.

a. Telescopes must be able to bring this radiation together. Telescopes that observe electromagnetic radiation need to bring the radiation together in one place, termed the focus of the telescopes.

The light-gathering power of a telescope is measured by the area of its aperture (that is, how big is the 'light bucket').

Magnification is the factor by which an obejct viewed with a telescope can be made to cover a wider angle than as observed by the naked eye.

2. Telescopes that detect electromagnetic radiation have two general methods for focussing light:

Refracting telescopes send the electromagnetic radiation through a lens, which bends the light towards the focal point (focus).

Reflecting telescopes reflect the light off of a curved surface, thus bringing the light to a central focus.

Note that refracting and reflecting telescopes do not have to be telescopes that work with visible light. Radio telescopes are reflecting telescopes, and there are infrared telescopes that use lenses that are opaque to visible light.

3. Kinds of foci (plural of focus)

The image produced by a telescope is the result of light from all parts of an object being focussed by the telescope. Most telescopes will focus an object in a plane that is perpendicular to the direction that light travels in the telescope. This is called the focal plane, and the term objective is used as the general name for the lens or mirror that produces the first refraction or reflection.

There are few fundamental restrictions on where the focal plane has to be located relative to the actual position of the objective. In a reflecting telescope (the most common type), the focus can be in front of the mirror (called the 'prime' focus), or it can be on the side of the telescope, parallel to the prime focus, called the 'Newtonian' focus, or it can be reflected back through a hole in the mirror, termed the 'Cassegrain' focus, or ... .

4. Associated Terms

Angular resolution: the smallest angle for which two objects can be seen as two distinct objects. The best angular resolution of a telescope, in the absence of an atmosphere, depends on the size of its aperture and the wavelength of light the telescope is trying to image. Angular resolution decreases with increasing wavelength, and increases with increasing aperture size.

Focal length: the distance between the objective and the focal plane. The increase in angular size of an image is determined only by the focal length of the telescope. Thus, a telescope with an aperture of 100" and a focal length of 800" will produce an image of the moon (or a planet, or ...) that has the same angular size as the image produced by a telescope with an aperture of 25" and a focal length of 800".

Image brightness: the amount of light contained in the image per unit area. In the above example, the 100" telescope concentrates (100/25)2 times = 16 times as much light into its image as does the 25" telescope. So, although the image of Jupiter will be the same size in both telescopes, it will appear 16 times brighter in the 100" telescope.

5. The Atmosphere and Telescopes

Transparency: In order to be able to detect stars with our telescopes, the atmosphere must be relatively transparent. Clouds obviously make the atmosphere very opaque, but even apparently clear days vary in the transparency of the atmosphere.

Seeing: Even if an atmosphere is clear, it still acts like a lens, and can shift the image of a star or planet relative to the observer on the ground. Such shifting occurs continuously in the atmosphere, and smears the image of a star into a much larger image. The degree to which this smearing takes place is called the seeing. Under the average conditions at Kitt Peak, the seeing image of a star is about 1.5 arcseconds in size; the best conditions in the world sometimes give seeing images of 0.25 arcseconds in size. The importance of this difference can be understood when you realize that a star has a true angular size of 1/10,000 of the seeing 'disk', so that reducing this image down by a factor of 6 concentrates the light of a star into a disk 36 time smaller and, hence, 36 times brighter.

6. Detectors for Optical Telescopes

Once the light is brought to a focus in a telescope, one must have a way of measuring the electromagnetic radiation. Until 1850, only the human eye registered visible light. In the 130 years since, we have come a long way in our ability to use all the light a telescope gathers.

Two things limit a detector: how much of the light that hits the detector is recored, and for how long can the detector record light.

The human eye records about 1 out of every 100 photons that hits our retina, and can store information for about 1/16 of a second (that is why the movies are 'moving').

Although a typical photographic plate can measure only 2 photons out of every 100, it can store information for hours, or up to 500,000 times longer than the human eye.

Photomultiplier tubes use the interaction of photons with the electrons in atoms to measure light, and can detect up to 10 photons out of every 100 photons, and can store information as long as a photographic plate.

Silicon diodes also use the interactions of photons with the electrons in atoms, and can detect up to 30 photons out of every 100.

Charge-coupled devices (CCDs) are basically more sensitive detectors made of silicon, and they can detect up to 60 photons out of every 100.

In terms of historical development:

up to 1850: the human eye
since 1850: the photographic plate, efficiency about 2%
since 1930: the photomultiplier tube, efficiency about 10%
since 1965: the silicon diode arrays, efficiency about 30%
since 1980: CCDs, efficiency about 60%.

Note the trend: over the past 130 years we have made our detectors more and more efficient. However, the best we could possible make a detector is 100% efficient. Over the past 130 years we have improved our efficiency by a factor of 30; for the coming millenia we can improve only by 50%!

7. Methods of Measuring Light

The electromagnetic spectrum can be measured in a variety of ways, all of them dealing with how finely we wish to divided up the spectrum. Terminology that is used also differs depending on the wavelength of the radiation:

a. Optical.

Spectroscopy: expand the spectrum and recored the amount of energy received at every wavelength between a beginning wavelength and an ending wavelength. The kind of instrument that can expand the spectrum is called a spectroscope. Spectroscopes vary according to the wavelengths of light that they expand (including X-ray, ultraviolet, optical, infrared, radio), and how much they expand the spectrum (their ’spectral resolution’).

Photometry: measure all of the energy contained in a range of wavelengths. Again, photometry can be done at all wavelengths. Photometry is more useful when one has to observe faint objects, for which expanding the spectrum would result in too few photons at each wavelength to produce a reasonable spectrum. Since photometry can measure much more total light, it is also used to get accurate measures of the total light coming from stars, planets and galaxies.

8. Astronomy by Wavelength

Astronomy is perhaps the only science in which most of the scientists identify themselves by the wavelength of light they most commonly use, rather than by the objects that they study. Thus, we have radio astronomers, infrared astronomers, optical astronomers, ultraviolet astronomers, X-ray astronomers and gamma-ray astronomers. The subcategories are the fields of study: for example stars or galaxies. Finally, planetary astronomy has recently separate itself from the rest of astronomy by one basic fact: the planetary astronomers have spacecraft that have physicially sampled the atmospheres and surfaces of planets. The rest of astronomy can only use the electromagnetic radiation (and, sometimes, cosmic rays and neutrinos) emitted by other objects to study them. A big distinction, as we shall see when we contrast what we now know about the Solar System with what we used to know before we sent up spacecraft.

VI. BASIC DATA FOR STARS AND HOW WE GET IT

A. Measuring Distances

1. Trigonometric Parallax and the Definition of the Parsec

The most direct way of measuring distances to stars is by direct triangulation of the star, using the Earth's orbit as a baseline. The principle behind this is: If one views a distant object from two different positions, the object will appear to move relative to a 'fixed' background (which, in reality, are simply other objects that are much, much further away from you). The apparent angle that the object moves is directly related to the distance between the two observing positions, and the distance one is from the object. (Even though this may be a new concept to you, it is not a new concept to your brain. Our eyes are separated on our face to give us 'binocular' vision to enable our brain to triangulate the distances of everything around us. All of this information is 'hard-wired' into the brain, so that we do not conciously think about it. This is also why optical illusions work: our hard-wired brain can be fooled by sending it information that looks correct, but is not.)

The angular movement of an object, due to the changing position of the observer, is called parallax. The straight line joining the two positions of the observed is called the baseline. Lines drawn from each position of the observer to the object make a triangle with the baseline. Since one uses trigonometry to deduce distances, this method of obtaining distances is usually called trigonometric parallax. (If you hold your finger up in front of your face, and altenately wink your eyes, the finger will appear to move relative to the background. This is an example of parallax. Try holding your finger at different distances from your eyes, and judge how much it appears to 'jump' at these different distances. The amount of the 'jump' is directly related to the distance of the object, since the distance between your eyes is the same.)

The largest baseline that can be reliably measured by observers here on the Earth is the diameter of the Earth's orbit (or, more precisely, the semi-major axis of the Earth's elliptical orbit). Since this particular baseline is the most direct measure of distances in astronomy, this baseline is given a special name, the Astronomical Unit. Stated another way, the Astronomical Unit is defined to be the semi-major axis of the Earth's orbit around the Sun.

In principle, any object in the universe has a parallax (since all objects we see are a finite, if sometimes very large, distance from us). In practice, we actually observe the parallax only of those objects that are relatively near us. In the case of our eyes, we can reliably judge parallax to a few hundred yards (Napolean used this fact in one of his battles to march his soldiers backwards towards his opponent. The opposing soldiers, seeing the backs of their enemy, but not being able to judge apparent motion because of the distance, thought Napolean's army was retreating. Napolean won the battle as a result.)

In the case of the Earth's orbit, we have to be more precise in our definitions. One can always think of the parallax angle in terms of being part of circle. Similarly, one can alway think of the Earth's semi-major axis (also known as the Astronomical Unit) in terms of being part of a the same circle. For a very 'skinny' triangle, the radius of the 'circle' is the distance of distance of the object. Following the notation of the book, if the parallax angle of an object is p, and the distance of the object is r, then putting the words above into an equation,

p/360o = (Earth's Baseline)/(2 r)   ,

because the circumference of a circle with radius r is 2 r.

If we rearrange this proportion, we get

r = 360o x (Earth's Baseline)/(2 p)

We could measure the angle p in units of degrees, but that would be very clumsy. It is much more convenient to measure p in terms of seconds of arcs. There are 60 arcseconds in an arcminute, 60 arcminutes in a degree, and 360 degrees in a circle. 60x60x360 = 1,296,000 seconds in a circle. We must then divide this number by 2 , and we get

r = 206,265 x (Earth's Baseline)/p"   ,

where p" indicates that angle is measured in arcseconds. Since Earth's Baseline is defined to be 1 A.U., this equation becomes

r = 206265/p" A.U.

An object with a parallax p" = 1 will have a distance of 206,256 A.U. The nearest star has a parallax of 0.75", and therefore has a distance of 206,265/0.75 = 275,020 A.U.

As always, having to cite big numbers is clumsy and, after all, this is the distance to the nearest star. If trigonometric parallax is the fundamental way we measure distances, it makes sense to define a distance based on this method. Define a parsec as being the distance at which an object will have a parallax of one arcsecond. Thus, a parsec has a length of 206,265 A.U. and, since 1 A.U. has a length of 149,597,871 km, and since 1 km has a length of 100,000 cm, then (whew!)

one parsec = 3.0857 x 1018 cm,

ten parsecs = 3.0857 x 1019 cm,

100 parsecs = 3.0857 x 1020 cm,

1,000,000 parsecs = 3.0857 x 1024 cm and so on.

The A.U. is the standard unit of length when studying the Solar System. The parsec is the standard unit of length for studying the rest of the universe.

All other methods for measuring distance in astronomy depend on one or more assumptions, either about the physical property of the object (for example, how bright is a star) or about the motions of objects (stars that are close appear to move faster). There are very few other methods of measuring the distances to stars that, like trigonometric parallax, involve only geometrical measurements.

2. Measuring Motions of Stars as We See Them

All of the stars that we see, including the Sun, are moving. Most of the motion of the stars we see is in terms of rotational motion around the center of the Galaxy. However, not all stars in the neighborhood of the Sun rotate around the center of the Galaxy with the same velocity (that is, not all stars have the same semi-major axis of rotation around the Galaxy). Some stars move faster than the Sun, some move more slowly. In addition, the gravitational attraction of nearby stars can change the motions of stars over the lifetime of a star. All in all, if we observe the positions of stars over many years we see that their relative positions in the sky change. The 'fixed' stars are not really fixed in space.

The three-dimensional motion of a star can be thought of in terms of two kinds of motions:

Proper motion: If a star changes position in the sky, relative to more distant objects (that is, either much fainter stars or even more distant galaxies), the direction and size of change with time, as it appears to us, is called 'proper motion'. The proper motion of a star is usually given in terms of its motion in right ascension and its motion in declination. Combining these two parts of its motion gives its total proper motion.

Radial Velocity: If a star is moving straight towards us, or straight away from us, it will have no proper motion, but it will have a measureable Doppler shift in its light. Since the Doppler shift can measure only motion towards or away from you, on a 'radial' line between you and the object, this velocity of a star is called its radial velocity.

The measurement of proper motions is limited by the angular resolution of the telescope and detecting equipment (such as a CCD). The measurement of radial velocity is limited only by the spectral resolution of your spectrograph. Thus, one can measure radial velocities much more easily that one measures proper motions.

If one knows the distance to a star, one can convert the measurement of proper motion into a transverse or 'tangential' velocity. If we express distances in terms of parsecs, and proper motion in terms of arcseconds, then

$$\text{Tangential Velocity} = 4.74\times(\text{proper motion} \times \text{distance}) \quad \text{km/s}$$

$$= 4.74\times(\text{proper motion}/\text{parallax}) \quad \text{km/s} \quad ,$$

where proper motion is measured in terms of arc seconds/year. The number 4.74 km/s is the speed of an object at a distance of one parsec, that gives the object a proper motion of one arscecond in one year.

Finally, if one knows the distance to a star, its proper motion and its radial velocity, one can determine its full 'space' velocity, V, by the equation

$$V^2 = V^2_{radial} + V^2_{tangential} \quad ,$$

where Vradial is the radial velocity of the star, and Vtangential is the tangential velocity of the star.

(In thinking about proper motion, radial velocity, tangential velocity and space velocity, it might be helpful to think of the motion of an airplane as you see it. The apparent motion of the plane in the sky is its proper motion. How fast the airplane comes towards you or away from you is its radial velocity. How fast the airplane is actually moving depends on its distance from you. In the case of an airplane, most of us know how big is an airplane, so we judge its distance based on its size.)

3. The Solar Motion

If the Sun and all of the surrounding stars move relative to one another, we have to think of a rational way of describing all of these motions. At the bare minimum we would like to know if the apparent motion of a star is due to that star moving, or due to the Sun moving relative to that star, or to a combination of the motions of both the Sun and that star. (Confusing? How many of you have ever been in a bumper car rink? It is easy in that situation to confuse your motion for the other car's motion, and vice-versa. The scene only makes sense to someone looking from the sidelines who is at rest relative to the average motion of the cars.)

Since we reliably measure trignometric parallaxes to distances of 100 pc or less by present techniques, it makes sense to define an average of motion of stars that are within that distance (including the Sun). This average motion is called the 'local standard of rest', in that the net relative motion of all stars is zero within this volume. This does not imply that this volume is not moving in space, but only that, on average, all stars in this volume have the same motion. (Using the bumper car analogy, all of the bumper cars together have a net motion of zero relative to outside observer, but, since the Earth rotates on its axis, and revolves around the Sun, the bumper cars still share much larger motions.)

In general, each star has a motion relative to the local standard of rest (just like each bumper car has a relative motion). The Sun is no exception, but, since its motion is our motion, the motion of the Sun occupies a special place in astronomy. The motion of the Sun relative to the local standard of rest is called, appropriately enough, the 'solar motion'. This motion is 20 km/s, or 4.14 A.U./yr. The direction towards which Sun appears to move, relative to the local standard of rest, is called the 'solar apex', and the opposite direction is called the 'solar antapex.'

With the motion of the Sun specified, the motions of other stars relative to the local standard of rest can be determined, if their distances are known. Since this kind of motion is relative to a 'standard', it is termed 'peculiar' velocity. Note: The calculation of peculiar velocity involves both a size and direction.

B. Measuring Light from Stars

The rest of the universe communicates with us by means of electromagnetic radiation. Almost everything we learn about the rest of the universe comes from interpreting the light coming from objects at all wavelengths (from gamma rays to radio waves). As such, it is very important in astronomy to have precise measurements of how much light we see coming from an object.

1. The Magnitude System

Our eye does not respond to differences in the brightness of objects in a linear manner. Rather, the eye (and the ear, for that matter) responds in a logarithmic manner. That is, the eye responds to ratios of brightness.

Since the eye measured brightnesses of stars until 1850, all of the measurements of brightness were in terms of ratios. Following the ancient Greek terminology, these ratios of brightness were called 'magnitudes', with a star of the 'first' magnitude being brighter than a star of the 'second' magnitude, a star of the 'second' magnitude being brighter than a star of the 'third' magnitude, and so on. Unfortunately, we must still use this system of measurement today, as all astronomical data over the past 2000 years is in this form.

The magnitude system is a system of ratios, or a logarithmic system. A difference of one magnitude is equal to a ratio of 2.512:1; a difference of two magnitudes is equal to a ratio of 2.512x2.512 = 6.3:1; a difference of three magnitudes is equal to a ratio of 2.512x2.512x2.512 = 16:1. Table 23.1 of the book gives a list of how ratios correspond to differences in magnitudes.

Note something about ratios: The ratio 16:1 is the same as the ratio 320:20. The only difference is that the numbers in the ratio 320:20 have values that are 20 times the numbers in the ratio 16:1. However, in both cases the corresponding difference in magnitudes is three.

One thing we inherited from the ancient astronomers is the manner in which the magnitude system is related to luminosity. Brighter stars have smaller magnitudes, since a star of the 'first' (=1) magnitude is brighter than a star of the 'second' (=2) magnitude.

2. Measuring Brightnesses of Stars

a. How do we do it?

The brightness of a star depends on what wavelengths of electromagnetic radiation you use to observe it. Our eyes see a certain range of optical wavelengths, a photographic plate sees a different range of optical wavelengths, and a CCD can see a still different range of infrared wavelengths. These are only three of many examples that can be used. We can measure certain ranges of wavelengths by using glass filters in the optical, and other kinds of filters for other wavelengths of light. Some of the optical filters are similar to those in high-priced sunglasses (which typically block out E-M radiation with wavelengths shorter than 4000 Ao.

Since the magnitude system is based on the ratios of brightnesses, all measurements are relative. An 'absolute' system of measurement is needed so we can discuss the brightnesses of stars with other people. There are two parts to setting up such an 'absolute magnitude' system:

i) We all have to agree that certain stars, measured in the same way in the same wavelength region, have the same magnitude as we observe them from Earth. (This kind of magnitude is called 'apparent' magnitude, because this is how bright the stars 'appears.') This is done by choosing a set of 'standard' stars, and someone spending many years measuring these stars very accurately. By convention, astronomers have defined the apparent magnitude system so that the 'first' magnitude stars of the ancients have magnitudes in the visual (that is, how we would see them with our eyes) of about 1.00.

ii) Not all stars are at the same distance from us. In fact, almost all stars are at different distances from us. If we want to compare the brightnesses of two stars on an absolute basis (say, like comparing the brightnesses of a 25 watt bulb with a 100 watt bulb), we would have to place these two stars at the same distance (due to the inverse-square law of light). We can do this with light bulbs on Earth, but we can't do it with the stars in the sky.

Instead, we use the inverse-square law of light to our advantage. If we define a certain distance as being the distance at which we measure absolute magnitudes, then we can, in our mathematics, ask what the apparent magnitude of a star would be if we moved it to that particular distance. If we do that for all stars (or galaxies, or planets or ...), we can then define the absolute magnitude of a star as the apparent magnitude the star would have if it were at that particular distance. The particular distance is defined to be a distance of 10 parsecs.

Thus, the absolute magnitude of a star is defined as the apparent magnitude a star would have if it were at the distance of 10 parsecs. By convention, apparent magnitudes are ususally denoted by a little 'm', with a subscript to denote what range of wavelength is being measured. Absolute magnitudes are usually denoted by a capital 'M', again with a subscript to denote what range of wavelength is being measured.

b. The relation between apparent magnitude, absolute magnitude and distance.

Remember:
1) Differences in magnitudes correspond to ratios of brightness;
2) The inverse-square law relating apparent brightnesses of the same object at different distances; and
3) Absolute magnitude is defined as the apparent magnitude a star would have it were at a distance of 10 parsecs.

If we combine all three concepts, then the difference between the absolute magnitude and apparent magnitude of any star depends just on the distance of the star (ignoring other factors, such as interstellar 'fog'), in the following manner:

Suppose star Eo is at a distance of Q (where Q stands for any distance that I would like to use), and has an apparent magnitude mEo. If we choose to call the absolute magnitude of star Eo to be MEo, then

mEo - MEo = magnitude difference corrsponding to ratio of (Q/10)2 .

(The more rigorous definition of the right hand side of this equation is 5log(Q/10), as shown in the footnote on pg. 432 of the book. This term is called the 'distance modulus'. It may be easier for those of you who are comfortable with logarithms to remember the formula for distance modulus.)

Three examples of calculate mEo - MEo:

i) If Q = 10 parsecs, then (Q/10)2 = (10/10)2 = (1)2 = 1.

The magnitude difference corrsponding to a ratio of 1 is 0. In this case mEo = MEo

ii) If Q = 50 parsces, then (Q/10)2 = (50/10)2 = (5)2 = 25.

Using Table 23.1, the magnitude difference corrsponding to a ratio of 25 is 3.5. In this case, mEo is 3.5 magnitudes fainter than MEo. If mEo is

7.26, this means that MEo = 3.76. If mEo = -0.34, this means that

MEo = -3.84.

iii) If Q = 160 parsces, then (Q/10)2 = (160/10)2 = (16)2 = 256.

Using Table 23.1, the magnitude difference corrsponding to a ratio of 256 is very close to 6.0. Let us assume that it is 6.0. Then mEo is 6.0 magnitudes fainter than MEo. If mEo is 9.57, this means

that MEo = 3.57. If mEo = 18.12, this means that MEo = 12.12.

Now generalize, and call the apparent magnitude of a star m, its absolute magnitude M and its distance D. If you know m and M, you can calculate D; if you know m and D, you can calculate M; if you know M and D, you can calculate m. You have to know at least two of these quantities to calculate the third.

In almost all cases, we measure m, we estimate M, and thereby estimate D.

c. Colors and Bolometric Magnitudes

Remember the discussion of the relationship of color and temperature for a blackbody. Stars do not radiate precisely like a blackbody, but they do share some of the characteristics of a blackbody: Stars which put out more red light than blue light ('red' stars) are cooler than stars which put out more blue light than red light ('blue' stars). The 'color' of a star is then a useful property to measure. We do this by measuring the magnitude of a star over a specific range in wavelength, say from 4000 - 4800 Ao, and compare this to the magnitude of a star over another range in wavelength that is redder, say from 5000 - 5800 Ao. A 'red' star will have a brighter 5000-5800 Ao magnitude than a 4000-4800 Ao magnitude; a 'blue' star will have the opposite.

If we want to measure all of the light coming from the star, we have to measure its light at all wavelengths. This kind of measurement is called the bolometric magnitude of a star.

d. Stars that vary in Luminosity

Certain kinds of stars vary in their brightness. Some vary in a regular way, some vary in a haphazard way. The ones that vary in a regular way do it by literally pulsating: first growing bigger and cooler, then shrinking and becoming smaller and hotter. The variation in time of the luminosity of a star is called its 'light curve'.

Why discuss variable stars? Because certain types of pulsating variables, called 'Cepheid stars' and 'RR Lyrae stars' have their period of pulsation related to their absolute magnitudes (the 'period-luminosity' relationship). Since we can measure the period of pulsation without first knowing their distances, we use this period to predict their absolute magnitudes. Combining their absolute magnitudes with their apparent magnitudes, we then know their distances. Along with trigonometric parallax, pulsating stars are one of the primary means of measuring distances in the Universe.

C. Stellar Spectra

The spectra of tens of thousands of stars were available for study thirty years before the structure of the atom was beginning to be understood. Astronomers, especially Annie Jump Cannon of Harvard University, arranged these spectra by the strengths of their hydrogen Balmer absorption lines (which were known at that time to come from the element hydrogen). The stars with the strongest Balmer absorption lines were termed A type, the next strongest B type, etc. These categories are called spectral types or spectral classes.

After nearly 100,000 stars had been classified, it was discovered that the color of a star was related to its temperature (at the end of the nineteenth century, after Planck introduced his 'law'). The spectral types of stars were then arranged in terms of the temperature of the stars, rather than the absorption-line strengths of the Balmer lines. In the process, most spectral classes were consolidated into a few classes.

The modern spectral classes are arranged from the hottest stars to the coolest stars, with the corresponding spectral types being: O,B,A,F,G,K,M. (The reason that the strengths of the Balmer lines do not correspond strictly to temperature is that hydrogen becomes ionized in the hottest stars.) The spectrum of a star is determined by a thin outer layer of the star's gaseous atmosphere. The interior of a star is hidden by this atmosphere, much as the crust of the Earth hides the interior of the Earth.

The hottest stars are the O stars, with surface temperatures ranging from 25,000 - 60,000 oK. B stars have a range of 11,000-25,000 oK; A stars a range of 7,500-11,000 oK; F stars a range of 6,000-7,500 oK; G stars a range of 5,000-6,000 oK; K stars a range of 3,500-5,000 oK; and M stars a range from about 2,000 -3,500 oK. Spectral classes are divided into tenths, and order with 0 being the hottest of the class and 9 being the coolest. So, an F0 star is hotter than a F5 star, which is hotter than a F9 star, which is hotter than a G1 stars, which is ... .

The spectra of stars are used to:

1) get the temperature of the star
2) measure the pressure in the atmosphere of the star (which, in turns, measure the gravity of the star)
3) measure the kinds of atoms and molecules in the stars atmosphere (termed chemical composition).
4) measure the radial velocity of a star
5) measure how fast the star rotates
6) measure the strength of the magnetic field of a star
7) detect ejection of gas from a star

D. Binary Stars: Getting Masses and Diameters of Stars

About half of the stars in our galaxy are part of gravitationally-bound binary or multiple star systems. This means that the stars are in mutual orbit around each other. How we view a binary system from the Earth depends on three things: a) how far apart the two stars are at a given time (remember, they are in orbit around each other); b) how far the two stars are from the Earth; and c) what is the inclination of the plane of the stars' orbit relative to the angle from which we view them. If a true binary star is close enough for us to see both stars, it is called a visual binary. If it is too far for us to see two stars, but we take a spectrum and we see the spectra of two stars, it is called a spectroscopic binary. If the plane of the orbit of a binary is oriented so that we see one star go in front of the other (that is, one star eclipses the other), this is called an eclipsing binary.

If we can measure the period and semi-major axis of the orbit of a binary star, we can, by Kepler's third law, get the sum of the masses of the two stars. If we can further determine the center of masss of the orbit, we can get the relative sizes of the two masses and, thus, get the mass of each star. This is the fundamental way masses of stars are measured.

When we measure masses of stars in this manner, we find that the mass of a star is related to its absolute luminosity. This is an important clue to understanding how stars form and how they evolve with time.

Eclipsing binary stars can also be used to determine the diameters of stars, simply by timing how long it takes one star to eclipse the other star. Another way of measuring diameters is to use the absolute luminosity and temperature of a star (measured by the colors of the star), since the total luminosity of a star is

L = 4 R2 T4 .

E. Summary of Basic Data for Stars

We have gone over the way in which astronomers measure absolute luminosities of stars, the temperatures of stars, their masses and their absolute sizes. We now need to combine this information in a meaningful way, one which will give use clues as to how stars are born, how the evolve with time, and how they die. How we do this is the story of the following sections.

VII. STELLAR POPULATIONS

A. Finding the Luminosity Function

The brightest stars in the sky are not necessarily the nearest stars to our Sun. Suppose we live in a region of space that can only have stars that have absolute magnitudes either 5 or 0. If I look at a star with an apparent magnitude of 5, it would have to be either a star of absolute magnitude 5 at a distance of 10 pc, or a star of absolute magnitude 0 at a distance of 100 pc. But the volume of space out to 100 pc is (100/10) x (100/10) x (100/10) = 1000 times larger than the volume of space out to 10 pc. So, if there were as many stars of absolute magnitude 0 near the Sun as stars of absolute magnitude 5, stars of apparent magnitude 5 would be almost all stars of absolute magnitude 0.

We can generalize this situation, and put in stars with a full range of absolute magnitudes, at a range of distances from the Sun. We would always see that stars at any apparent magnitude will tend to be mostly (but not completely) stars with the brighter absolute magnitudes at larger distances from the Sun.

The effect of this ’bias’ in the way we select stars is that, if we want to find all of the stars that are near us, we have to look at stars with faint apparent magnitudes. This is difficult for obvious reasons (there are simply too many faint stars to measure the parallax of each star), so that nearby stars are usually found by looking at their ’proper motion.’

If we take a certain volume of space, say 100 cubic parsecs, we would like to know how many stars there are in terms of their absolute magnitudes. We would want to know how many stars there are in this volume with absolute magnitudes between 15.0−14.5; 14.5−14.0; 14.0−13.5; ... ; 1.5−1.0; 1.0−0.5; 0.5−0.0; 0.0 to −0.5; −0.5 to −1.0; and so on. This ’distribution’ of stars with absolute luminosity is called a ’luminosity function.’ When we do this, we find that most of the stars near the Sun have fainter absolute magnitudes than the Sun. Since the mass of stars varies only slowly with absolute magntiude, this means that faint stars have most of the mass near the Sun.

However, since one star of absolute magnitude 0 puts out the luminosity of 1,000,000 stars of absolute magnitude 15, we also find that the total light put out by this volume of space is dominated by the fewer, brighter stars.

B. The Hertzsprung-Russell Diagram

The Hertzsprung-Russell Diagram (or H-R diagram, for short) is one of the most fundamental tools of astronomy. Why?

Around 1910 Hertzsprung (in Denmark) and Russell (in the US) independently plotted a new kind of diagnostic diagram for stars. The spectral classes had been recently revised to the modern, temperature system; and a number of the nearer stars had measured parallaxes, which meant measured distances and measured absolute magnitudes. Hertzsprung and Russell independently realized that a graph of absolute magnitude versus spectral type would distinguish stars of large absolute diameter from stars of smaller absolute diameter, and would also show how the absolute luminosity of a star is related to its temperature.

Today, any diagram which plots a temperature-related quantity on the horizontal axia (such as spectral type; temperature; color) and a luminosity-related quantity on the vertical axis (absolute magnitude in any color, or bolometric absolute magnitude) can be called an H-R diagram.

By graphing absolute magnitude versus temperature, one can discover the sizes of stars. In a real sense, the size of the star is the 'hidden' quantity of the H-R diagram. This fact comes again from the equation:

L = 4 R2   T4 .

L is measured by absolute magnitude; T is measured by spectral type. If two stars have the same R, L and T, they will fall in the same place in the H-R diagram. If two stars have same L and T, but different values of R, the larger star will have a brighter absolute magnitude and, hence, the larger star will be plotted higher in the H-R diagram than the smaller star, but both stars will have the same temperature. By convention, an H-R diagram is made by having temperature on the horizontal axis and increasing from right to left as you view the graph, and absolute magntiude on the vertical axis increasing from the lower part of the graph to the upper part.

If we plot all of the nearest stars on the H-R diagram, we find that stars do not have random sizes, luminosities or temperatures. What we find is that most stars lie in a main sequence in the H-R diagram. Such main sequence stars are termed dwarf stars, and the Sun is considered a G2 dwarf main sequence star.

Some stars are seen to be 3-6 magnitudes brighter than main sequence stars of the same temperature. These stars must have larger diameters than the main sequence stars, and so they are called giant stars. An even fewer number of stars are found be 10-15 magnitudes brighter than main sequence stars of the same temperature, and even 5-10 magnitudes brighter than giant stars at the same temperature. These stars are called supergiant stars. Finally, some of the hotter stars are found to be 10 magnitudes fainter than main sequence stars of the same temperature. These stars must be of much smaller size than main sequence stars. Since these stars tend to be hot, they are given the name white dwarf stars.

VIII. GAS, DUST AND THE STRUCTURE OF OUR GALAXY

A. As through a fog (or on a smoggy day you can see ...)

The Milky Way is such an obvious feature of the night sky that all ancient socities included it into their cosmologies. It was not until the invention of the telescope, however, that the Milky Way was shown to be made up of literally millions of individual stars. Thus, when Herschel counted the apparent density of stars (that is, the number of stars per square degree) on the sky in the late 1700's ('star-gauging'), it was logical that the highest density should be towards the Milky Way, but he also found the lowest densities of stars to be in directions that were perpendicular to this band. Herschel therefore concluded that the stars around us were arranged in a disk-like fashion, and that the band of stars we call the Milky Way was simply our viewing this disk along its long axis.

However, Herschel, and many who followed him, thought that this disk was about 500 pc thick and 2000 pc in diameter. It was not until 150 years later, in the 1930's, that astronomers realized that the small apparent size of this disk was the result of obscuration of dust in our galaxy. This is much like trying to examine the mountains that surround Los Angeles while looking through the Los Angeles smog. The smog can be so bad that the mountain cannot be seen. The effect of smog is much like the effect of dust in our galaxy: it cuts down on the amount of light we get (an effect astronomers call 'extinction') and it makes that light appear redder by selectively absorbing blue light (an effect we call 'reddening'). Measuring how much dust is along a line of sight in astronomy is very tricky, since we know a priori neither the amount of dust there should be, or what we should see in the absence of this dust. The fact that dust reddens light provides one of the main clues of how much dust is present.

Once we account for the obscuring effects of dust, we can better define the structure of our Galaxy, the disk-like system of stars, gas and dust of which the Sun is one star among billions of stars. This structure will be discussed at length in future lectures. The concept is introduced now in order to introduce the concept of an 'interstellar medium', namely the gas and dust that exists 'between' the stars. As we will see, stars are born when this gas and dust collapses under the force of gravity. The evolution of stars returns much of this gas back into the interstellar medium, after 'processing' some of the gas through nuclear fusion reactions in the centers of the stars. Both the births and deaths of stars can be noisy, energetic events, releasing energy to churn up the gas and dust around the stars.

As a result, our Galaxy is a very active place, with a constant parade of stars being born, living and dying; with dust being created and destroyed; and with dust and gas being blown around in sometimes regular, and sometimes chaotic ways. All in all, our Galaxy is constantly changing in detail, but remaining similar in general: a dynamic place. We simply cannot see this dynamic behavior during our lifetimes, since it takes place over periods of 1000's of years to million's and billion's of years.

B. The Interstellar Medium

On average in the interstellar medium there is one atom per cubic centimeter. There are regions that are much denser than this, and regions that are of much lower density. For reference, a cubic centimeter of air at sea level has about 1024 atoms, and the best vacuum that can be made on the ground has about 1016 atoms/cubic centimeter. Yet, even at these low densities, our Galaxy is large enough that interstellar gas and dust make up about 10% of the mass of our Galaxy: The size of our Galaxy compensates for the low density of gas.

1. Hydrogen

Most of the atoms in the interstellar medium are hydrogen: hydrogen is present as a gas in the form of single atoms, separate protons and electrons, molecules of hydrogen (two hydrogen atoms chemically bound together), and hydrogen combined with other elements (for examples, water, ammonia, methane). About 90% of all atoms in the interstellar medium are H; about 9% are He, and all other elements make up the remaining 1%.

Astronomers have separate names for the three forms of hydrogen found in the interstellar medium:

a. Neutral Hydrogen, or HI - hydrogen gas in which the atoms are neutral

b. Ionized Hydrogen, or HII - hydrogen gas in which the atoms are ionized (that is, the electron in an H atoms has been separated from the proton).

c. Molecular Hydrogen, or H2 - two atoms of H chemically-bound together.

Terminology: In astronomy, the neutral form of an element is designated as a 'I' (Roman numeral one), an atom once-ionized is 'II', an atom twice-ionized (that is, two electron separated from the atoms) is 'III', an atom three-times ionized is 'IV' and so on.

Hydrogen can be ionized either by collisions of atoms (requiring gas of fairly high densities), or by electromagnetic radiation of wavelenths shorter than 912 Ao. All stars produce such short wavelength radiation, but, from Planck's Law, the hottest stars produce the most amount of short wavelength radiation. Bright, hot stars are therefore usually surrounded by regions of ionized hydrogen, or 'HII regions'. As a result, all other kinds of regions of hydrogen are called 'HI regions'.

Most of the hydrogen atoms in our Galaxy are in the form of HI. In its lowest energy state, the electron in an hydrogen atom can emit a photon of wavelength 21 cm about once every 10 million years. However, since there are about 1056 hydrogen atoms in our Galaxy, about 1040 hydrogen atoms are emitting 21 cm radiation every second! 21 cm emission is the strongest radio emission line that can be observed from our Galaxy. Since it is an emission line, we can measure the motion of the HI gas from its Doppler shift.

2. Dust

When many different molecules 'stick together', the result is generally termed 'dust.' We think that most of the dust in the interstellar medium is made of ices that have coated smaller balls of 'dust' (made up of silicate and carbonate molecules), much like frost forms on a window. The major kinds of ices in the interstellar medium are water, methane and ammonia. The ratio of atoms in dust to atoms in gas is about 1/100.

Dust affects how we view the Galaxy in many ways. For this class, we will be concerned primarily with four of these:

a) extinction - the absorbing and scattering of light along the line of sight, about 1 magnitude per kpc of distance.

b) reddening - selectively absorbing blue light more than red light

c) scattering - displacing the path of light; you always appear to be in the center of a fog because the fog particles scatter the surrounding light. Light from a star can be reflected off of dust, resulting in a 'reflection nebula.'

d) absorption - absorbing radiation, thereby increasing the motion (temperture) of the dust grain

Dust appears to be formed either in very cold regions of space, or when the properties of the gas are changing relatively rapidly, such as in a shell of gas that is expanding from a star.

3. Molecules

Many of the atoms in the interstellar medium are found as individual atoms, since it requires either high densities or very cool temperatures to form molecules. However, there are regions of very cool, relatively high densities in which molecules can be formed. Quite naturally, astronomers call such regions 'molecular clouds'. By some mechanism, still not well-understood, these molecular clouds grow to be as large as 50 pc in size, with masses of 10 million solar masses of gas. Latest estimates indicate that the total mass of molecular gas is about 50% that of the atomic gas.

Molecular gas and dust are formed under the same physical conditions, so it is not surprising that dust and gas are well-mixed. This directly affects how we can study the formation of stars:

Stars must form out of what is initially very dense, very cold gas - molecular gas. Such molecular gas is always accompanied by large amounts of dust, which therefore obscure the formation of the star. However, the thermal energy of the dust is related to how stars form, and this can be observed at infrared wavelengths. It has only been over the past decade that both observational techniques, and infrared telescopes in space (IRAS, for example), have permitted us to 'view' the process of stellar birth.

4. Absorption Lines and Emission Lines from Gas

A cool gas in front of a hot star will produce absorption lines. The stellar absorption lines in the star are produced by cooler gas on the surface of the star. However, cool gas in the interstellar medium also produces absorption lines in the spectra of stars. The interstellar absorption lines are different from the stellar absorption lines in two ways: interstellar absorption lines are narrower in wavelength range and are usually doppler-shifted in wavelength from the stellar lines.

Interstellar gas in general is too cool to produce emission lines. However, the 'recombination' of hydrogen atoms in HII regions produces emission of radiation corresponding all of the transitions of the electron, among them the Paschen series, Balmer series, Lyman series. This kind of emission is called 'flourescence.'

IX. The Sun as a Star

The Sun has a diameter of 800,000 miles and a mass of

2,000,000,000,000,000,000,000,000,000,000,000 grams, or 1033 gms.

1. Basic Structure

a. Outer Layers:

Photosphere: the region of the Sun that we see as the disk of the Sun, since it is not transparent to radiation coming from the interior of the Sun. It has an average temperature of 5,500 oK, a thickness of about 250 km, and a higher temperature and density with increasing depth in the atmosphere.

Chromosphere: the region of the Sun that lies just above the photosphere, about 3000 km thick. This region is hotter (up to 100,000 oK) but is much less dense than the photosphere and relatively transparent to most radiation. Emission lines seen in Solar eclipses come from this region, and their various colors gave the name to this region.

Corona: The region above the chromosphere that can extend for more than a million miles from the photosphere, has a temperature of 1,000,000 oK or more, and is very tenuous. This high temperature of the corona gives rise to a 'solar wind' of fast atomic particles being emitted from the Sun.

b. Phenomena observed on the Surface:

As we see it, the surface of the Sun is in constant motion. The Sun rotates about an axis, but not as a solid body. It rotates faster near the equator, and slower near the poles. Since the polar axis of the Sun is nearly perpendicular to the ecliptic, we see very little of the polar regions of the Sun.

The most notable features on the surface of the Sun are sunspots. Sunspots are areas of lower temperature in the photosphere (typically 1500 oK cooler than the surrounding surface), and hence appear darker (from Planck's law). However, if you could remove a sunspot from the surface, it would appear to be a bright orange in color. Sunspots only appear small by comparison with the rest of the Sun; a typical sunspot is many times the diameter of the Earth.

If we look more closely at the surface, we can see that it is not smooth, but divided into 'cells' that give the surface the appearance of 'granulation.' Each cell is the result of motion within the region of the Sun below the photosphere (see discussion of convection, below).

The Sun has a strong magnetic field which influences many of the features we see. As discussed in the text, we know that this magnetic field causes sunspots, as well as many other very energetic phenomena. The source of this magnetic field is poorly understood.

2. The Solar Cycle and possible other variations

The magnetic activity of the Sun has a regular variation with a period of 22 years. Every 11 years the polarity of the Sun's magnetic field reverses itself. At the beginning of each 11 year portion of the cycle, the magnetic field is strong, many sunspots are produced, and the Sun is very active on its surface. This activity increases the strength of the Solar Wind, which in turn can affect the upper atmosphere of the Earth. As the cycle continues, the magnetic field gradually becomes weaker, and the number of sunspots gradually becomes smaller. In the middle of the 11 year cycle, about five years after it begins, the magnetic field and sunspots all but disappear from the surface. This is called the period of the 'quiet' Sun. After this period of the quiet Sun, the magnetic field grows in strength again, sunspots increase in number, and the polarity of the poles have reversed. At the end of one 11 year cycle, the magnetic poles have reversed on the Sun. It takes a total of 22 years for each pole to return to the same polarity, with each 22 year cycle containing two 11 year cycles of 'active'-'quiet'-'active'-'quiet'-active.'

Could the Sun have other, longer term variations? We only discovered the 22 year Solar cycle 200 years ago. If longer term variations of the Sun's luminosity exist, we may not have collected enough data to measure them.

The evidence of possible longer-term variation centers on a 50 year period in Europe from 1650-1700, when the climate was much colder than usual, and apparently few sunspots were observed. This 'Maunder' minumum in sunspots has been suggested by some solar astronomers to signal a somewhat cooler Sun. A change of just 1-2% of the Sun's luminosity can be the difference between an ice age or severe heat on the Earth.

Very recent studies of the luminosity of the Sun again point out how naive interpretations in astronomy are often incorrect: The discussion of the effects of sunspots on the Sun's luminosity on pg. 513 correctly points out that increased numbers of sunspots should lower the luminosity of the Sun, all other factors being equal.

However, all other factors are not equal. Studies of the solar luminosity made over the past five years (or after this section of the book was written) show that the average luminosity of the Sun increases when the number of sunspots increases! How can this be? Remember that sunspots are simply one result of the Sun's magnetic field. Since they are cooler regions, they stand out more readily than regions of higher temperature. Evidently, the total energy output of the Sun increases when its magnetic field is stronger. Thus, it becomes more plausible that the Maunder minimum in sunspots corresponded to an overall lower luminosity of the Sun, resulting in colder weather here on the Earth. This leaves a very open question, namely are the Ice Ages the result of very long-term variations in the luminosity of the Sun?

3. The Interior of the Sun - The Structure of a Star

All matter has to contend with the force of gravity: Gravity wants to bring everything together to a central point. The more matter, the stronger the force of gravity. The force of gravity inward must be countered by a pressure outward. It takes energy to generate this pressure. In the case of the Earth, the chemical energy from atoms is sufficient. In the case of the Sun and all stars, nuclear energy must be used. Remember: the force of gravity is always present. Stars must find ways to generate an outward pressure, or they will collapse to higher and higher densities until they find a way to generate a sufficient outward pressure.

Certain concepts are important in understanding the structure of a star:

a. Hydrostatic equilibrium (or, pressure balance): The pressure exerted by a gas is the result of its temperature (that is, how fast are the atoms and molecules moving) and its density (how many atoms are moving per cubic centimeter). For example, the photosphere and chromosphere of the Sun are in hydrostatic equilibrium, since the hotter chromosphere has a lower density.

b. Opacity: Opacity is the technical term that refers to what percentage of radiation at a specific wavelength is absorbed by the atoms and molecules of a mass. For example, ordinary glass absorbs very litte visible light (that is why we can see through it), but glass absorbs most ultraviolet radiation. The opacity of glass is therefore small in the visible, but large in the ultraviolet. Absorption of radiation in stars is usually due to electrons either bound to atoms or separate from atoms. If bound to an atom, an electron can only absorb radiation at specific wavelengths (quantized absorption); if free, an electron can absorb radiation at any wavelength (continous absorption).

c. How Energy Moves: Energy can be transported from place to place in one of three ways that are familiar to everyone:

i) Radiation - Energy can be moved by electromagnetic radiation in the form of photons. The heat from a lamp, or the Sun on our skin, are examples of the transfer of energy by radiation.

ii) Conduction - The hotter something is, the faster the atoms and molecules are moving in that object. If one places a hot iron next to a cooler finger, the finger will get very hot because the atoms in the iron are moving very fast, and collide with the atoms in the finger. Conduction is the name given to the transfer of energy by direct contact of moving atoms.

iii) Convection - The movement of a mass of hot material into a mass of cooler material is called convection. It differs from conduction in that a much larger number of hotter atoms are physically transported into a cooler region, thereby heating the cooler region much more quickly. The boiling of water is a very familiar example of convection, as well as the afternoon formation of a thunderhead in the Valley during the Summer.

Of the three kinds of energy transport, radiation is the least efficient, conduction is intermediate in efficiency, and convection is the most efficient. Convection usually occurs when a very hot object comes in contact with a cooler object, and the energy of the hot object cannot be removed by either radiation or conduction.

d. Thermal equilibrium: If the solar luminosity is very constant over a long period of time, it is not enough for the energy generation of the Sun to be relatively constant. The temperature of every part of the Sun must remain relatively constant.

Parts of the interior of the Sun transport energy by radiation, part transport energy by convection (convection near the surface of the Sun creates the granulation pattern in the photosphere).

4. Generation of Nuclear Energy

The gravitational collapse of a gas like the Sun produces thermal energy simply from the fact that the atoms of gas are squeezed more tightly together and, hence, will move faster (like the rapids on a river). Gravitational collapse can only keep the Sun at its present luminosity for 100,000,000 years.

Over the past fifty years, we have realized that the Sun gets its energy not from gravitational collapse, but from nuclear fusion.

a. Nuclear Fusion and the Strong Nuclear Force

Of the four fundamental forces in nature, we are familiar with two: gravity and electromagnetism. The other two forces operate only the scale of the size of the nucleus of an atom: The strong nuclear force holds the nucleus of protons and neutrons together, and must obviously be much stronger than electromagnetic force over the size of a nucleus. The weak nuclear force also is important for the structure of the nucleus.

It takes energy to bind two protons together with the strong nuclear force. This ’binding’ energy is greatest for atoms with a mass near that of the iron nucleus; it gets less with both lower mass atoms and higher mass atoms. Thus, if one combines two lower mass atoms together, such as four hydrogen atoms to form a helium atom, the lower energy of the combination will result in energy being released - nuclear fusion. Conversely, if one splits an atom much heavier than an iron atom into two smaller atoms, energy will also be released - nuclear fission.

The most abundant atom in a star is, of course, hydrogen. Nuclear fusion of hydrogen to helium is therefore the primary source of energy for a star. There are two ways this is done:

i) proton-proton (p-p) chain: Protons are directly combined in a series of steps to make helium nuclei.

ii) Carbon-Nitrogen-Oxygen (CNO) cycle: The carbon nucleus serves as a site on which four protons can combine to form a Helium nucleus.

The CNO cycle is more efficient at generating energy than the p-p chain, but it also requires a higher density and temperature than the p-p chain to work.

The action of nuclear fusion is often called nuclear ’burning’ in astronomy. All main sequence stars burn hydrogen to helium in their central cores. The main sequence is the longest period in the lifetime of any star because the hydrogen in the central region of a star is the largest supply of nuclear fuel available to a star.

5. The Russell-Vogt Theorem

From what has been discussed, the structure of a star should depend only on two things: The mass of the star (that is, how much gravity), and the chemical composition of the star (that is, how many particles are moving). Two stars of precisely the same mass and chemical composition will therefore have precisely the same evolution. This principle of stellar structure is called the Russell-Vogt Theorem, and assumes that the only factors in the evolution of a star are gravity, thermal energy and nuclear energy.

One consequence of the Russell-Vogt Theorem must be immediately obvious: In order to generate energy, the chemical composition of a star must change continuously: nuclear fusion makes fewer, heavier particles out of lighter particles. Thus, the internal structure of a star must be continually changing. For a main sequence star, these changes go very slowly, but are nonetheless there: The Sun is now about 50% brighter than when it first became a main sequence star 5 billion years ago. The Sun will become another 50% brighter over the next 5 billion years, until it exhausts its core of hydrogen.

In reality, the Russell-Vogt Theorem is not a iron-clad law: Stars rotate at different speeds, and a star that rotates fast will have a somewhat different structure than a star that rotates slowly, even though the two stars might have the same mass and chemical composition. Second, the magnetic field of a star contributes energy to support against gravity. Third, since a star like Sun is composed of 1056 or so particles, it is improbable that two stars will have precisely the same mass.

IX. THE FORMATION AND EVOLUTION OF STARS

A. Initial Collapse and Formation of a Star

The formation of a star typically begins when one of the dense, cold molecular clouds begins to collapse under its own gravitational force. This can happen either in a small, isolated molecular cloud, producing a single star like the Sun, or it can happen in a Giant Molecular Cloud. The gravitational collapse of a large cloud will cause it to fragment into smaller clouds, ranging in size from 0.1 times the mass of our Sun to over 100 times the Sun's mass.

As each of these 'fragments' begins to gravitationally collapse, its center becomes denses, hotter and more condensed, and the collapse proceeds more quickly. The collapse process for a star can be followed by placing its observed properties in the H-R diagram, if we can see the collapsing star. Since star formation takes place in very dusty regions, much of what we know about the very first stages of star formation comes from studies in the far-infrared and at radio wavelengths.

In general, all stars are first identified in the H-R diagram when they are very big and very cool (with temperatures of a very cool M star). Throughout the collapse process, the very center of a 'proto-star' is continually heating up. In these intial stages, proto-stars generate energy from gravitational collapse, and their outer envelopes are cool and relatively opaque. The combination means that proto-stars transport energy by convection (literally, a slow boiling of their interiors). Such a 'convective' star has a combination of luminosity, temperature and radius properties that must collapse in a certain way, in a certain region of the H-R diagram (known as the 'Hyashi track', after the astronomer who first defined it).

The actual evolution of a proto-star depends very directly on its mass. A relatively massive star (say, more than 10 times the mass of the Sun) will collapse relatively quickly (in 100,000 years or less). During its collapse, the radius of a high mass star will gradually shrink. The combination of rising temperatures and shrinking star will 'move' this 10 solar-mass star nearly horizontally across the H-R diagram with time, moving from a cool, M-type supergiant to a very hot, O or B type main sequence star.

The collapse process of a star with mass like the Sun is somewhat different: Lower mass stars do not become as hot in their centers as higher mass stars and, as a result, the proto-star must collapse to higher densities to eventually initiate nuclear burning. This has another consequence later in the life-history of this kind of star, but in this stage of evolution, it requires the star to become both smaller and not as hot as a higher mass star.

B. Main Sequence Evolution - General Concepts

The life history of a star depends princpally on the mass of the star when it is 'on' the Main Sequence (that is, when it is 'burning' hydrogen into helium in its core). The chemical composition of the star is of secondary importance for main sequence stars, but becomes more important in post-main sequence evolution.

A higher mass Main Sequence star:

```
is hotter than a lower mass Main Sequence star
is brighter      "             "
is larger        "             "
has a shorter lifetime than a lower mass Main Sequence star.
```

This last result (shorter lifetimes for more massive stars) is a consequence of the fact that higher mass stars can produce nuclear energy more efficiently than lower mass stars, combined with the fact that higher mass stars use their energy much more quickly.

When a proto-star begins to burn hydrogen into helium, it must change its internal equilibrium (since nuclear burning is a much more efficient energy source than gravitational contraction). This causes the outer surface of the proto-star to change slightly in temperature and size, resulting in a small 'movement' in the H-R diagram.

The definition of the Main Sequence is that it is formed by the observed properties of stars of different mass, when these stars are burning hydrogen to helium in their cores. Since this is the main 'fuel' of the star, the Main Sequence phase of stellar evolution is the longest period in the life of any star: up to 95% of the life of a star is spent on the Main Sequence. This, of course, is why the Main Sequence is such a striking feature in the H-R diagram: in any random sample of stars, most stars will be in the Main Sequence phase of their evolution.

Since the evolution of a star depends on its Main Sequence mass, it is convenient to talk about the evolution of two kinds of stars: a star with a Main Sequence mass near that of the Sun (generally, 2 solar masses or less); and a star with a Main Sequence mass much more than that of the Sun (generally 10 solar masses or more). Please keep in mind that the evolution of these two kinds of stars are being used to typify the the Main Sequence evolution of stars in general.

C. Evolution of a Solar Mass Star

Once a star with the mass of the Sun (in other words, a star of one solar mass) begins nuclear burning, there is enough hydrogen fuel to keep this nuclear burning going for about 10 billion years. During this time, the star continually evolves on a very slow time scale, gradually growing in size, becoming slightly cooler, and growing slowly in luminosity. Over the 5 billion years since our Sun was born, it has increased in luminosity by about 50%. The result of these slow processes is that a one solar mass star remains in about the same place in the H-R diagrm during it main sequence lifetime.

1. Exhaustion of Hydrogen burning in the core

Through nuclear burning, a solar mass star will enventually convert most of its hydrogen into helium in its center. When this core hydrogen fuel is exahausted, the core can no longer support the gravitational mass of the star, and the core begins to collapse again. For a solar mass star, this collapse occurs after about 10 billion years (the Sun still has about 5 billion years to go).

When the core of this Main Sequence star begins to collapse, several things happen at the same time:

the core becomes denser and, through energy gained from the collapse, the core becomes hotter

the outer envelope of the star responds to the core collapsing and getting hotter by expanding (Newton's third law) and, as a result, getting cooler.

the shell of hydrogen that surrounds this helium core is compressed by the gravitational mass of the star, and hydrogen burning now begins in a shell. This is now the main source of energy for the star.

2. The 'Journey up the Giant Branch' - Red Giant Stars

During this collapse process, the surface of a star (that is, what we see) becomes bigger and much cooler. The interior of the star becomes convective, and the star continues to grow much bigger and cooler. The increase in size of the star more than compensates for the star becoming cooler, so the star increases greatly in luminosity (by up to a factor of 1000). If we plot the changes of the star in the H-R diagrm, it would appear to 'move' from the Main Sequence towards the mid-upper right of the diagram, along a line similar to a 'Hyashi' track (but this time increasing in luminosity).

When the solar mass star reaches this stage, it is now a 'red giant', and the path it takes to reach its maximum luminosity is usually termed the 'giant branch' (for how it looks in the H-R diagram). Meanwhile, hydrogen continues to be burned into helium in a shell around the core, adding to a helium core that is still collapsing, searching for a source of energy to stop its collapse.

3. The Helium 'Flash'

As the collapse proceeds, the core continues to grow in mass and, being almost pure helium, gets hotter and denser. When the core becomes hot enough (100,000,000 oK), it can begin a nuclear burning process that combines three helium nuclei together to make on carbon nucleus. This burning process is usually called the 'triple-alpha' cycle (since the nucleus of a helium atom was originally named the 'alpha' particle in the early days of nuclear physics). However, the stars of the triple-alpha in the core of this kind of star is literally like an explosion! This explosion is due to the 'degenerate' nature of the core of this star at this stage; see Chap. 32.1 and 32.2 in the book.

This explosion in the core is termed the 'helium flash', and serves to counteract the core collapse and expand the core. After the helium flash the star will begin to burn helium in its core, and continue to burn hydrogen to helium in a shell around the core. In reaction to the helium flash and resulting core expansion, the star loses some mass (about 0.4 solar mass, via a very strong stellar wind), decreases somewhat in size, and becomes somewhat hotter. In the H-R diagram the star will appear to 'slide' down the giant branch towards lower luminosities (by about a factor of 10) and hotter temperatures, but still remain 'close' to the giant branch. (Stars with much lower abundances of elements heavier than helium [very old stars] become much hotter at this point in their evolution, but that will not be discussed here).

4. Post-Giant Branch Evolution

A star of one solar mass has enough helium fuel in its core to live as a bright red giant for about 100,000,000 years (or only 1% of its Main Sequence lifetime). After the helium core is mostly converted to carbon, the core of the star begins to collapse again. However, this time the core will not find another nuclear burning cycle to counteract the force of gravity, since the core cannot get hot enough, and the core continues to collapse, again becoming an electron degenerate gas. Shells of hydrogen burning into helium, and helium burning into carbon, surround the core and provide energy for the star. As the core collapses, the envelope of the star expands again and becomes cooler, 'moving' the star to even higher luminosities.

5. Planetary Nebulae

As the core is collapsing, the shell of helium burning is inherently unstable, producing a series of 'flashes' that gradually lift the outer envelope of the star away from the core. Over a period of a few million years the outer envelope of this star is expelled, leaving the very hot, small, collapsing core in the center of this ejected material. This very hot, small star ionizes the surrounding material, causing it to fluoresce. We observe these shells as 'planetary nebulae' (such as the Ring Nebula).

6. White Dwarf Star

The core of this star, by this stage, is so dense that the electrons in the core are packed extremely tightly together. In this condition, all of the electrons in the core 'act' as if they belonged to one 'big' atom. Certain quantum mechanical laws of physics then become important, including electron degneracy. Electron degeneracy simply means that electrons cannot be packed any closer together without enormous energy input. Such energy input does not come from the core collapse, so the core stops collapsing.

The result is a very dense, very small, very hot star that now begins to cool (since it is no longer generating energy by either nuclear burning or gravitational collapse). Such a star will be initially very hot, but it rapidly cools. Following the planetary nebula phase in the H-R diagram, the star must go from being very hot and luminous (about 1000 times the luminosity of its Main Sequence phase), to being hot and less luminous, to gradually cooling and becoming very faint. So a star moves from the upper left in the H-R diagrm, and descends diagonally downward towards luminosities 1/1000 to 1/10,000 that of its Main Sequence phase.

The end product of this evolution is a moderately hot star, gradually cooling that is very small - a 'white dwarf'. Only cores with of 1.4 solar mass and less can become white dwarfs - even electron degeneracy cannot counteract the force of gravity for a more massive core. Moreover, degenerate matter, like the electrons in a white dwarf, become smaller with increasing mass! This property is opposite that of common matter in stars. A white dwarf of one solar mass has about the size of the Earth.

7. Binary Stars, White Dwarfs and Novae

In a binary star, the two stars are almost never precisely the same mass. As a result, the more massive of the two stars must evolve off the Main Sequence first. If both stars are relatively low mass stars, there is a good probability that the more massive of the two stars will evolve into a white dwarf star while the less massive star is still on the Main Sequence. The changes that can occur in a binary star system under these circumstances are discussed in Chap. 32.3c and d in the book. Basically, when the less massive star becomes a Red Giant, its outer envelope expands, and part of it is gravitationally drawn to the white dwarf companion. This gas forms a ring of hot material around the white dwarf (termed an 'accretion ring'), and some of this gas can begin to actually accumulate on the surface of the white dwarf. If the density of this extra gas becomes high enough, nuclear burning of hydrogen to helium can occur on the surface of the white dwarfs. Such nuclear burning occurs explosively, and we see the result as a 'nova', a star that brightens by 10,000 - 50,000 times in luminosity over a few days time (see discussion below).

8. Rotation and Magnetic Fields

Much of our knowledge of the details of stellar evolution come from computer modelling of the basic physics of stars. Yet, such computer modelling is still very limited. For example, in the discussion of the evolution of a star like the Sun, we have ignored two properties that we know a star like the Sun possesses: rotation and a magnetic field. Both processes are potentially important in the evolution of a star: Rotation provides extra support against gravitational collapse, and magnetic fields provide an extra source of energy in the envelopes of stars. How important? We do not yet know, since the effects of rotation and magnetic fields are very difficult to model in our computers. A proper modelling of these effects will take a much larger computer than currently exists.

D. Evolution of a 10 Solar Mass Star

As stated before, the evolution of a star of 10 solar masses or greater is very different from that of lower mass stars. For example, we take the evolution of a star that has a Main Sequence mass of 10 solar masses.

1. Main Sequence Evolution

Very fast (by astronomical standards). A massive star will burn all of the hydrogen in its core very quickly. A star of 10 solar mass will take less than 10 million years. As the hydrogen is turned into helium, the core of the star is gradually becoming denser and hotter. The star's envelope reponds by increasing in size and becoming somewhat cooler - so the star becomes more luminous and cooler with age on the main sequence, like the solar mass star.

2. Post-Main Sequence Evolution

Unlike the solar mass star, the core of the 10 solar mass star is not dense enough to become degenerate after it has exhausted its hydrogen - the core was hot enough to burn hydrogen without getting too dense. Instead, when the hydrogen is completely used up, the core of the star contracts further, the envelope of the star expands quickly, and the star is observed to become much cooler.

As the triple-alpha cycle ignites in the core, the core expands gradually, and the envelope of the star gradually contracts and becomes hotter. Around the core, hydrogen is being burned into helium, continually adding to the core. The star is still much redder than when it was on the Main Sequence, but now its even brighter because of the expansion of the envelope before helium burning began.

The 10 solar mass star is now a red supergiant. When the helium in the core is burned to carbon, the core again collapses, the envelope expands and cools, until the core can gain enough temperature and density to 'burn' carbon in nuclear reactions. The shell of helium burning around the core is somewhat unstable, causing the envelope of the star to lose mass. Thus, this red supergiant star is continually ejecting matter from its surface after the core has become carbon.

After carbon is 'burned', successive elements are used in nuclear burning cycles, each producing a nucleus that is heavier than the last: carbon to neon, neon to oxygen, silicon to iron. Each new burning process is even more 'touchy' than the last, causing the interior to 'burn' rather erratically. Eventually the interior of the star will by 'layered' according to the elements that have been burned: a hydrogen shell eneveloping a helium shell enveloping a carbon shell enveloping a neon shell, ...

3. Towards the end results ...

The end product of all this nuclear burning is the creation of the element iron. At the instant the core becomes iron, a remarkable change occurs: Iron will not produce energy when fused together; instead, it takes energy to fuse iron nuclei together. As a result the iron core shrinks almost instantaneously (a core of 3-4 solar masses, with a size of the Sun), absorbing energy from the collapse, and causing the collapse to proceed even faster. The mass of the core is too great for electron degeneracy to overcome the force of gravity, so that the electrons and protons are forced together to form neutrons. In this process, the energy that went into forming the neutrons is released by neutrinos, which simply escape from the star. This cools the core even more rapidly, and the very center of the core collapses either into a neutron-degenerate center about 20 miles in diameter, or if the core is too massive, into a black hole (more about which will be discussed later). All of this takes place literally in the wink of an eye - about one second of time.

The formation of the neutron star sends a tremendous shock wave back through the star, off of which the still-collapsing outer core bounces. The result is that 3 solar masses of material is sent outward from the center at speeds approaching the speed of light. The result of this is that the rest of the stars literally tears itself apart in the space of less than a day, spewing material out at velocities of 20,000 km/s or more. The light produced by this outpouring material is bright enough to outshine 10,000,000,000 stars - this is what we call a supernova.

Until last year, all of the supernovae that had been seen by man and recorded in history occured before the invention of the telescope. This changed in February, 1987, when a supernova exploded in another galaxy that is very close to our own - the Large Magellanic Cloud. The LMC supernova provided the first real test of this model of supernova explosion: Neutrino detectors here on Earth detected neutrinos coming from that supernova (a total of 12 neutrinos were actually detected, out of the 1050 neutrinos produced by the supernova), confirming the model that a neutron star is formed in the core of a supernova.

D. The Phenomena Associated with Supernovae

1. Supernova Remnants

Supernovae are among the most energetic events in our galaxy. Through their ejected material they 'seed' the rest of the gas with elements that were created in the first few hours of the explosion: All of the elements much heavier than iron and cobalt were created in this kind of supernova explosion - that includes gold, silver, platinum and uranium. We are literally made of star-stuff. Not only does the supernova provide enough energy to make these elements, it also provides a near-perfect delivery system - a straight injection of material far into the interstellar medium. In the first several thousand years after a supernova explosion, this ejected material will expand from the now-exposed neutron star.

In known supernova remanants, such as the Crab nebula, we see this ejected material still moving at 2,000 km/s nearly 1,000 years after the exposion. The spectra of the Crab nebula shows emission lines due to heavy elements, such as silicon and sulfur. The energy of the supernova also goes into making X-rays, ultraviolet rays and radio waves. The most energetic particles from the supernova move at nearly the speed of light, escape the supernova completely, and eventually become cosmic rays. Just by these interactions alone, supernovae play a very important role in the evolution of other stars in our galaxy.

2. Neutron Stars - Pulsars

If a neutron star is left as the result of a supernova explosion, it will be rotating very fast. The very rapid contraction of the core must conserve angular momentum, with the result that a neutron star of 20 miles diameter will, when initally formed, rotates as fast as 100 times a second! In addition, the rapid contraction of the core traps whatever magnetic field was in the core of the parent star, also condensing it and making it tremendously strong (over a billion times stronger than that on the surface of the Sun).

This combination of rapid rotation and very strong magnetic field causes the gas around the neutron stars to be wrapped into a rapidly rotating thick ring, or torus, of material around the equator of the neutron star. This ring will have holes at the magnetic poles of the neutron star. The gas torus will be very hot, and this hot gas will radiate primarily in the radio region, along the axis of the magnetic poles. However, in general the axis of the magnetic poles will not be aligned with the rotation axis of the neutron star itself, so this radio beam will appear to 'circle' around the rotation pole. To an observer far away from the neutron star, the radio beam will appear as a sharp pulse with a very fast, regular period - a 'pulsar'.

The existence of a neutron star was predicted in the 1930's, but the first neutron star was discovered as a pulsar by Jocelyn Bell Burnell in 1967. The interaction of the magnetic field of the neutron star with its surrounding gas means that the star gradually loses energy, and must start slowing down its rotation. It is believe that, over 10's of thousands of years, a pulsar will essentially stop spinning and, as result, no longer pulse.

3. Black Holes

If the central part of the core of a supernova has a mass greater than three solar masses, even neutron degeneracy will not prevent the force of gravity from completely collapsing the core. The result of this complete collapse is an object with a mass of more than three solar masses, but a radius of less than 10 km. A mass so condensed has a gravitational field so strong that the escape velocity from its surface is greater than the speed of light. Thus, no light can be emitted from its surface, and no matter or light, once it comes very near its surface, can escape. This is a 'black hole.'

4. Cosmic Rays

Cosmic rays are nuclei of atoms such as helium and carbon that are moving at velocities very close to the speed of light. We think that most of the lower energy cosmic rays are produced in supernova explosions. Moving at nearly the speed of light, these particles can travel over very large distances, so what we observe as cosmic rays are particles that have been emitted by supernovae in our galaxy over the past few million years - about 100,000 supernovae.

E. The Phenomena Associated with Condensed Objects

Geometrical considerations show that the graviational field of a spherical mass attracts other masses as if all of its mass were located at its center. As a result, the Sun, a one solar mass white dwarf, a one solar mass neutron star, and a one solar mass black hole, all have the same gravitational force far away from the object. However, the diameter of the Sun is 800,000 miles, the diameter of the one solar mass white dwarf is 8,000 miles, that of the one solar mass neutron star 30 miles, and that of the one solar mass black hole, 4 miles. The acceleration produced by gravitational attraction is cumulative, and increases rapidly with decreasing distance from the center of the attracting mass (try dropping an egg into a deep well, compared to dropping into a shallow bowl).

Hence, matter falling onto a one solar mass white dwarf will gain much more energy than falling onto the Sun - about 10,000 times more energy. Matter falling onto a one solar mass neutron star will gain 50,000 times more energy than falling onto a one solar mass white dwarf, and 50,000,000 times more energy than falling onto the Sun. Matter falling onto a one solar mass black hole will gain 50 times the energy of matter falling onto a one solar mass neutron star, and 2,500,000,000 times more energy than falling onto the Sun. All of this, simply because matter can fall further into the deep gravitational 'well' created by gravity.

Of course, matter does not just simply 'fall' onto a neutron star or a black hole. Matter, typically in the form of gas, will collide with itself as it falls, and convert this increase in gravitational energy into thermal energy. If the falling gas has some angular momentum around the 'condensed' mass, as is most likely the case, the gas will form a ring of material around the condensed object.

1. Condensed Objects in Binaries

The evolution of a massive binary star provides a ready source of fuel for white dwarfs, neutron stars and black holes. We have already discussed how, when a white dwarf is formed in a binary star system, an 'accreation' disk of material can build up around the white dwarf when the other star beomces a red giant. Such a situation can lead to a nova.

The same situation arises when two massive stars are in a binary star system. The more massive of the two stars will evolve first, become a supernova, and leave a rapidly spinning neutron star. When the initially less massive star becomes a red supergiant, it will feed material towards the neutron star, and this material will also become an 'accretion' disk.

The size, density and temperature of an accretion disk depends on how condensed the central object is (that is, white dwarf, neutron star or black hole), how much material is being released by the companion star (a red supergiant of 10 solar masses releases more mass than a red giant star of one solar mass), and how fast this matter is being released.

If a lot of matter is released very fast onto a very condensed object, the result can be an accretion disk that is very hot and very dense. Such a disk will emit most of its energy in the X-ray region of the spectrum. If somewhat less matter is released at a somewhat slower pace, the accretion disk will be less hot and less dense, and might release most of its energy in the optical part of the spectrum. All factors combine in such a way to produce accretion disks with a wide range in density and temperature. In most cases, however, the temperatures and densities in the accretion disks around condensed objects can become high enough to initiate nuclear burning.

2. Novae

While supernovae are the most spectacular explosions that occur in the galaxy, novae (the plural of nova) are the most common explosion. We believe that all novae come from the accumulation of hydrogen and helium gas on the surface of a white dwarf star. This most likely happens in binary stars, and almost all of the known novae are in binaries. But it could also happen to white dwarf stars that are in dense interstellar clouds. When the hydrogen on the surface of a white dwarf ignites, it does so in a degenerate gas, resulting in an explosion. This explosion ejects the accumulated envelope, and a great deal of energy along with it. The result is that the white dwarf brightens by a factor of up to 100,000. If the white dwarf is in a binary, the accumulation of mass on the white dwarf will continue after the nova outburst, eventually resulting in another nova outburst! Indeed, many novae are known to be recurrent, recurring every 50-100 years!

3. The Special Case of SS 433

The phenomena associated with condensed objects can lead to some objects that appear to be very bizarre by normal standards. The variable star SS 433 is one such case. It is an object that is literally both 'coming and going' at the same time. The current best model for SS 433 is that of a binary system containing a massive neutron star with a 20-30 solar mass companion in its post-main sequence phase of evolution. The result is a tremendous influx of matter near the neutron star, which forms a very thick accretion disk of material around the neutron star. Some of the matter flowing towards the neutron star is channeled by the very strong magnetic field onto the pole of the neutron star, from which nuclear burning ejects matter that is channeled outward along a very narrow beam - a jet of material moving a velocities close to that of the speed of light.

The result is that we see jets of material that come from both poles of the neutron star, jets going in opposite directions at about 0.3 times the speed of light. Besides being of intrinsic interest itself, SS 433 has been an object of intense study every since it was discovered for a more basic reason: Many astronomers feel that that physical mechanism that produces the jets in SS 433 is the same as that which produces the million times more powerful jets in the centers of galaxies and quasars. If so, then SS 433 is the only such 'engine' that we can currently study 'up close.'

X. THE OBSERVATIONAL EVIDENCE FOR STELLAR EVOLUTION: STELLAR POPULATIONS

Historically, the key to deciphering stellar evolution was the fact that many stars are found in star 'clusters', which are regions of space in which a large number of stars are formed out of a single, large, gravitationally-bound cloud of gas (sometimes with 1,000,000 solar masses of gas). These clusters of stars contain stars of different masses that were born in close proximity to each other at the same time.

1. Star Clusters

Our galaxy has apparently made clusters of stars since it was formed, and stars clusters are still being formed today. Such star clusters permit us to study a wide variety of phenomena associated with star formation: i) How stars of the same age, but different mass, evolve with time. ii) Once knowing that, calculating how old star clusters are. iii) Knowing (i) and (ii), being able to determine how the star formation rate, and how the abundance of chemical elements in the galaxy, have changed with time (the latter due to the steady enrichment of heavy elements formed in novae and supernovae, and ejected into the interstellar medium).

The types of star clusters in our Galaxy are classified as one of three types:

Open Cluster (also called galactic cluster) - a gravitationally-bound cluster of stars found in the disk of our galaxy. They usually contain 100-1000 stars, and are irregular in shape.

Globular Cluster - a gravitationally-bound cluster of stars usually found outside the disk of our galaxy. They can contain 1,000 - 1,000,000 stars, and are spheroidally-shaped (hence the name 'globular').

Associations - loose groupings of stars that may or may not be gravitationally-bound. They usually contain less than 200 star, and are always found in the disk of the Galaxy.

2. The H-R Diagrams of Clusters

The stars in a cluster differ principally only in the mass that they had when they were formed (their 'initial' mass). If we measure the parameters necesary to construct the H-R diagram for stars in a typical open cluster, such as the Pleiades or the Hyades, we would see the the Main Sequence would contain stars with a relatively wide range in temperature, ranging from warm stars near A0 to cool stars near M. We would not find Main Sequence stars hotter than A0. In addition, the open clusters would have a few bright, cool giants stars near M spectral type, but few, if any, warm giants from F-K.

If we determine the H-R diagrm for stars in a typical globular cluster, we would find no warm stars on the Main Sequence. The stars with the highest temperature that we find on the Main Sequence will have G0-G3 spectral types. This is because this kind of cluster is old enought that all stars with masses higher than the G0 stars have already evolved into neutron stars and white dwarfs. In addition, we would also see stars that have luminosities and temperatures that form a 'branch' from the warmest remaining part of the Main Sequence - the 'giant branch.' The giant branch exists because the post-main sequence evolution times of these low mass stars are about 10% as long as their Main Sequence lifetimes.

The H-R diagrams of open clusters are obviously different from those of globular clusters. This fact was known for about 20 years before it was understood why these H-R diagrams were different (in the 1950's). The stars in open clusters differ from the stars in globular clusters in two fundamental properties:

a) Open clusters are relatively young, typically being the age of the Sun or younger. Globular clusters are old, typically 3 times the age of the Sun.

b) The abundances of elements heavier that hydrogen and helium are much higher in the stars in open clusters than in stars in globular clusters (by factors that range from 10-100). The abundances of these 'heavy' elements in open cluster stars are similar to that measured for the Sun.

3. Stellar Populations

The concept that large groups of stars in our Galaxy can have different ages and chemical composition is usually termed 'stellar populations.' The concept of stellar populations was begun in 144 by Walter Baade, and over the past 45 years has increasingly grown in complexity. Originally, stellar populations were broken down into two categories:

Population I were the young, heavy-element rich stars in the disk of our Galaxy

Population II were the old, heavy-element poor stars in the galactic halo

This terminology is still convenient today, by only in a very general sense. The true stellar populations in our galaxy require a classification scheme that is much more sophisticated; a requirement that has become gradually and steadily more complex as we learn more about the stars in our Galaxy.

It is important to understand, however, that the differences in the H-R diagrams of open clusters and globular clusters were the driving force behind the development of modern concepts of stellar evolution.

4. Pulsating Variable Stars (revisited)

The classes of variable stars that vary in luminosity in a regular manner - RR Lyrae stars and Cepheid variables - do so by pulsating. That is why their period of pulsation is related to their size and luminosity - a smaller star can pulsate faster, a larger star must pulsate more slowly. This kind of pulsation will happen to any star that has the right combination of temperature and luminosity. Most post-main sequence massive stars pass through the Cepheid pulsation phase; only the more heavy-element poor, low mass giant stars pass through the RR Lyrae phase.

XI. GALACTIC STRUCTURE

A. Getting a sense of scale

Astronomers, just like everyone else, do not like to deal in saying or writing down very long numbers, such as 1,325,256,790 inches (feet, miles, etc.). Instead, units of distance (and time and mass) are used so that quantities in these units are kept at a comfortable size. For example, one would not measure a room in km, or the distance from ASU to London in inches.

In astronomy, we use the A.U. to measure distances in the solar system, and the parsec to measure distances to the nearest stars. The size of our galaxy, however, is measured in thousands of parsecs, or kiloparsecs (kpc). The distances to the nearest galaxies are measured in millions of parsecs, or megaparsecs (Mpc); and the size of the universe is measured in terms of billions of parsecs, or gigaparsecs (Gpc).

B. The Different Forms of Matter in Our Galaxy

In general, matter in our Galaxy (our home Galaxy is always capitalized) comes in one of four forms:

1. Stars: Individual stars; binary stars; mulitple stars; clusters of stars - open, globular, association.

2. Gas: The hydrogen atom - neutral and ionized; hydrogen molecules, atoms of helium, oxygen, carbon and all of the other elements, in roughly the ratio that we find in the Sun; molecules made from these elements, including organic compounds. This gas is produced partly by the deaths of stars, and is often ionized by the radiation of starlight, resulting in the fluoresence of the gas (for example, HII regions). (discussed previously).

3. Dust: Agglomerations of atoms and molecules forming large grains

(disussed previously).

4. Dark Matter: There exists matter in our Galaxy whose gravitational force we can measure, by which apparently does not emit any electromagnetic radiation that we can detect. That some matter in our galaxy would be 'dark' to us is not surprising: if the Galaxy were filled with comets, we could not detect them. What is surprising, however, is that as much as 95% of all matter in the universe could be dark! What this dark matter could be will be discussed in later lectures.

C. The Physical Components of our Galaxy

The basic shape of our Galaxy is that of a very flattened disk, coexistent with a spheroidal distribution of stars. The stars in the disk emit about three times as much total light as the spheroid. The structure of our Galaxy is commonly described in four parts:

The nucleus: The very center of the galaxy, the densest region of stars and gas in the galaxy. The nucleus has a size of about 3 parsecs. We can only observe the nucleus in very long, or very short wavelengths, such as infrared, radio and X-ray. As best as we can determine, the nucleus is comprised of stars, gas and dust, with active star formation, and the possibility of containing a central black hole with a mass of 1,000,000 solar masses.

The Bulge: The inner, most visible parts of the spheroid are generally termed the 'Bulge', since they define a rather bulge-like feature relative to the very flat disk. The center of the bulge is coincident with the center of the nucleus. The bulge is comprised of primarily old stars.

The Disk: The disk is the main visible component of our Galaxy. It is a very flattened object - the thickness of the disk near the Sun is about 1 kpc, but the disk is over 30 kpc in diameter, or a ratio of 30 to 1 in thickness. For comparison, a table of thickness one inch would have to be 30 inches long to have a ratio of 30 to 1. The disk contains both old and new stars, gas dust and active star formation. The center of the disk is coincident with the centers of the bulge and of the nucleus.

The Halo: The outer parts of the bulge, that extend to distances well beyond the disk, are commonly termed the 'halo'. We trace the halo by the globular clusters and stars that we can see, but we also strongly suspect that dark matter contributes much of the mass of the halo. As such, I have found it convenient to adopt a terminology introduced by Gerry Gilmore, of England's Cambridge University:

Halo of the First Kind: The stellar halo, comprised mostly of old stars and globular clusters.

Halo of the Second Kind: The dark matter component, whose composition and distribution we are still trying to figure out.

The disk, halo, bulge and nucleus are all have common center. The density of matter in each component is greatly concentrated towards the center: for example, a cubic parsec of the material that lies in the disk in the center of the disk is 10,000 times the density of a cubic parsec of disk material at the edge of the disk.

Much of the following discussion presents a supplemental view of our Galaxy, to complement the view given in the book:

D. Some Helpful (?) Analogies

The objects that belong to the disk (that is, stars, gas and dust) and halo stars can all be found anywhere in the disk in our galaxy. The reverse statement is not true, however - stars in the disk are only found in the disk, not in the halo. It is the density of such objects which determines how we see that component from a distance, and it is the motions of the objects which determine which component they belong to.

Consider the following analogy:

Suppose I construct a disk of sand particles, 30 feet across and one foot thick. Like the disk of our Galaxy, let me make the density of sand particles very high in the center, decreasing outwards: If in the center I have 10,000,000 sand particles per cubic centimeter, at the edge I would have only 1,000 sand particles per cubic centimeter. Among this disk of sand particles, let me make small clouds of smoke, along with some clouds of household dust. Of course, the air molecules are present all the time. Then let me form some more sand particles into a spheroidal shape (that is, like a squashed basketball), with its center coincident with the center of the disk. Finally, let me put a very dense, very tiny steel ball right at the center of the disk.

The analogy is this: The sand particles are stars; smoke is the hydrogen gas in its various forms; dust is dust; air molecules are very hot gas; and the steel ball is a 1,000,000 solar mass black hole. The sand particles in the disk are analagous to the stars in the disk of our Galaxy; the sand particles in the 'squashed basketball' are analagous to the stars in the Galactic halo.

Note that in our model, the sand, dust, smoke and air can coexist in the same volume of disk space, as well as some sand from the halo. In the same way, stars of the halo can coexist in the same volume as the stars of the disk - it is just that there are far fewer spheroid stars where there are disk stars, so they are much harder to pick out.

A Scaled-Down Disk

For the purposes of describing the properties of the Galactic disk in a manner with which most of you are familiar, let me make the following 'scale model' of the Galactic disk:

1 mile = 5 parsecs; then the diameter of the Earth = 40,000 pc, or about the size of our Galaxy.

It takes the Sun 250,000,000 years to rotate once about the center of our Galaxy, so let 1 Galactic year = 250,000,000 Earth years.

With this scaling, the nearest star would be about 0.2 mile away; the Sun would move around the center of the Earth at a velocity of 1.4 miles/ Galactic hour. A supernova will eject matter at velocities of 60-150 miles/hour into a gaseous medium that itself is moving at a mere 3 feet/hour! In other words, the supernova explosion scales down to a hurricane force wind! Novae eject material at velocities of 30-60 miles/hour. The Sun would last 40 years, but a 20 solar mass star would last only 7 days before it became a supernova. All around you, supernova would be going off a a frequency of about 1 per Galactic day, and novae at about 20 per day. A pretty active place!

I hope this gives you a feel of what the environment of our disk is like on Galactic scales: a dynamic place, with gas moving at both large and small velocities, and much turbulent motion.

E. Brief Supplemental Description of Individual Components of our Galaxy

1. The Nucleus

Our current picture of the nucleus is one of a dynamic, very dense, very turbulent inner 3 parsces of our Galaxy. Energy is apparently being supplied both by supernovae occurring on a regular basis, due to large regions of star formation, and there is mounting evidence that a black hole with a mass of 1,000,000 solar masses sits at the very center. The actual measurement is this: From measuring the rotation of gas very near the center, and using Kepler's third law, we find that there is 2,000,000 solar masses of matter within the central 2 parsces of our Galaxy. For comparison, there are only about 3 solar masses of matter in a sphere of 2 pc centered on the Sun.

2. Bulge

The stars that comprise the central bulge are older than most of the stars in the disk. They also have a high chemical abundance of heavy elements, indicating that they we formed from gas that had been through many generations of supernovae, in contrast to the heavy-element poor stars in the halo. How old are the bulge stars? We know that they are older than 7 billion years, but we do not yet know if they are as old as globular clusters (15-18 billion years; see below).

Note that bulge stars do not fit the general desciption of Population II stars - they are old stars, but they are not heavy element-poor. It was this knowledge, gained over the past 15 years, that has partly lead to the more complicated description of stellar populations used today (but which will not be used in this class).

3. Halo of the First Kind

The stellar halo is, naturally enough, defined by stars:

Globular Clusters: We can observe about 150 globular clusters in our Galaxy, and we can estimate that our Galaxy has a total of perhaps 300 clusters. Measurements of the H-R diagrams of globular clusters shows that they are very old - about 15-18 billion years old. The globular clusters in our Galaxy are not the oldest known objects in our Galaxy, they are also the oldest known objects in the known universe.

Halo Stars: RR Lyraes are relatively easy to spot in the halo; recently studies have been made of fainter, harder-to-study stars in the halo. Our view of halo stars has been changing over the past five years. We are now finding important differences in the chemical abundances of halo stars relative to globular cluster stars. These differences have made many astronomers re-think the relationship between the globular clusters and the halo that we see. This re-thinking process is going on now - but no clear cut answer yet.

4. Halo of the Second Kind

The term 'dark matter' is a bit misleading, since this matter may just be enormously 'under-luminous', such as comets, bricks, bats, bottles, or just about any object smaller than Jupiter. Dark matter could also be black holes or very low mass, 'black' dwarf stars.

Not many contraints on the form of dark matter come from studies of our own Galaxy. Through studies of other galaxies, and clusters of galaxies, better constraints lead to more plausible sources of dark matter, and these will be covered in the lectures concerning the large-scale structure of the universe.

5. The Disk

The disk of our Galaxy contains the majority of all stellar mass in the Galaxy. The disk is also the site of almost all of the star formation that takes place in the Galaxy (the nucleus being the other site).

If we could view our Galaxy from a position perpendicular to the plane of the disk, the most noticeable thing we would see in the disk are the characteristic spiral arms. Why disks should contain spiral arms will be discussed later. Here we note that the spiral arms contain hot, blue (that is, young) stars, concentrations of clouds of Hi and H2 and dust. The rest of the disk is not empty, however. Sprial arms are only a 10-20% enhancement over the rest of the disk, and the disk itself is composed of young and old stars.

F. The Sun's Place in the Galaxy

The Sun is situated about halfway out in the disk of our Galaxy, using the most recent estimates of the Sun's distance to the center of the Galaxy - 7 kpc, and the radius of the Galaxy - 15 kpc. The Sun, like the rest of the disk, rotates about the center of the Galaxy (see below), completing one revolution in about 220,000,000 years. At the present time, the Sun is sitting in a relatively quiescent part of the Galaxy; no active star formation is taking place within 50 parsecs of the Sun, and there is little dust within 100 parsecs of the Sun. However, if one were to travel just 150-200 parsces from the Sun (or about 40 miles in our 'model' galaxy), one would encounter massive areas of dust, HI, HII, H2 and, of course, star formation. We see one of these active areas as the stars forming the Orion constellation.

In fact, if the Sun were located just 200 parsecs closer to the center of the Galaxy, or 200 parsecs further away), the dust around us would be so thick that it would be like living in a deep fog, and we would visually see little of the rest of the Galaxy, not to mention very little of the universe!

By our best estimates, the current position of the Sun relative to these dust clouds is by chance. I estimate that we had a 15% chance to be in our current 'clean' area, and a 15% chance to be in one of the 'dusty' areas. Hence, the fact that, at this time and place in mankind's histroy our Solar System exists in a relatively transparent part of the Galaxy is, to my mind, as spectacularly coincindental as the angular sizes of our Sun and Moon being of similar sizes to produce total and annaular solar eclipses!

G. The Rotation of Our Galaxy

The stars, gas and dust that make up the disk of our Galaxy must rotate around the center of the Galaxy, much the same way that the planets must either revolve around the Sun, or they would fall into the Sun. If we measure the rotation velocity as a function of distance from the center of the Galaxy, we can measure the amount of mass interior to each point using Kepler's third law:

(m1 + m2)P2 = a3, with mass in solar masses, P in years, and a in A.U.

Use of this equation is given in the book on pg. 606.

As we move from the center of our Galaxy outwards, we encounter more total mass, and our rotation velocity will increase. Eventually, however, we begin to encounter less and less total visible mass (the disk is getting much less dense with distance from the center), so that the rotation velocity of the Galaxy should eventually stop increasing with distance from the center, and begin to decrease with distance from the center.

For 50 years this is what astronomers believed should happen. However, during the 1970's more sensitive measurements of rotation velocities in the outer parts of spiral galaxies, including our own, indicated that the rotation velocity did not fall. Instead, the rotation velocity remained at a rather constant, or even slightly increasing, level. This lack of 'falling rotation curves' directly implies the presence of mass that was not being seen - dark matter.

H. Differential Rotation

From a distance of abour 4 kpc from the center to a distance of 15 kpc from the center, the disk of our galaxy rotates with about the same rotation speed. This constancy of rotation velocity influences how gaseous material is distributed in the disk.

Consider the following example: Suppose one arranged HI and molecular gas clouds to lie in a straight line from the center of the Galaxy out to the distance of the Sun. After a few million years the clouds near the Sun will have moved only a small fraction of their orbit, while the clouds closer to the center of the Galaxy will have moved a much larger fraction of their orbit, simply because the circumference of their orbit is smaller. This is true, even when the rotation velocity is constant with distance from the center. In fact, a straight line will always turn into a spiral line in a galaxy that has differential rotation. Differential rotation just means that a galaxy does not rotate like a record (that is because a record rotates so that all of its parts are moving at the same angular speed, not speed in cm/sec.)

I. Why Only Two Spiral Arms?

Many other spiral galaxies that we can see have two main, dominant spiral arms in a rather symmetric pattern. We suspect that our own Galaxy has such a two-armed pattern as well. Although differential rotation explains why the shape of arms should be spiral-like, it does not explain why many spiral galaxies just have two arms. There must be some physical effect that causes spiral galaxies to have two main arms, and to sustain these arms over a long period of time (otherwise, the arms would 'wind-up' in one or two rotation periods, typically 100-300 million years).

It is now geneally thought that the two main spiral arms are caused by the presence of a spiral density wave pattern in the disk of the galaxy. A density wave is a change in the local gravitational force that is caused by material being compressed together (the slow-down in traffic along a highway caused by an accident is a 'density wave' in the traffic flow).

The two-armed density wave pattern can be produced by gravitational interactions between neighboring galaxies, or by large irregularities in the gravitational field, such as due to a bar-shaped distribution of stars in the disk (see the discussion of kinds of galaxies).

J. The Galactic Magnetic Field

The Galaxy possesses a magnetic field, just like the Sun. Although the magnetic field of the Galaxy is about 1 millionth that of the average magnetic field of the Sun, it covers a volume 1029 times that covered by the Sun and, as such, has 1017 times the total energy of the Sun’s magnetic field.

How does the magnetic field of the Galaxy originate - we do not know. We do know that it influences how charged particles travel within the galaxy (such as cosmic rays), and that it influences how gas clouds collapse to form stars (by restricting the manner in which they collapse). The presence of a magnetic field also produces a different kind of electromagnetic radiation from charged particles - synchrotron radiation, as discussed in Chap. 34.4a.

K. The Mass of Our Galaxy

If we apply Kepler’s third law, we find that our galaxy has between 1-2 x 1011 solar masses of material interior to the Sun. However, when we try to sum all the mass in stars, gas and dust, we only come up with about 1/2 of this mass - another indication that much of the matter in our Galaxy is dark. In our galaxy, as in most other galaxies like our own, about half of the mass is of matter that we see, and about half is dark matter, within the parts of the galaxy that we see. What exists beyond the parts that we see? We do not yet know, but there could very well be much more dark matter outside the visible matter.

L. How Did Our Galaxy Form?

In trying to understand how our Galaxy, and the billions of other galaxies like it formed, we have to consider several basic points: (1) Originally our galaxy, as well as the rest of the universe, was gaseous. (2) Our galaxy has a disk that dominates the visible matter, so that most of the gas out of which the galaxy formed had to be rotating when it collapsed to form the galaxy. (3) Our galaxy has a spheroid made of old stars, so that some of the visible matter collapsed into stars early-on, and did not make a disk. (4) Present evidence suggests that the oldest disk stars are about 6 billion years younger than the oldest halo and bulge stars, implying that the halo was formed first, the disk later.

These are the basics, but going beyond them requires a more detailed understanding of the one part of our Galaxy that we cannot easily study - the dark matter. Since at least half of the mass of our Galaxy is dark matter, this omission is obviously important. Theorists are just now coming to grips with this problem.

XII. EXTERNAL GALAXIES

A. Historical Perspective

We are most familiar with the structure of galaxies as seen in visible light (for example, either with the eye or with photographs). In visible light, the light from galaxies comes primarily from stars and flourescence of HII regions and planetary nebulae, as modified by the extinction and reddening of dust (if dust is present; not all galaxies have either dust or a disk).

Other parts of the electromagnetic spectrum emphasize other components of a galaxy: for example, the X-ray emission from a galaxy comes from very hot (greater than 1,000,000 oK) gas; the gravitational mass comes from measuring the motions of stars and gas in a galaxy; HI is measured with radio telescopes (via the 21 cm emission line).

However, the measurement of all of these other components in galaxies requires methods and kinds of telescopes that have only been used in the past 10-30 years. In contrast, one could look at galaxies through a telescope since the invention of telescopes (1610), and one could take photographs of galaxies since 1850. Thus, our knowledge of the structure of galaxies is greatest at visible wavelengths, although, as we have seen and shall see, the structure of galaxies at other wavelengths could be different.

Until the invention of photography, galaxies and HII regions in our own Galaxy both appeared as 'fuzzy' or 'nebulous' images in a telescope, with most galaxies having a spiral shape, and most HII regions having an irregular shape. Astronomers of that era called the spiral images 'spiral nebulae' (nebula is the singular), and the irregular images 'diffuse nebulae.'

With the invention of the spectrograph in the mid-1800's, it was soon realized that the spiral nebulae were made mostly of stars, while the diffuse nebulae were mad of gas. The issue then became, were the spiral nebulae contained within our own Galaxy, or did they constitute separate 'island universes' (in the words of Immaneul Kant, who so hyphothesized in 1755)?

This debate raged off and on for nearly 50 years, confused by the changing interpretation of the size and structure of our own Galaxy (due to the obscuring problems due to dust). The controversy over this issue culminated in a 'debate' in 1920 between two of the principal researchers in this field: Harlow Shapley for spirals being local and in our Galaxy; H.D. Curtis for spirals being separate galaxies. The results of such a 'debate' are themselves inconclusive, but serve to highlight the points of difference on the issue. In any event, the issue was being settled during this time by Edwin Hubble, who, using the new 100" telescope at Mt. Wilson, had measured the brightnesses of Cepheid stars in nearby galaxies. Hubble's measurements showed that these 'spiral nebulae' were really very distant compared to the globular clusters, thus ending the debate.

B. Types of Galaxies

As photographs of galaxies were accumulated in the early 1900's, it was easy to notice that the apparent 'forms' of the galaxies could be grouped into a very few classes, much like the spectra of stars could be classified into very few classes. A number of different classification schemes for galaxies were invented in the period 1915-1930, and different classification schemes for galaxies are still being invented to this day. However, the classification scheme most widely used is the one partially due to Hubble, and is given his name.

The Hubble classification of a galaxy is also called the 'Hubble Type' of a galaxy, and the sequence of galaxies so defined is called the 'Hubble Sequence':

Elliptical galaxies (E): Galaxies that are elliptical in shape, ranging from nearly circular to being somewhat elongated. The classification of an elliptical galaxy is given by its apparent flatness, with an E0 being a circular galaxy and an E6 being an elongated galaxy. Ellipticals are mainly stellar in appearance, with few, if any, signs of active, on-going star formation, and no visible disk.

S-zero (S0) galaxies: Galaxies that have both an elliptical-like component (now given the genreal name of ’spheroid’) and a disk component. S0’s share the common property with ellipticals that they have few signs of active star formation, and S0’s share the common property with spiral galaxies that they have a visible disk.

Spiral galaxies (S): Galaxies, such as our own, with a spheroid and a disk, with active, on-going star formation, some of which occurs in the shape of spiral arms.

Spiral galaxies are subdivided according to three criteria based on the visual appearance of the galaxies on photographic plates: a) How much do the spiral arms appear to wind-up; b) how knotty in appearance are the arms; c) how big does the spheroid appear relative to the disk.

The subdivisions of spirals are as follows:

Sa galaxies: arms that are tightly wound around the center of the galaxy; arms with little resolution into knots; a relatively large spheroid relative to the disk.

Sb galaxies: arms less tightly wound; greater resolution of arms into knots; smaller spheroids relative to disk, but spheroid still noticeable.

Sc galaxies: arms loosely wound, high degree of knottines of arms, spheroid small.

Irregular Galaxies (I): Galaxies that have no definite form. Irregular galaxies have no evidence of a spheroid, and have large, obvious regions of active star formation.

Barred and Non-barred Galaxies: Galaxies that have a disk component, which includes all galaxies except Elliptical galaxies, can also have a visible component within that disk that has the shape of a ’bar’. This bar component is only seen when the disk of a galaxy is facing towards the viewer (that is, towards the Earth). When the disk is seen edge-on, the bar is hidden by the dust in the rest of the disk. In face-on galaxies, about half of all galaxies with disks have bars in their disks. The luminosity of a bar can be small, or it can dominate the luminosity of the disk. We do not yet understand why some galaxies have bars in their disks, and why others do not.

C. The Tuning-Fork Diagram

1. The Hubble classification of galaxies can be arranged in the form of a ’tuning fork’:

```
                        S0  -  Sa  -  Sb  -  Sc  -  Irr (or I) (disks, no bar)

     E0  -  E3  -  E6 (no disk)

                        SB0 -  Sba -  SBb -  SBc -  IBm   (disks, with bars)
```

2. The physical properties of galaxies differ systematically along the Hubble tuning fork:

E galaxies have: no visible disk; little or no visible star formation (hence they are made mostly of old stars); little HI observed at 21 cm.

S0 galaxies have: a visible disk, but little or not visible star formation (hence they, like elliptical galaxies, are made of primarily old stars); somewhat more HI than elliptical galaxies, but not much, in general.

Spiral galaxies have: visible disks; the progression from Sa to Sb to Sc is a steady progression in increasing size of the disk relative to the spheroid, visible amount of on-going star formation, and higher percentage of mass in HI and H2.

Irregular galaxies: are essentially all disks, with high star formation rates and high percentages of mass in HI and H2.

3. Absolute Magnitudes of Galaxies

The absolute magnitude of a galaxy is measured in the same manner as for a star. On this scale:

E galaxies can have absolute visual magnitudes that range from -24 to -10. Those E galaxies with absolute magnitudes of -24 to -17 are usually termed 'giant' E's; those with absolute magnitudes fainter than -17 are usually termed 'dwarf' E's.

S0 and SB0 galaxies can have absolute magnitudes that range from -23 to -18.

Spiral galaxies (Sa, SBa, Sb, SBb and Sc, SBc) can have absolute magnitudes that range from -23 to -17.

Irregular galaxies can have absolute magnitudes that range from -20 to -12.

4. Mass-to-light ratios of Galaxies

We can measure the mass of spiral and S0 galaxies by measuring how fast their disks rotate, and using Kepler's third law. We have to resort to other methods to measure the mass of elliptical galaxies (which do not, in general, rotate) that are based on Newton's laws of motions.

The ratio of the mass of a galaxy, expressed in terms of solar mass, to the luminosity of a galaxy, expressed in terms of solar luminosity, is commonly termed the 'mass-to-light' ratio of a galaxy (by definition, the mass-to-light ratio of the Sun is 1). When we compare this mass-to-light ratio with the amount of light given off by a galaxy, we find that about half of the mass of a galaxy is not emitting electromagnetic radiation - the dark matter that was discussed in the context of our own Galaxy.

The same kinds of techniques can be applied to galaxies that are part of galaxy clusters (in the same way that stars can be part of star clusters). Again, we find that the gravitational field of a galaxy cluster is much stronger than we would have estimated on the basis of the light coming from the cluster. Again, the existence of dark matter. We shall need to discuss dark matter again when we talk about the structure of the universe as a whole.

D. Active Galaxies, Peculiar Galaxies and Mergers

About 30,000 galaxies have been classified in the Hubble system; of this number, 93% can be placed in the simple tuning-fork diagram. The other 7% of the galaxies are peculiar in form in one way or another. Most of these peculiarities appear to come from the gravitational interaction of one galaxy with another. Such gravitational interactions can: tidally-disrupt a galaxy, strewing gas and stars away from the galaxy; dump gas onto a galaxy, causing a strong burst of star formation; cause physical collisions between galaxies, during which the two galaxies can merge into a new, bigger galaxy. As will be seen, many astronomer think that one of the standard Hubble classes of galaxies - elliptical galaxies - are formed by this last process, a process termed 'mergers.'

1. Active Galaxies

Most galaxies are observed to have some kind of energetic phenomenon going on in their nuclei, much like what is going on in the center of our own Galaxy. However, the energy that is generated by the nucleus of a galaxy can vary enormously from galaxy to galaxy: the nucleus of our Galaxy is about 100 times more energetic that the nucleus of our neighboring galaxy, M31, the Andromeda galaxy; yet the nucleus of our Galaxy is 1,000 to 1,000,000 times less energetic than the nuclei of 'active' galaxies.

The designation 'active' galaxy is derived from the fact that the energetic activity in the center of the galaxy is so strong that it actually dominates the stellar luminosity of the inner part of the galaxy, if not that of the whole galaxy!

Active galaxies can be spiral galaxies, elliptical galaxies, S0 galaxies or irregular galaxies. The energetic activity is seen at visual wavelengths, radio wavelengths, infrared wavelengths and X-ray wavelengths and, hence must involve a wide range of temperatures at which energy is generated.

What is the source of this energetic activity? Why are some galaxies much more active thatn others (only about 3-5% of all galaxies are very 'active')? We do not definitely know the answers to these questions, but there is mounting evidence that two kinds of phenomena can produce 'activity': i) A massive black hole (from one million to one billion solar masses) that resides in the nucleus, around which is a massive accretion disk. ii) A highly concetrated, large number of massive stars in a very small area around the nucleus. Types of active galaxies include Seyfert galaxies and N galaxies.

2. Mergers and Cannibalism

If the collison between two galaxies is too direct, their two separate gravitational fields combine to produce a single gravitational field, into which the stars and gas from both galaxies must fall. Computer models of such collisions indicate that the product of this 'merger' will look rather similar to an elliptical galaxy after about 2 billion years. Mergers like this have almost certainly played a role in the formation of cD galaxies, which are supergiant elliptical galaxies that exist in the centers of clusters of galaxies.

E. Determining Distances to Galaxies

Galaxies are the brightest objects in the universe that can readily be identified. As such, how galaxies are distributed in the universe by and large determines how we define the structure of the universe.

However, before we can proceed with our use of galaxies to define the large-scale structure of the universe, we must have reliable means of measuring distances to galaxies.

Methods that are used to measure distances within our own Galaxy work primarily with individual stars, usually Cepheid or RR Lyrae variable stars.

Individual stars can currently be detected only for galaxies within about 2 megaparsecs (Mpc). This constraint will loosen somewhat when the Hubble Space Telescope is launched, hopefully in 1989, as it is expected that the HST will be able to detect bright individual stars in galaxies as distant as 20 Mpc (or a volume of space 1,000 times that of the current available measurements).

The most reliable method we now have for measuring distances to galaxies more distant than 5 Mpc is one which relates the gravitational field of a galaxy to its apparent luminosity. Why this method works is still unclear - our knowledge of the structure of galaxies is still too little to give an interpretation in terms of physics. That this method works is, by now, reasonably established. The other distance-measuring methods discussed in the book for distances greater than 5 Mpc are of much lower reliability, as summarized in Table 35.1.

F. Galaxies: A Census

1. The Local Group

Within a radius of one Mpc of our Galaxy (formally called the Milky Way Galaxy) there are about 30 galaxies, mostly clustered about the two dominant galaxies in this group: our Galaxy and M31. The nearest galaxies to our galaxy are the Magellanic Clouds, which are only 50 - 70 kpc distant (it was the Large Magellanic Cloud that contained the supernova of 1987).

M31 is about 700 kpc from the Milky Way. The galaxies within a Mpc from our own galaxy are thought to be gravitationally bound to the dominant pair of M31 and the Milky Way and, to a lesser extent, to the smaller Sc spiral galaxy M33 (about 900 kpc from the Milky Way). Most of these Local Group galaxies are very faint, dwarf ellipticals; some are faint irregular galaxies (like the Magellanic Clouds); some are faint Sc spirals.

2. Clusters of Galaxies

Galaxies tend to cluster together over a very wide range of spatial scales in the universe: The Local Group is about one Mpc big, and within it galaxies cluster on scales of 100 kpc or so (for example, the Magellanic Clouds near the Milky Way). The Local Group, in turn, tends to cluster with about 10 other similar-sized groups of galaxies on a scale of 10 Mpc. In turn, these groups of groups tend to cluster together through their mutual gravitational attraction to form a cluster of galaxies, containing perhaps 10,000 galaxies, that might be on a scale of 20 Mpc. In turn, clusters of galaxies tend to collect together into superclusters of galaxies, on a scale of 50 Mpc, and superclusters of galaxies tend to collect together into ’clusters’ of superclusters on a scale of 100 Mpc. At this point, our ability to measure even larger structures in the universe is limited, so we cannot really say if larger structures exist.

This layering of galaxies on ever larger scales is thought to be due primarily to one force - gravity (again!). If galaxies originally formed in small groups, the force of gravity will tend to clump these groups together over the lifetime of the universe, forming ever larger and larger clustering.

3. ’Girders’, ’Pancakes’ and ’Sponges’

My current picture of the universe, as defined by galaxies, is one of mostly empty space, pervaded by long ’strings’ and ’pancakes’ defined by clusters of galaxies. Looking out into the universe is much like standing at the base of an unfinished skyscraper, with partially completed, transparent floors, and a girder superstructure. Everywhere one looks, one sees 'strings' of galaxies, and occasionally, a 'pancake' structure.

In a recent set of papers, Richard Gott and colleagues have proposed that the structure of our Universe is similar to that of a sponge - with interconnected areas that are empty of visible matter and interconnected areas of visible matter. This description promises to be a useful concept for understanding the origin of this large-scale structure. Remember, however, that absence of visible light in our Universe does not necessarily mean absence of gravitational mass - dark matter could very well exist in the regions between galaxies.

XIII. THE EXPANSION OF THE UNIVERSE

1. The 'Redshift'

If one measures the Doppler shift of galaxies by taking spectra, one quickly discovers that essenatially all galaxies are moving away from us, in all directions we look. This motion is so universal that it is commonly referred to as the 'redshift', since the light from the galaxies will be Doppler-shifted to the red.

Moreover, when early measurements of distances to nearby galaxies were made, it was found that the redshift of a galaxy was directly proportional to its distance from us:

$$V = Ho \times d,$$

where V is the redshift velocity of the galaxy, d is the distance of the galaxy, and H is called the 'Hubble constant', and is a number that relates distance with redshift velocity.

The constant Ho has the interesting units of km/sec/Mpc. This means that, for every Mpc increase in distance from us, a galaxy will, on average, be moving Ho km/sec faster from us. Current estimates of the value of Ho range from 50 to 100. The uncertainty in the value of Ho is related to the fact that we do not yet have a good understanding of galaxies, so that our methods for measuring distances are not yet completely understood. If we take Ho to be equal to 50 km/sec/Mpc (as an example), this means that a galaxy that has a redshift velocity of 1000 km/s will be, on average,

1000 km/s divided by 50 km/s/Mpc = 1000/50 Mpc = 20 Mpc distant.

The Hubble constant really has units of 1/time - it is one of the 'clocks' by which we can measure the age of the universe.

2. Why is the Universe Expanding from Us?

The movement of essentially all galaxies away from us indicates that the Universe is expanding: Consider the common analgous example: Suppose you lived on the surface of a balloon, and the balloon began to expand. You would see all other parts of the balloon move away from you, you yourself would increase in size, and the further the part of the balloon from you, the faster it would move. Everyone living on the balloon would see all other parts of the balloon moving away from them.

The surface of a balloon is a two-dimensional world expanding into three dimensions. However, our Universe is already in three dimensions, yet it is expanding. Where is the center of the expansion? - literally, everywhere, since the Universe itself is expanding - you, the Sun, our Galaxy, everything in the Universe. Why do we not see ourselves expanding? Because the universe is so big, and the expansion so slow on our local size scales, that the chemical forces that hold our molecules together easily overcome the expansion. On a large scale, the force of gravity must overcome the expansion rate of the Universe to form galaxies and clusters of galaxies.

Right now, the expansion of the Universe is correlated with time - as time increses, the Universe expands. This will not necessarily always be so: If there is enough matter in the Universe, the force of gravity of all of that matter can slow down the expansion and eventually either stop the expansion, or force the Universe to begin to collapse again (if the latter, it will take about 50 billion years, long after the Sun has become a white dwarf star).

2. ’Peculiar’ Motions of Galaxies and Clusters of Galaxies

As discussed above, the force of gravity that comes from local concentrations of mass can modify the local rate of expansion of the Universe, and can cause matter to contract, rather than expand. We use that fact to measure the mass of our Sun and Earth, and the masses of galaxies. We can also use that fact to measure the mass of the Universe as a whole, but first we need to have a way of measuring distances to galaxies that is independent of the Hubble expansion.

Until 10 years ago, no such reliable methods existed, and most astronomers thought that the Universe expanded in a rather smooth, quiet manner. Over the past ten years, applying the newer methods of measuring distances to galaxies discussed above, we have found that the expansion of the Universe is not completely quiet. There apparently exist sufficiently massive superclusters of galaxies that contain enough mass to change the motions of galaxies near them by about 5% relative to the overall expansion of the Universe. The difference between the redshift of a galaxy due to the expansion of the Universe, and that due to more locally-generated gravitational force, is called the ’peculiar’ motion of a galaxy.

3. A Sense of Time

Because the speed of light is finite - 300,000 km/s in a vacuum, the further an object is from you, the younger that object was when it emitted, or reflected, that light. In the case of you as a spectator in a football game, it takes about 1/10,000,000th of a second for the light reflected off the players on the field to reach your eyes, if your are sitting in the stands. You do not see the players instanteously, just the same way you do not see anything else instanteously. However, since our eyes cannot distinguish differences in time of less than about 1/100th of a second, it simply looks instantaneous to us.

In the case of the nearby galaxy M31, we always see that galaxy as it was about 2,300,000 years ago; in the case of the most distant known galaxies, we always see them as they were 10 billion years ago! As we look out, we must always be looking at further distances from us, and we therefore must always be viewing galaxies in progressively younger stages of their evolution. Over a 60 million year time period one does not expect galaxies to have changed very much; but over a 5 billion year time period, one does expect large changes (remember, the Sun is only 5 billion years old).

XIV. EXPANSION AND QUASARS

A. A Quick Overview of the Structure of the Universe

To summarize the above disussion:

We live in an expanding universe, one in which all galaxies are moving away from us with velocities that are directly proportional to their distance from us.

Galaxies occupy only a very small fraction of space in the Universe, but galaxies are the only practical means by which we can measure the structure and size of the Universe. We can measure the motions of galaxies, and also measure how many galaxies there are.

If one traces the expansion of the universe back to its beginning, then the current best estimates of the age of the Universe are in the range of 12-18 billion years old. Essentially, the best estimate for the age of the Universe comes from the obvious fact that the Universe cannot be older than the globular clusters in our Galaxy.

B. Quasars - The Most Distant Objects in the Universe that we can see

What is a quasar? The name 'quasar' is shortened form of the longer name 'quasi-stellar object'. These are objects that appear stellar on an ordinary photographic plate, but are moving away from us with velocities comparable to that of the speed of light. Quasars have relatively bright apparent magnitudes, so that if they are at the distances implied by their redshift, they are, as a class, by far the most luminous objects in the Universe. The first discovered quasar, 3C 273 is 40 times more luminous than the brightest galaxy. About 2,000 quasars have now been discovered by various techniques; it is estimated that we could observe about 40,000 quasars to an apparent magnitude limit of 21.

Why are quasars important? If quasars are at the distances implied by their redshift, then many of them are both the most distant objects known, and the most luminous objects in the Universe. The highest redshift quasar could be over 15 billion light years away, and the light that we see wold have been generated less than 1 billion years after the Universe was formed. Thus, quasars are probes of the earliest times in the Universe.

If quasars are that far away, they have to emitting tremendous amounts of energy, up to 10,000 times that of single normal galaxy. How do they do it? We think they do it by gravity (again!): Remember, the gravitational field around a condensed object can create tremendous amounts of energy if matter is falling into it. Most astronomers now think that very large black holes exist inside some galaxies. If, in an early part of their history, these black holes were 'fed' large amounts of gas (say, by the merger of two young spiral galaxies), the massive accretion disk around such a black hole could generate the luminosity of a quasar.

The realization that large, condensed objects could generate this much energy came about in the late 1970's and early 1980's. Before this time, the lack of a viable source of energy for quasars led some astronomers to think that quasars did not get their redshift velocity from the expansion of the universe, and that they were somehow ejected from nearby galaxies with those velocities. The 'quasar controversy' raged for over 15 years, but today there is nearly overwhelming evidence that quasars are at the distances of their redshifts.

C. Quasars and Gravitational Lensing

Most quasars are much more distant than most of the galaxies that we can see (because the quasars are intrinsically much brighter). Since there are perhaps 100 billion galxies and 100,000 quasars that we could detect with present telescopes, the probability is high that a galaxy would be seen in projection along the line-of-sight to a quasar. When this occurs, the gravitational field of the galaxy would act as a 'lens' on the light from the quasar, and either multiple images of the quasar might be seen, or the light from the quasar may be amplified.

The first indentification of a gravitational lens was made in 1979, using the discovery of a 'double quasar', in which the 'two' quasars had precisely the same physial properties. The two quasars are actually two images of one quasar being gravitationally-lensed by a foregound galaxy.

At least seven gravitational-lensed quasars have now been found, with an the discovery of another, multiple-lensed quasar just recently announced. However, as reported several years ago, several of the 'lenses' have no apparent galaxy counterpart! Could the gravitational lenses be concentrations of dark matter? The answer is still not known.

XV. COSMOLOGY

A. Theories of the Origin of the Universe

1. The Cosmological Principle

The Cosmological Principle is the assumption that the Universe is homogeneous and isotropic everywhere. Homogeneous means that the distribution of matter is similar is all directions, as long as one takes a large-enough volume of space. Isotropic means that the properties of the Universe are the same no matter in which direction we look.

2. The Steady-State Theory

The Steady-State theory was proposed in the later 1940's, and is based on the further assumption of the 'Perfect Cosmological Principle' - the Universe is homogeneous and isotropic everywhere, for all time. In the Steady-State theory, matter must be created out of vacuum in order to continually supply a constant average mass density to the expanding Universe.

3. The Big-Bang Theory

The Big-Bang theory was first proposed in the 1930's, soon after the expansion of the Universe was discovered. The Big-Bang theory proposes that the Universe began with an enormous explosion.

B. The 3 Degree Blackbody Cosmic Background Radiation

The Steady-State theory and the Big-Bang theory stand at opposite poles of the possible histories of our Universe. For about 25 years, no hard evidence for or against either of the theories was found, since both (by necessity) incorporated an expanding Universe.

However, the Big-Bang theory did make one prediction that the Steady-State theory did not: Since we are part of the 'Big Bang', we should still be enveloped in the 'afterglow' of the explosion. A prediction of the temperature of this 'afterglow' was made in 1965, which suggested that the Universe now would emit isotropic radiation with a blackbody temperature of 3 oK. Coincidentally, at about the same time, Arno Penzias and Robert Wilson had discovered a source of noise at their new, sensitive radio telescope, that was isotropic and had a blackbody temperature near 3 oK! This isotropic radio source was quickly indentified as being the 'afterglow' of the Big Bang.

Penzias and Wilso received the Nobel Prize for their discovery, but they only detected the 'cosmic background radiation' at two wavelengths. More recent experiments have completely mapped out all the wavelengths of emission of the background radiation in most directions of the sky. With these more recent measurements, the motion of the Earth relative to the background radiation (remember 'peculiar' motions - these are really motions relative to the cosmic background radiation) has been detected, since the 3 oK radiation will be blueshifted in the direction we are moving (that is, shifted to a slightly higher temperature) and redshifted in the direction from which we are coming (that is, shifted to a slightly lower temperature).

C. Our Peculiar Motion in the Universe

These recent measurements show that our Galaxy is moving in the Universe, relative to the cosmic background radiation, at a velocity of 600 km/s! This large velocity is much larger than either the Earth's motion around the Sun (30 km/s), the Sun's motion around the Galaxy (220 km/s), or the Galaxy's motion relative to our neighbor, M31 (80 km/s).

It is only over the past year that we have begun to understand the origin of our velocity of 600 km/s relative to the cosmic background radiation. Again (and by now it should come as no surprise) the source of motion is gravity, but this time the gravitational force on very large scales - up to 100 Mpc in size. I have been part of a team of seven astronomers (dubbed the 'Seven Samurai' by our colleagues) that have shown that these 'large-scale' motions exist.

D. Ho, qo and 3W - The three measurable parameters of the Universe

We can, in principle, independently measure three properties of the Universe:

1. What is the age of the Universe?

The Hubble constant, Ho has units of 1/time. Thus, 1/Ho is a time, called the 'Hubble' time. For a Hubble constant of 50 km/sec/Mpc, 1/Ho = 20 billion years; for a Hubble constant of 90 km/sec/Mpc, 1/Ho = 11 billion years. However, this Hubble time is not necessarly the 'true' age of the Universe, since it assumes that the Universe has been expanding at the same velocity since it was formed. Since we know this is not true, we also need to know

2. How fast is the Universe decelerating with time?

The parameter qo is a measure of how fast the Universe is decelerating with time, as explained in Chap. 37.2b. If we combine qo and Ho we can predict both the age of the Universe, and the mean density of matter in the Universe. However, we can also measure the mean density of matter in the Universe, by asking

3. How dense is the Universe?

The parameter 3W measures how dense is the Universe. We can measure 3W by measuring how much matter is in the Universe in a given volume of space.

Again, we use gravity to do this, but gravity on the largest scales that we can measure in the Universe. Making this measurement is tricky - you cannot measure the mass of objects that are outside your measuring volume - and we are still trying to do the observations correctly.

4. Putting it all together?

If we separately measure Ho, qo and 3W, we then have an independent means of finding the age of the Universe, and predicting the ultimate fate of the Universe. Unfortunately, measurement of these three quantities is very difficult, and horribly complicated. Our understanding of the age of the Universe and its ultimate fate is still unfolding, which makes this one of the most exciting fields of observational astronomy today.

E. The Inflationary Universe

Chap. 37.3, 37.4 and 37.5 give a reasonable description of the standard model of the Big Bang, and, by now, 'standard' models of the Inflationary model of the Big Bang. The Inflationary Universe is a further modification of the Big Bang theory.

Since the writing of this chapter (in about 1985), the theoretical work on understanding the Big Bang has continued to progress. We now understand that the original Inflationary model is not as unique as we once thought - there can be different 'kinds' of inflation, depending on precisely how the Universe is constructed. As such, the search for a more complete understanding of the Big Bang continues.

F. Physics vs. First Principles

Why do we want to find a theory that explains the Big Bang? Why not simply say, 'That's the way the Universe is.'?

The reason is very basic to science: The structures of the world around us that we see, and the structures of the world within us, are determined by certain physical properties of matter, and the existence of the four forces of nature. Is the structure of the Universe as a whole also determined by these 'laws of nature'? This is a central question of science, and one which we think we can begin to answer in our lifetimes.

In order to answer this question, we must continually ask, 'Can the physical laws of nature predict what we observe in the Universe?' It was because the Inflationary Hypothesis used physical laws to predict certain key features of the Universe, that it is regarded as a viable theory of the Big Bang. Our current understanding of the Universe in terms of physical laws is imperfect - we know that. Indeed, it is this fact that gives us hope that, with better observations and better understanding, we will eventually be able to understand the Universe in terms of its physical laws.

Please note, our search for an understanding of the physical structure of the Universe does not make the Universe any less wondorous to us. On the contrary, the simulataneous intricate and simple nature of the Universe is amazing: The old cliche' is continually be shown to be true - The universe is not only stranger than we think, it is stranger than we can think!

XVI. 30-MINUTES OF EINSTEIN AND RELATIVITY

You all have heard and read about Einstein the man, and the Theory of Relativity, but what does 'relativity' mean?

Einstein’s Theory of Relativity includes: The Special Theory of Relativity, which deals with objects moving at constant velocity; and the General Theory of Relativity, which deals in a general way with objects that are accelerating (the Special Theory is encompassed in the General Theory). Anyone with High School algebra, some calculus and some imagination can understand the Special Theory; understanding the General Theory requires somewhat more mathematics and physics.

The genius of what Einstein created is partially reflected in the simplicity of the assumptions that led him to his theories: Einstein assumed that 1) the laws of physics are the same for all observers experiencing uniform motion (that is, not being accelerated; the description of motion is relative); 2) the speed of light is independent of the velocity of the object emitting the light - the speed of light is ’constant’; 3) the Universe is everywhere the same on large scales - that is, the Universe is isotropic.

The logical consequences of the first two assumptions include: 1) The famous equation E = mc2, which equates mass with energy. 2) No object with mass can travel at the speed of light (meaning light is massless). 3) The realization that space and time are intertwined - the concept of ’spacetime’; this has certain consequences, such as the fact that simultaneous events to one observer will not necessarily appear to be simultaneous events to another observer that is moving relative to the first. 4) When one is travelling with a velocity relative to another, time is stretched space is condensed (’time dilation’ and ’space contraction’).

General Relativity deals with objects undergoing acceleration, such as any object being acted on by a force, like gravity. The concepts of ’inertial mass’ and ’gravitational mass’ are important here, as well as the Principle of Equivalence (which assumes that gravitational mass is equal to inertial mass). Gravity literally ’curves’ spacetime - the stronger the gravity, the more the curvature. Since light must follow the natural curvature of spacetime, light will bend near the surface of massive objects as seen from an outside observer. This is what causes a gravitational lens. The extreme case of this bending is around a black hole - light is bent so completely around the black hole that it cannot escape.

The predictions of General Relativity have certain consequences for our ability to travel to distant stars: If we go at speeds slow compared to the speed of light (300,000 km/s), then it will take an awfully long time to get anywhere (10,000 km/s = 4 hrs/A.U. = 300 years to the nearest star; the fastest any spaceship has gone todate is less than 30 km/s!). If we go at speeds comparable to the speed of light (0.999 ... times the speed of light), then time will be dilated (the clock will run more slowly) and distances contracted. At such speeds, in your spaceship time it would not take much longer to travel to the most distant galayx as to the nearest star (several years). The catch is, time back on Earth, or on any object not travelling at those velocities, would be much, much later, from millions to billions of years later!

Given our current understanding of physics and the structure of the Universe, there are certain restrictions on how we could explore the rest of our Galaxy (or how the rest of the Galaxy could explore us), let alone the many 100’s of billions of other galaxies in the Universe. This is not to say that those restrictions cannot be eventually circumvented, but just to say that we see no way of doing so now.

XVII. LIFE IN THE UNIVERSE

In this part of my notes, I usually add some personal observations to what is written in the book.

A. Origin of Life on Earth

Science attempts to explain how life could have originated, not why. Many laboratory experiments, and observations of life that exists in the ocean bottom and survives from heat generated by volcanic vents, leave little doubt that fairly complex organic molecules will form naturally out of the primitive conditions found on the early Earth (or on the moon Titan, or on Jupiter, or on ...). How such molecules combine to form life as we know it is not yet understood.

B. Life on Other Worlds?

In considering if life could (or does) exist on worlds other than the Earth, one must allow for two possibilities: 1) Life must always be based on the carbon molecule and, hence, be biologically related to life on Earth; or 2) Life can be based on almost any combination of molecules and situations and, hence, not necessarily be related to life here on Earth.

Several years ago, Carl Sagan coined the term 'oxygen chauvinism' to refer to the bias that we have towards life that requires oxygen to live. However, there is life on our Earth that is carbon-based to which oxygen is a poison - the anaerobic bacteria (such as the botulism bacterium). It is believed that such bacteria originated when the Earth did not have an oxygen atmosphere.

Given assumption (1) in section A above, what are the factors necessary to have life as we know it?

A planet close enough to a star to be able to have water exist in all three states - gas, liquid and solid - simultaneously on the surface for geologically-long periods of time. This requirement restricts such a planet to a narrow range of orbits around any particular star, since the temperature of the planet is sensitive to its distance from the star.

A star that provides heat that will last long enough for life to evolve. Life as we know it is unlikely around an O or a B star (their lifetimes are too short); if would be more likely around stars of spectral type F0 or later.

A concentration of Earth-like minerals. Although minerals comprise only a small part of the body mass of a living organism, they apparently perform many important biological functions. Life as we know it would probably be unlikely on the surface of Jupiter.

C. How many planets could exist that are suitable for life as we know it?

There are at least one billion stars around which a planet like the Earth could exist. How many stars actually have planetary systems? We do not know for certain, but recent discoveries have shown that at least 40 nearby stars have disks of dust around them - one of the necessary preconditions for planets to form. Much more sophisticated observing techniques, including use of the Hubble Space Telescope, might begin to be able to detect planetary systems around nearby stars.

D. What are the chances for Extraterrestrial Life?

Little modification has been made to the 'Drake' equation, since Frank Drake formulated it nearly 30 years ago. Basically, the Drake equation puts donw all the possible factros that could affect our chances of contacting extraterrestrial life.

Some of the factors that go into the Drake equation can be varied by a great amount, since we just do not know how to estimate the probability of life on another planet.

My personal opinion is that one of two situations exists in our Galaxy: 1) We are the only thinking civilization existing at this time. or 2) The Galaxy has thousands, if not millions of such civilizations. I do not know how to choose between these two very different possibilities.

E. Should We have been Contacted?

If the number of civilizations in the Galaxy are really in the millions, many people argue that we should already have been contacted by another civilization. Again, this reasoning requires that we impart human actions on hypthetical non-human civilizations. We will be able to decide if we should, or should not have been contacted when we begin to truly explore outer space, and the distances between the stars.

XVIII. Closing Questions

If you lived on the furthest quasar, and looked out into the Universe, what would you see? The same Universe you see now. Although the light you see from that quasar is only 1 billion years old, the quasar is as old as our Galaxy, in its own reference frame (remember relativity!).

That question has an answer, but the following questions do not. What would have happened if the dinosaurs had not been exterminated 60 million years ago? There is growing fossil evidence that man-sized dinosaurs walked on two legs and had begun to develop arms with fingers. Would dinosaurs have evolved into intelligent, oxygen-based beings in another 10 million years? If so, would they then have had 50 million years to develop and explore the Universe?