Extraterrestrial Life
Fifth Edition

Neal J. Evans II

The University of Texas
Austin, Texas

Pearson
Custom
Publishing

PEARSON
Custom
Publishing
www.pearsoncustom.com

Jesse C. White, M.Ed.
Senior Acquisitions Editor

3917 Dry Creek Drive
Austin, TX 78731
512-374-1557 • Fax: 512-374-1757
Email: jesse.white@pearsoncustom.com

Table of Contents

Page

Preface

These notes have grown out of a course on extraterrestial life that I have taught for the last six years at the University of Texas. As I learned more, and added more to the subjects discussed in class, students found it increasingly difficult to assimilate the material without a source they could turn to outside of class. Other published books on the subject followed approaches different from those used in this class or left out significant areas which I have found to be important. For these reasons, as well as perhaps a case of temporary insanity, I convinced myself to put together some notes from my class lectures. The notes are an amalgam of components of varying quality: some are nearly verbatim transcriptions of lecture presentations, complete with bad syntax, incomplete sentences, and other horrors which return to haunt me whenever I correct the English style on a student's term paper; others have been rewritten with at least an attempt at correct style and an occasional illustration. In view of this situation, I apologize to the students for the inconsistencies of style and the paucity of illustrations. As partial compensation, the chapters and the subsections do coincide with the outline notation used in my class lectures.

I would like to thank some people for contributions of one kind or another. Tom Barnes initiated the course on extrarrestrial life at UT and provided much initial assistance in my development of the course. John Scalo also contributed through an abortive attempt at a co-authored textbook - back when none existed and the market was ripe. Oh well. More recently, Marycarol Rossiter gave the biological aspects a careful look, supplied the much improved discussion on protein synthesis, and generally encouraged me in believing that I wasn't grossly distorting modern biology. Dave Bloch and the students of Botany 394 gave me much insight into the complexities of biochemistry and recent research in *in vitro* nucleic acid replication and the possible role of clays in the origin of life. Various teaching assistants, notably Lee Mundy and Mike Scholtes, have contributed research in some areas and Steve Federman has fed me relevant articles from his beloved *New York Times*. On another level, I thank my son, Danny, for playing quietly by himself while Daddy sorted out such things as the relationship between Pongidae and Hominidae (pretty close, as it turned out). Most of all, I thank the many students who have taken my course, especially those whose term papers opened new areas of exploration for me and those who encouraged me to persevere.

Chapter 1

The Nature of the Subject Matter

A. Overview

No one knows whether life exists elsewhere in the Universe. This uncertainty means that a certain amount of speculation is unavoidable. The freedom to speculate has been exploited heavily by science fiction writers. Because this is a science class, we will attempt to inform the speculation with the facts that we do have. One objective of the class is to bring a scientific perspective to a subject more often considered to be in the realm of pure fantasy.

The subject of extraterrestrial life necessarily involves the study of both the Universe outside the confines of the Earth (the "extraterrestrial" part) and life. Since the only life that we know about exists on the Earth, we will have to base our speculations about extraterrestrial life on a study of life on the Earth— how it arose and evolved. The assumption that underlies this approach is that life anywhere will share some features with life on the Earth. This assumption may be incorrect and we will have to be wary of "Earth chauvinism." On the other hand, we will discover some deep connections between life on Earth and the Universe; understanding these will be a second objective of the course. In the process of pursuing these connections, we will see that a few principles, notably those involving conservation and transformation of energy, can be used to unify many apparently disparate subjects.

Because we are concerned not only with life but particularly with intelligent life that we could communicate with, we will study how life evolved from the very simple organisms that first appeared on the Earth into more complex organisms, such as our own species. Consequently, we will discuss Darwinian evolution in some detail. This can be a controversial topic and we will not avoid the controversy. However, I want to make it clear from the beginning that we will take an evolutionary point of view. In addition to the usual understanding of this term as biological evolution, we will look at evolution in a much broader sense, including cosmic evolution. Very simple forms of matter emanating from the Big Bang that began our Universe changed into more complex forms of matter, eventually leading to a transition from non-living molecules to life. Thus, evolution is a major focus of the course. In a sense, the course will present the scientific version of the creation stories of religions, a scientific look at where we came from and how we got here. In the process, we will find that there are numerous points in this evolutionary story that are poorly understood. The scientific "Genesis" is not a finished product, but a "work in progress."

While examining the evolutionary processes that led to intelligent life on Earth, we will lay the groundwork for estimating the chances that such life exists elsewhere. Since there is much room for speculation on this point, every student can make his or her own estimates; you will be tested on how well the estimates are backed up by the facts that we have at our disposal and how consistent the arguments are. After arriving at these estimates, we will examine the means that could be used to communicate with other civilizations in our Galaxy and assess the possibility of traveling to another star, including the idea that we have been visited by extraterrestrial beings. Since this latter topic is also quite controversial, I should make it clear that I am very skeptical about claims of visits to the Earth by aliens.

The rest of this chapter is devoted to a brief discussion of the mathematics needed for the class (minimal), a look at the nature of the subject (a review of the dimensions and structure of the Universe and of definitions of life), a discussion of the chemical basis for life, and an introduction to the Drake equation. This equation provides much of the underlying structure of the course.

B. Notes on Mathematics and Units

This is not a course that requires much mathematics, but two concepts are important. The first is equations, including the Drake equation. The student will not have to do any significant algebra with these equations but will use the equations to express relations between things and compute things of interest. As an example of an equation, consider the most famous of all equations: $E = mc^2$. What this compact equation expresses is the fact that matter and energy are related; a given amount of mass (m) has associated with it an amount of energy (E) which is calculated by multiplying the mass by the square of the speed of light (c). We may also use equations to specify a value for something; thus $T = 100$ means that the temperature is 100 degrees. A variation on this use is to specify a rough value; thus $T \simeq 100$ means that the temperature is not known very precisely but is *about* 100 degrees. Another expression that we sometimes use is the sign for "proportional to" ($\propto$). This technique allows us to focus on how one quantity depends on another without worrying about the possible effects of other variables. For example, the luminosity (energy per unit time) produced by a star is proportional to the fourth power of its mass ($L \propto M^4$). All other things being equal, a star twice as massive as the Sun will produce $2^4 = 16$ times as much luminosity. This approach avoids complex equations, difficulties with units, etc.

The second mathematical technique that we will use is scientific (or exponential) notation. This is simply a convenient way to express very large or very small numbers. There is an explanation in Appendix 3, which you should study in some detail if you are not sure about this notation.

We will be using metric units in the course and you should be familiar with them. They are summarized in Appendix 4, along with some useful quantities. The primary quantities in any system of units are time, distance, and mass. Fortunately, time in the metric system is the same as in everyday usage. The fundamental unit is a second (abbreviated by an s), and there are about 3×10^7 seconds in one year (yr). In the metric system, the primary unit of length is the meter (abbreviated by m). One meter = 100 centimeters (a centimeter is abbreviated by cm); also, one meter = 1,000 millimeters (mm). One kilometer (km) = 1,000 meters (10^3 m). Most often, we will use the light year (ly) as a measuring unit; it is defined as the distance that light will travel in the time of one year. Light travels at 186,000 miles per second (or 3×10^{10} cm per second). Multiplying the speed by the number of seconds in a year gives us the result that light travels about 10^{18} cm per year (the mathematically inclined will notice that I have rounded off 9×10^{17} to 10^{18}; this will be a common practice in this course, which is not concerned with high precision). A parsec, which you may encounter in some books, is about three light years. The fundamental unit of mass is the kilogram (kg), which is 10^3 grams (g). In discussing astronomical objects, we often use a larger unit of mass; we speak in terms of solar masses. One solar mass (denoted $M_\odot$) is the mass of the Sun, 2×10^{33} grams.

We will use the Kelvin temperature scale where temperature is measured in kelvins. Absolute zero, where all motion halts, is zero on the Kelvin scale. Water freezes at 273 kelvins (273 K); this temperature is also zero degrees Celsius and 32 degrees Fahrenheit. Water boils at 373 K (or 100 degrees C, or 212 degrees F). When you think of room temperature, think of 300 K.

C. Dimensions and Structure of the Universe

The Universe is immense and incredibly empty. Matter in the Universe is clustered in structures known as galaxies; a galaxy is a collection of a large number of stars (often a galaxy may contain some 100 billion stars). Matter between stars is called "interstellar matter." We use light travel time as a measure of distance because nothing in the known Universe travels faster than the speed of light. For example, it takes about 1.25 seconds for light to travel from the Moon to the Earth; it takes about 5.5 hours to travel from the Sun to the outermost planet, Pluto (which circumscribes the size of our solar system). It takes 4.3 years for light to get to us from the closest star, Alpha Centauri, while it takes about two million years for light to reach us from the nearest large galaxy, the Andromeda galaxy. The object we have sent farthest into space has only traveled about 9 light hours, and people have traveled only 1.25 light seconds (to the Moon).

When thinking about communication with extraterrestrial life, we must remember that radio waves travel at the speed of light (since they are part of the spectrum of electromagnetic waves) and radio waves have been traveling outward from the Earth for about 60 years. Thus, any sufficiently advanced civilization within about 60 light years could know of our existence. Our Galaxy, the Milky Way, is 10^5 light years across, and it is about 500 light years thick (our Galaxy looks like a very thin disk). There are about 12 galaxies within three million light years of us, known collectively as the "Local Group." Other galaxies are clustered in much larger numbers; for instance, there are 2,500 galaxies in the Virgo cluster alone. The Virgo cluster galaxies are about 3×10^7 light years away from us. The Hercules cluster contains about 10^4 galaxies, and these are about 3×10^8 light years away from us. Remember, there are billions of stars in one galaxy, and there are billions of galaxies in the Universe. It is estimated that there are at least 10^{20} stars in the known Universe, and it may well be that the Universe contains an infinite number of stars.

When we observe an object at a given distance in light years, we see the object as it was in the past. A signal from a civilization 100 ly away would have been sent to us 100 years ago. For example, if a civilization 30 ly away that received our first, faint broadcasts replied immediately, that reply would just now be reaching us. The great distance between stars provides a kind of quarantine; as we shall discuss in Chapter 10, interstellar travel is very difficult. If we are to explore other solar systems for life, we will have to use our imagination for the present.

D. The Definition of Life

To introduce the part of the course dealing with life, we address the question of how we define life. In some cases, it is hard to tell just what is a living thing and what is not. Traditional biological attributes of life are often used to try to decide if a thing is living; these attributes are the following:

1) Living things are composed of **organic molecules**. We shall see that these molecules are based on carbon chemistry.
2) Living things engage in **metabolism**, the exchange of matter and energy with the environment, resulting in production of the organic molecules of life.
3) Living systems engage in **reproduction**, yielding offspring more or less identical to the parents.
4) Because of **mutation** and other processes, the offspring are not precisely like the parents.
5) Living things exhibit **sensitivity** to changing conditions in the environment.

Several caveats accompany this list. The first rule might just be an example of Earth chauvinism; life on other planets might have a different chemical basis. Also, some primitive organisms do not engage in metabolism during part of their life cycle, yet they are living. Also, reproduction is not by itself a necessary condition: for example, a mule is sterile and cannot reproduce, and yet it is a living thing. Thus, one can find counterexamples for each of the attributes. However, we believe that some combinations of these rules must apply during some stage of the organism's life cycle for it to be considered a form of life.

An alternative approach to listing these properties of individual organisms is to consider the inter-relatedness of life— the ecological aspect. Here, the flow of energy from one form to another is emphasized in an attempt to derive a very general definition of life. An example given by Folsome, describing the work of Onsager and Morowitz, leads to the following definition of life:

> Life is that property of matter that results in the coupled cycling of bioelements in aqueous solution, ultimately driven by radiant energy to attain maximum complexity.

This approach has the advantage of avoiding details that may be peculiar to individual situations (the sterile mule is clearly alive by this definition) and of fitting into modern ideas in which the flow of energy through a system results in increased complexity. On the other hand, it still has some remnants of Earth bias (an "aqueous solution" implies life based on water) and is so esoteric as to mean nothing to the average intelligent person. Also, because of its generality, it might apply to situations that we would not generally associate with life.

One interesting aspect of Folsome's definition is that it refers not to an individual organism, but to "coupled cycles." This approach emphasizes that no organism, or even species, exists in isolation; instead the life of each is intertwined with the lives of other species through ecological cycles. Plants change inorganic raw materials into organic compounds using solar (radiant) energy; herbivores eat the plants; and the carnivores eat the herbivores. When the carnivores (and herbivores and plants) die, bacteria assist their decomposition, returning the raw materials (bioelements) to the environment. These great biological cycles link all life forms into the planetary phenomenon of life.

Not only biological, but also geochemical, cycles are involved. Some of the elements needed for life are lost from the biological cycles; they are eroded into the ocean, they settle to the bottom, and are buried as sediments. Over the life of the Earth, enough vital elements would have been lost that life should have ended. Life has not ended because the sediments do not stay put. The sea floors spread, and when the ocean sediments slide under the continents, they are driven downward to a layer of molten rock. There, the great heat and pressure modify the sediments, and the vital elements return to the surface as gases and rocks produced through volcanic activity and mountain building. The rocks are eroded to soil and the bioelements are available once more to the plants.

Thus, in some sense, we can describe life as a planetary property involving and relating all organisms, the atmosphere, the oceans, and the continents. This total system has been called Gaia, after the Earth goddess of ancient Greece. James Lovelock and Lynn Margulis suggest that Gaia can be thought of as a living being.

A less poetic name for Gaia might be the biosphere. Another definition for life is offered by Feinberg and Shapiro in their intriguing book, **Life Beyond Earth:**

> Life is fundamentally the activity of a biosphere. A biosphere is a highly ordered system of matter and energy characterized by complex cycles that maintain or gradually increase the order of the system through an exchange of energy with its environment.

Feinberg and Shapiro consider the biosphere to be the level at which the phenomenon of life is easiest to recognize, whether or not the biosphere is itself a living being.

We will leave the definition of life unresolved, but the subject will arise again when we discuss the experiments to search for life on Mars and ideas for detecting evidence of life through observations of planets around other stars (Chapter 5).

E. The Chemical Basis of Life

Here, we want to analyze the first criterion for life, based on the life we find on our planet. What is the chemical composition of life on Earth? Robert Davies has provided the following interesting facts. The average human being contains about 6×10^{27} atoms, including at least one atom of every stable element and some atoms of unstable (radioactive) elements, such as ^{14}C, ^{3}H, and ^{40}K. During a typical 70-year lifespan, at least 10^{12} atoms of radioactive ^{14}C decay in our bodies. Each of us probably contains some atoms from every species that ever existed. Of the roughly 90 stable elements, about 27 appear to be essential for humans (or any other mammal). For a bacterium, about 17 elements appear to suffice, and even fewer are required for a virus. Phosphorus (P) and potassium (K) appear to be the least accessible elements in nature.

Some elements are much more abundant in living things than others. The elements exist in the following proportions in living things: hydrogen, 60%; oxygen, 25%; carbon, 10%, and nitrogen, 2% (we give the percentages by number of atoms). There are also smaller amounts of calcium, phosphorus, and sulfur in living organisms. These percentages are characteristic of essentially all life. The chemical composition of humans and bacteria is similar, the main difference being that we contain much more calcium (still only 0.23%) because we have bones. There are traces of other elements as well, such as iron, which forms an essential part of hemoglobin. Magnesium is also a key part of chlorophyll (which plants use to convert sunlight into usable energy).

The composition of life is very different from the chemical composition of the Earth. The crust of the Earth is mainly made up of oxygen and silicon (oxygen forms 47% of the crust and silicon 28%); if we include the interior of the Earth, iron would be a major component. We are much more similar in composition to the ocean than to the Earth (67% hydrogen and 33% oxygen); in fact, we are mostly water, suggesting a clue to our origins. The current atmosphere of the Earth is mostly nitrogen (78%) and oxygen (21%), with smaller amounts of carbon (0.011%). We shall see later that the atmosphere when life arose was quite different. The composition of the ocean and atmosphere, as well as the composition of life on Earth is roughly similar to the chemical makeup of the rest of the Universe. The Universe and the Sun are mainly composed of H (hydrogen, 93%), He (helium, 6%), O (oxygen, about 0.06%), C (carbon, 0.03%), and N (nitrogen, 0.01%). Helium is not present in life in any abundance because it does not form molecules. A little later in the course, we will see where all of these chemical elements come from and how they were formed in the Universe. The origin of chemical elements is part of one of the two main themes in the course— the evolution of matter from fundamental particles into complex life forms and civilizations. The other major theme is the Drake equation.

F. The Drake Equation

We will refer to this equation extensively during the rest of the class. The Drake Equation is the following:

$$N = R_* \, f_p \, n_e \, f_\ell f_i f_c L,$$

indicating that N is the product of all the other symbols. The symbols have the following definitions:

N = number of communicable civilizations in our Galaxy. N is the product of the following factors.
R_* = rate at which stars form.
f_p = fraction of stars that have planetary systems.
n_e = number of planets, per planetary system, that are suitable for life.
f_ℓ = fraction of planets suitable for life on which life actually arises.
f_i = fraction of life-bearing planets where intelligence develops.
f_c = fraction of planets with intelligent life that develop a technological phase, during which there is a capability for and interest in interstellar communication.
L = average lifetime of communicable civilizations.

Once we have calculated N, we can also calculate r, the average distance to the nearest civilization.

Frank Drake suggested this equation in 1961 at an international conference. In principle, it allows us to calculate the number of civilizations in our Galaxy with which we could communicate (to avoid this awkward construction, we will refer to them as communicable civilizations). In practice, it provides not so much an answer to the question of how common extraterrestrial civilizations are as it does a guideline for formulating the relevant questions. In this class, we will examine some of the facts relevant to estimating each factor in the Drake equation; the final estimate of each factor will be up to each individual student. A work sheet with places to fill in your estimate for each of the factors is included as Appendix 5 of this book.

Chapter 2

Cosmic Evolution: 1. Protons to Heavy Elements

In this chapter, we will discuss the origin of galaxies and stars, but our main focus will be on the origin of the elements needed for life (see Chapter 1E). Three of the four most critical bioelements, carbon, oxygen, and nitrogen, became available only after the birth and death of early stars. Consequently, we will discuss briefly the origin and life of these early stars, leaving detailed discussion of the birth of stars to the next chapter. The role of stars in producing the bioelements is our first example of the deep connections between the Universe and life. We use the term **cosmic evolution** to describe the process by which matter in the Universe increases in complexity; in this chapter, we examine the first step of this process.

A. The Origin of the Universe and Galaxies

The accepted theory for the origin of the Universe is the Big Bang. The Big Bang occurred about 13 billion (1.3×10^{10}) years ago. At early times (within a few minutes after the Big Bang), the Universe was very dense and very hot, and matter existed in only a few types of simple particles. At **very** early times, the particles were probably quarks and other exotic objects. By the time one millisecond (10^{-3} s) had passed, the quarks were bound into more familiar particles, such as protons and neutrons. We begin our story during this era. The particles that existed then are listed below, along with their symbol, electric charge, and mass:

Name	**Symbol**	**Charge**	**Mass**
proton	p	+ (plus)	1.7×10^{-24} g
neutron	n	0	1.7×10^{-24} g
electron	e	– (minus)	1×10^{-27} g
photon	γ	0	0
neutrino	ν	0	~0

Note the use of Greek letters to represent the photon (gamma) and neutrino (nu). Greek letters are commonly used in science, and it is important to distinguish them from English letters to avoid confusion. At the present time, protons and neutrons form the nuclei of the elements. A few minutes after the Big Bang, the basic building blocks of the elements were available but unassembled into nuclei.

Within a few minutes of its birth, the Universe was extremely hot. Temperature refers to how fast particles are moving around in a random way. The more rapid the motion, the higher the temperature; more rapid motion implies that the particles have greater kinetic energy, or energy of motion. As long as a particle is not moving too close to the speed of light, its kinetic energy is given by the equation,

$$E = \frac{mv^2}{2},$$

where m is mass of the particle and v is its velocity. At a temperature of T, measured in kelvins, the average kinetic energy of a particle is given by

$$E = \frac{3kT}{2},$$

where k is called the Boltzmann constant. If we combine these two equations and use a little algebra, we can compute the average speed of a particle in a gas of temperature T to be

$$v = \sqrt{\frac{3kT}{m}} = \left(\frac{3kT}{m}\right)^{\frac{1}{2}},$$

where the superscript 1/2 means to take the square root of the quantities in the parentheses.

When particles collide, the outcome depends on the kinetic energy of the collision. Imagine two billiard balls colliding; at ordinary speeds or kinetic energies, they will bounce off each other. At high enough speeds, one or both will shatter. For the first 3 to 4 minutes after the Big Bang, the collisions between particles were so energetic that nuclei more complex than the particles listed above could not exist; they were broken apart as quickly as they formed.

Ever since the Big Bang, the Universe has been expanding; when a hot gas expands, it cools. As the Universe expanded, it cooled and became less dense. As the temperature dropped below about 10^9 K, some slightly more complex nuclei were able to hold together. These nuclei, combinations of protons and neutrons, accumulated for about 30 minutes, until the temperature dropped below 3×10^8 K and all nuclear reactions stopped. The composition of the Universe at this time was about 94% protons and 6% alpha particles. An alpha particle is a combination of two protons and two neutrons. There were also enough electrons to balance the positive charges in the protons and alpha particles. Eventually, about 3×10^5 years later, the protons became hydrogen atoms by combining with an electron, and the alpha particles became helium atoms by combining with two electrons.

As the Universe expanded and cooled, clumps of gas formed. In a poorly understood process, these clumps eventually formed galaxies. Smaller clumps within galaxies formed into star clusters. The first stars in our Galaxy formed before our Galaxy had settled into its disk structure. Later, stars formed near the center of the Galaxy and star formation then occurred throughout the rest of the Galaxy. The oldest stars in the disk of the Galaxy appear to be about ten billion years (10^{10} yr) old. There was no carbon, oxygen, or nitrogen in the Universe at that time. Consequently, life was not yet possible.

B. First Generation Stars and Nuclear Energy

We now want to consider the first generation of stars that formed from the clump of gas that made our Galaxy. The original clump of gas that formed our Milky Way Galaxy contained about 10^{12} solar masses ($M_\odot$). For comparison, stars now range in mass from about 0.1 to 100 $M_\odot$. While it is likely that the first generation stars were more massive than present-day stars, each represented an extremely small fraction of the total mass. Thus, we must imagine the original, galaxy-sized clump breaking into many, much smaller clumps.

Now let us consider one of the smaller clumps that is going to form a star. We will consider the process in much more detail in the next chapter; for now our focus is on the transformation of matter in the stars that formed early in the history of our Galaxy. A clump of gas contains an energy source, gravitational potential energy. It is called potential energy because it has the potential to become a more active form of energy, like the kinetic energy discussed earlier. It is called gravitational because it arises from the force of gravity, which causes any two pieces of matter to attract each other. It is this gravitational attraction that causes the clump to collapse, with all the pieces falling toward the center.

Since gravity is a force we all experience, though we may not think about it much, an example from our terrestrial experience may help to explain the concept of potential energy. Consider a reservoir holding water at the top of a hill. The Earth attracts this water, leading to the familiar observation that water runs downhill. It would do so until it reached the center of the Earth, if it were not prevented from doing so. Any water (or anything else) not at the center of the Earth has gravitational potential energy, some of which it can give up by getting closer to the center. Since the principle of **conservation of energy** tells us that energy cannot be created nor destroyed, the lost gravitational potential energy turns into other forms of energy. If we open the floodgates and let some water run out of the reservoir, it acquires kinetic energy— the energy of motion. If we put a hydroelectric generator in the path of the water, we can convert some of this energy into electric energy. As the water reaches the bottom, it will be very stirred up, with large portions moving randomly; this is called turbulent energy. The turbulent energy gradually dissipates into motion on smaller and smaller scales until it becomes heat energy, the random motion of individual particles of water. We can analyze the events of water flowing from a reservoir as a series of transformations of energy, which always conserve the total amount of energy.

Very similar transformations occur when a clump of gas collapses to form a star. Although no hydroelectric generators exist in the clump of gas, an analogous process produces electric energy. For now, we defer the details of star formation and consider only the transformation of gravitational potential energy into heat energy. As the clump collapses to make a star, it eventually heats up; the individual atoms move faster and faster in random ways. Eventually, the temperature in the center of the clump of gas reaches 10^7 K. At this temperature, which had not existed in the Universe since the first month after the Big Bang, nuclear reactions again become possible. You may note that this critical temperature, 10^7 K, is lower than the temperature at which nuclear reactions stopped during the rapid expansion of the Universe. This difference occurs because the density and temperature were dropping rapidly after the Big Bang, so reactions needed to be fast, but they are maintained at very constant values in the core of a star, allowing time for rather slow nuclear reactions. The density and temperature remain constant because the nuclear reactions release enough energy as heat so that the outward pressure of the gas balances the inward pull of gravity. The clump of gas has become a star that can maintain a nearly constant size, density, and temperature for more than a million years. During this stage of the star's life, it is referred to as a main sequence star.

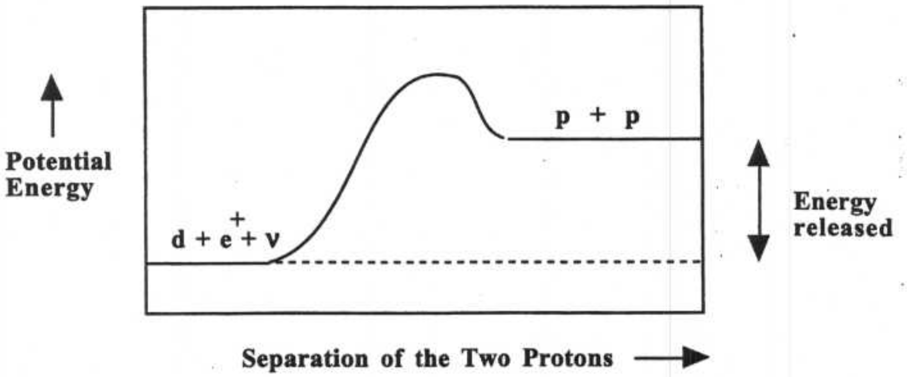
To understand how nuclear reactions produce heat, we must consider forces other than gravity. There are four basic forces of nature: gravity, the electromagnetic force, the weak nuclear force, and the strong nuclear force. For our purposes, we can consider the strong and weak nuclear forces together as the nuclear force. Each of these forces has associated with it a potential energy, analogous to gravitational potential energy. Thus, we can refer to nuclear potential energy; because nuclear forces are less familiar to us than gravity, we can use our example of gravitational potential energy to understand nuclear potential energy.

The star gets the pressure to balance gravity from the release of nuclear potential energy. The process that releases the potential energy is nuclear fusion— the combining of fundamental particles or simple nuclei into more complex nuclei. The first step of this process is the fusion of two protons to make a deuteron, a positron, and a neutrino, which is indicated in the following reaction:

$$p + p \rightarrow d + e^{+} + \nu.$$

The p stands for proton; the d is for the deuteron (a combination of a proton and a neutron); the e^{+} indicates a positive electron, or positron; and the ν represents a neutrino. The positron is an example of antimatter, being the anti-particle of the electron.

The essential point is that the nuclear potential energy of the deuteron, positron, and neutrino is less than that of the two separate protons. Thus, the reaction releases energy, just as water running downhill releases gravitational potential energy. We can make this analogy visual by drawing a figure in which potential energy increases upward and the separation of the two protons increases to the right. As the two protons come closer together, the energy changes along the curve from right to left.



As the two protons "fall" together, their separation decreases until they "fall" into the form of d, e^{+}, and ν, which, taken together, represent a state of lower potential energy. The difference is released in the form of kinetic energy, as the d, e^{+}, and ν move off at very high speeds. As they collide with other particles, they make the other particles move faster and so on until the motions are random and heat has been produced.

A slightly different way of looking at this process is provided by Einstein's most famous equation, $E = mc^2$. The energy released in nuclear fusion corresponds to a decrease in mass; the combined mass of the deuteron, positron, and neutrino is slightly less than that of the two protons. The lost mass has been converted to another form of energy.

Now we are left with the following question: if what we have said is true, why don't all protons immediately combine into deuterons, positrons, and neutrinos? To use our analogy, water runs downhill! We can also approach the answer to our dilemma through our gravitational analogy. By erecting appropriate barriers (e.g., dams), we can impound water and prevent it from running downhill. In a similar way, the protons must overcome the energy barrier drawn in our diagram. Unless the temperature is very high (at least 10^7 K), the kinetic energy of the protons will not be sufficient to overcome this barrier. They will approach and climb part way up the barrier only to fall back again. Thus, only at the high temperatures of the very early Universe and in the cores of stars are the conditions right for nuclear fusion.

What is the source of the energy barrier? It is caused by the electromagnetic force. The two protons both have positive charges, and like charges repel each other. If they collide with enough energy, they will get very close before the electromagnetic force pushes them apart. If they get close enough, the much stronger nuclear force takes over and pulls them together. This corresponds to making it to the top of the barrier. At a temperature of 10^7 K, the average kinetic energy is still not enough to overcome the barrier; only the few particles that are traveling with much more energy than the average are able to react.

Let us summarize. A collapsing clump of gas releases gravitational potential energy, some of which becomes heat— the random motion of particles. When the temperature reaches 10^7 K, nuclear fusion begins. The release of nuclear potential energy maintains the high temperature without further contraction of the star. The pressure associated with this high temperature balances the pull of gravity so that the star remains stable for a long time; during this time we call it a main-sequence star.

Finally, we should recall that these first generation stars formed with only hydrogen and helium. Consequently, they could have had no Earth-like planets, made of silicon, oxygen, iron, etc., and certainly no life could have existed around them. Nevertheless, they are important to us because they themselves produced some of the elements essential for life.

C. Heavy Element Production and Dispersal

For astronomers, the term "heavy" element refers to anything that is heavier than helium. The heavy elements were created, not in the Big Bang, but in the process of evolution of the first generation stars that formed after the Big Bang. After two protons combined to make a deuteron, a positron, and a neutrino, the next nuclear process was the following:

$$d + p \rightarrow {}^{3}\mathrm{He} + \gamma \qquad (\gamma = \text{gamma ray, an energetic photon}).$$

In ^{3}He, the 3 refers to the total number of protons and neutrons in the nucleus. (^{3}He has two protons and one neutron in its nucleus.) Then,

$${}^{3}\mathrm{He} + {}^{3}\mathrm{He} \rightarrow {}^{4}\mathrm{He} + 2p \qquad ({}^{4}\mathrm{He} = 2p + 2n).$$

This fusion process was the main source of heat and pressure in the star. However, there was still no carbon, oxygen, or nitrogen in the star as long as it was on the main sequence.

First generation stars reached the end of their main sequence lifetimes when this process had converted 10% of the hydrogen they contained into helium. Then the first generation stars moved into the red giant phase of their lives. The envelope of the star expanded and cooled, while the core of the star shrank and heated up. When the core reached a temperature of 10^8 K, the following reaction can take place, producing carbon:

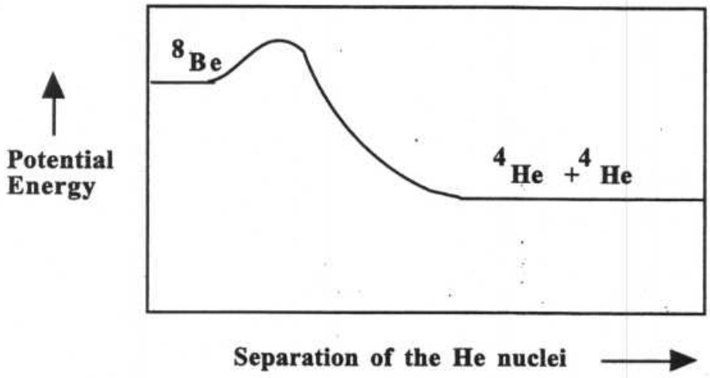
$$3\ {}^{4}\text{He} \rightarrow {}^{12}\text{C} + \gamma \qquad ({}^{12}\text{C} = 6\text{p} + 6\text{n}).$$

Actually, three ^{4}He nuclei are unlikely to collide simultaneously; instead, the following two-step process involving beryllium occurs:

$${}^{4}\text{He} + {}^{4}\text{He} \rightarrow {}^{8}\text{Be}\ ({}^{8}\text{Be} = 4\text{p} + 4\text{n}).$$

Then, ${}^{4}\text{He} + {}^{8}\text{Be} \rightarrow {}^{12}\text{C} + \gamma$.

The difficulty with this process is that ^{8}Be is unstable; it has a higher nuclear potential energy than two helium nuclei.



This is a critical step in creating all the heavy elements. If the density is high enough, another ^{4}He can come along and hit the ^{8}Be before it decays, creating ^{12}C. Once ^{12}C is available, oxygen can form:

$${}^{4}\text{He} + {}^{12}\text{C} \rightarrow {}^{16}\text{O} \qquad ({}^{16}\text{O} = 8\text{p} + 8\text{n}).$$

So, when first generation stars reached the red giant phase of their lives, C and O first appeared in the Universe (some nitrogen may have been produced in the first generation of stars, but not much). As these stars further evolved, their cores got hotter and denser, and eventually the oxygen they contained began to fuse. Then, the following reactions occurred:

$${}^{16}\text{O} + {}^{16}\text{O} \rightarrow {}^{32}\text{S} + \gamma \qquad (\text{S} = \text{sulfur})$$
$${}^{16}\text{O} + {}^{16}\text{O} \rightarrow {}^{31}\text{P} + \text{p} \qquad (\text{P} = \text{phosphorus})$$
$${}^{16}\text{O} + {}^{16}\text{O} \rightarrow {}^{28}\text{Si} + {}^{4}\text{He} \qquad (\text{Si} = \text{silicon}).$$

Eventually, iron (^{56}Fe) was produced from nuclear reactions. Iron has the lowest nuclear potential energy of any nucleus; it cannot produce energy. Since the star could not get energy by fusing iron, the core of the star collapsed, and the rest of the star exploded. This supernova explosion carried some C, O, P, S, and Si away from the star and out into the Galaxy.

Second generation stars form with some C and O in addition to H and He. In second generation stars, another energy cycle can operate, the CNO cycle, which uses C and O as a catalyst for the $H \rightarrow He$ reactions. As a by-product, some C and O are converted into N; this process created most of the nitrogen in the Universe.

Second generation stars also became red giants; they exploded, spreading nitrogen and other elements throughout the Galaxy. Thus, before life as we know it formed, at least two generations of stars exploded and spewed material into space. Eventually, this material spread throughout our Galaxy because of galactic rotation and random motion. However, you might expect to find more heavy elements near the center of the Galaxy than anywhere else in the Galaxy since stars formed earlier there. We have observed this trend, both in our galaxy and in others. Most of the heavy elements in the Sun and the Earth were produced by massive stars. These massive stars live fairly short lifetimes because they use up their nuclear fuel supply so rapidly. A massive star, for example, may live only millions of years or tens of millions of years, which is very brief compared to the age of our Galaxy. It turns out that low mass stars live much longer than high mass stars. Our own Sun is 5×10^9 years old and it is expected to live another 5×10^9 years. A massive star, however, will live only 10^6 to 10^7 years. So, the oldest one-solar-mass stars are just now dying because they are just now using up their fuel supplies; however, many generations of massive stars have already died.

Summary: The heavy elements, essential to life, were created by the early generations of massive stars. In essence, then, we can paraphrase Shakespeare, "We are such stuff as stars are made of, and our little life is rounded with a sleep." The basic forces of nature (strong nuclear force and gravitational force) led to the first evolution of matter: from protons to heavy elements.

Chapter 3

Cosmic Evolution: 2. Heavy Elements to Molecules

A. Interstellar Molecules

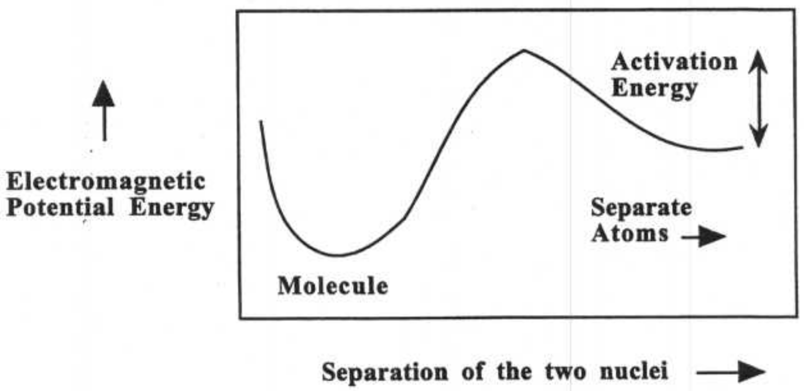
Before discussing the next step in cosmic evolution, we need to review some basic concepts. In the previous chapter, we discussed elements. Ninety-two (92) elements occur naturally in the Universe; about twenty more elements have been created in laboratories. The naturally occurring elements were almost all made in the interiors of massive stars or in the brief episode of a supernova explosion. If we take a sample of an element, say iron, and break it into smaller and smaller pieces, each of which is still iron, we would reach the limit at one atom of iron. Thus the smallest unit of an element that still has the chemical properties of that element is an atom. Atoms consist of a nucleus and electrons. In the process of creating the elements, we were mostly concerned with creating the nuclei. Deep inside stars where synthesis of nuclei takes place, temperatures are so high that electrons are stripped off in a process called ionization. A neutral atom must have the same number of electrons as positive charges (protons) in its nucleus. Otherwise, it is an ion and a superscript plus sign indicates the number of missing electrons. Thus C^{+2} would indicate a carbon atom missing two electrons. Note that the atom is held together by the electromagnetic force, since the positive charge of the nucleus attracts the negative charge of the electron. The nucleus, on the other hand, has only positive charges, which repel each other, so it must be held together by the more powerful strong nuclear force. The greater strength of this nuclear force is able to hold nuclei together, even in the very hot cores of stars.

Now we want to consider a more complex structure than an atom. If we combine two or more atoms in such a way that they are bound together without their nuclei merging, we have a molecule. A molecule is the smallest piece of a compound, just as an atom is the smallest piece of an element. If we subdivide a glass of water into smaller and smaller units, each of which is still water, we would reach the limit with a single water molecule. Any further subdivision would leave us with separate atoms of hydrogen and oxygen, the two elements that comprise water. Molecules permit an enormous jump in potential complexity. While there are only about 115 elements, there are millions of known molecules and probably an infinite number of possible molecules. Some of the key molecules of life contain billions of atoms. The next few steps in the story of cosmic evolution involve the formation of these complex molecules from atoms.

1. Chemical Reactions

To see how molecules work, let us consider the simplest molecule, H_2. This notation implies that two atoms of hydrogen have combined. Imagine taking two separate H atoms and bringing them together to make a molecule. There are four particles to consider: two protons and two electrons. The two protons will repel each other and the two electrons will also repel each other, so we have two repulsive forces. However, the electron of each atom will attract the proton of the other atom, so we also have two attractive forces. In this situation, unlike that of the atom, it is not obvious whether the attractive forces will outweigh the repulsive forces. Although molecules are also held together by the electromagnetic force, the binding is much more delicate. Consequently, not all atoms will combine into molecules; for example, helium makes no molecules, so it will not play a major role in the rest of our story, even though it is the second most common element in the Universe. Molecules can be broken apart much more easily than can atoms. This fragility gives them a flexibility that is essential for life. Life relies on chemical reactions, the breaking apart and formation of molecules.

Let us return to our example of the two hydrogen atoms. It turns out that the forces of attraction and repulsion can balance each other if the nuclei (protons in this case) have just the right separation. If the protons deviate from this optimum separation, the electrons quickly move to adjust the forces to keep the molecule from coming apart. Of course, if the protons get too far out of position, this electronic "glue" fails. The two electrons tend to congregate between the two protons; this sharing of two electrons is called a chemical bond. For molecules with many atoms, each with many electrons, keeping track of all the forces would be hopeless. Instead, we rely again on our concept of potential energy. Now we are dealing only with electromagnetic potential energy. Analogous to our analysis of the reactions of protons to form a deuteron, positron, and neutrino in stars, a molecule will be stable if it has a lower electromagnetic potential energy that the sum of its separate atoms. Also analogous to our previous discussion, there is usually a barrier to the formation of molecules from atoms. The barrier, in this case called the activation energy, exists because the first parts of the atoms to interact are the electrons, which, having like charges, will repel each other. Only if we push the atoms over this barrier will the electrons rearrange themselves to bring the forces into balance and form a molecule, as indicated in the potential energy drawing below.



While many similarities exist between chemical and nuclear reactions (Chapter 2), there are some very important differences. Because the electromagnetic force is so much weaker than the strong nuclear force, the energies involved are far lower. Thus, far less energy is released in chemical reactions, which involve only the electromagnetic force, than in nuclear reactions. This difference explains why nuclear weapons can produce so much explosive power from a small amount of matter, or why nuclear reactors can get so much more energy from a given amount of fuel than can be obtained by burning conventional fuels. It is also true that the activation energy barrier is much lower for chemical reactions than for nuclear ones. Thus, we do not need the enormous temperatures of stars to cause chemical reactions. In fact, the typical activation energy barrier can be overcome by collisions that occur at temperatures of 100's to 1000's of degrees, much lower than the $T > 10^7$ K needed for nuclear fusion. Many reactions can occur at "room temperature," which is about 300 K.

We will represent molecules in several ways. One is to write the elements involved with a subscript to indicate how many atoms of each element are needed. Examples are H_2 and CO_2; the latter is carbon dioxide, a combination of one carbon atom and two oxygen atoms. We may also indicate their chemical structures by drawing the separate atoms connected by lines; each line represents a bond, which in turn represents a sharing of two electrons. Since carbon is willing to share four electrons and oxygen two, the CO_2 molecule actually contains two "double bonds," drawn as a pair of lines, each representing a sharing of two electrons.

H_2 $\quad$ CO_2

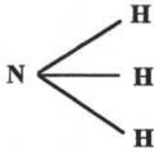
$H—H$ $\quad$ $O=C=O$

single bond $\quad$ double bonds

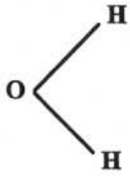
2. Origin of Interstellar Molecules

Where and when did this next evolution in the complexity of matter occur? We now know that it occurred in the space between stars, which we call interstellar space, probably after several generations of stars had created a substantial abundance of heavy elements. Although a few very simple molecules had been discovered in interstellar space in the 1930's, it was generally believed that very few of the atoms ejected from stars would form molecules. However, this belief was shattered in 1968 when some fairly complex molecules were found in interstellar space. The interstellar molecules are not liquids, which only exist in very special conditions. Instead they were in the gas phase, individual molecules floating freely in interstellar space. These gas-phase molecules can spin, end over end, emitting radio waves as they do so. Each molecule emits radio waves at particular wavelengths, mostly around a few millimeters. We identify an interstellar molecule by detecting radio waves at the precise wavelength known to be emitted by a certain kind of molecule. These radio waves are picked up by specialized radio telescopes, which look much like satellite dishes, but with much smoother and more accurately shaped surfaces. In recent years, several interstellar molecules have been discovered using infrared light. The molecules also emit or absorb infrared light when they vibrate, a process in which the nuclei jiggle back and forth around the separation with lowest potential energy.

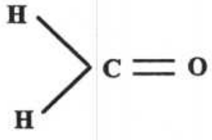
The molecules discovered in interstellar space as of this writing are listed in Appendix 2. Some of the first molecules to be discovered were ammonia (NH_3), water (H_2O), and formaldehyde (H_2CO). All of these molecules, whose structures are shown below, turn out be crucial for the origin of life. Water, in particular, is the solvent for all life on Earth and the most common molecule in all terrestrial organisms.



NH_3 (Ammonia)



H_2O (Water)



H_2CO (Formaldehyde)

Notice in the diagram that nitrogen can make three bonds, oxygen two, and, as previously noted, carbon can make four bonds, while hydrogen can make only one, since it has only one electron to share. These differences determine a great deal of chemistry. In particular the flexibility of carbon, resulting partly from its ability to make four bonds, accounts for the dominance of carbon chemistry in terrestrial life. Another interstellar molecule of note is carbon monoxide (CO), the most common interstellar molecule after H_2. Some of the biggest interstellar molecules are called cyanopolyynes, long linear chains of carbons, with a hydrogen on one end and a nitrogen on the other end. For example, $HC_{11}N$, with 11 carbon atoms, is the longest of a sequence of molecules beginning with HCN and including HC_3N, HC_5N, etc., up to $HC_{11}N$. These all contain energy-rich triple bonds, and some are important in the standard picture of the origin of life. Recently, acetic acid (vinegar) has been found. A related molecule of greater interest to many college students, at least in its terrestrial form, is CH_3CH_2OH, or ethyl alcohol. In a particular region near the center of our Galaxy, there is about one Jupiter mass worth of alcohol!

3. Three Lessons of Interstellar Molecules

What does the presence of these complex molecules in interstellar space tell us? There are three main lessons. First, we can see that molecules with as many as 13 atoms have evolved in places other than Earth. These molecules are seen in locations all over our Galaxy and in many other galaxies, extending far beyond the Earth the scale on which such complexity exists. Perhaps, still more complex molecules exist in space. It turns out that it is more difficult to detect complex molecules than to detect simple ones. Later, we will see that molecules called amino acids are important building blocks of life. For only the simplest of these, glycine, has the laboratory data been obtained to allow an interstellar search. This search has found tentative evidence for interstellar glycine. The abundances of the cyanopolyynes decrease with increasing number of carbons; nonetheless, we suspect that still more complex forms are present. In fact, there is increasing evidence for a class of very large carbon-based molecules, called polycyclic aromatic hydrocarbons (PAH's, for short), which are linked rings of carbon atoms, with hydrogens only on the outside of these rings. These compounds, also found in the exhaust of automobiles, may also play a role in the origin of life. Their detection would have been impossible with radio techniques; instead they were detected by the vibrations of the atoms in the molecule, which result in the emission of infrared waves. In this case, we are able to identify only the class of molecules rather than a specific form of PAH, and even this general identification is somewhat uncertain.

Second, the dominance of carbon in interstellar chemistry can be appreciated by examining the list of interstellar molecules (see Appendix 2). The dominance of carbon becomes increasingly evident as the complexity of the molecules increases. Carbon dominance encourages us that our concentration on the carbon chemistry of terrestrial life is not just a case of Earth chauvinism. It really seems that carbon's versatility makes it a likely basis for life elsewhere. The dominance of carbon is perhaps most dramatic in the long chains (cyanopolyynes) and rings (PAH's).

Third, a study of the formation and destruction of interstellar molecules will illustrate some of the problems in understanding chemical evolution on the early Earth. In both cases, we have to understand how simple molecules are driven to form more complex molecules and how these complex molecules are protected from destruction by ultraviolet light. Let us consider destruction first. The main reason that no one expected to find complex interstellar molecules is that ultraviolet light can easily break chemical bonds and destroy molecules. Massive stars produce a lot of ultraviolet light, and people believed that few molecules could survive. Molecules are indeed rare except in particular locations, called molecular clouds, where there is a sufficient concentration of molecules and dust to block out the ultraviolet light, just as the ozone layer does for the Earth.

4. Dust and Molecules

What are these dust particles? As the name implies, they are small, solid particles, smaller than a micrometer (10^{-6} m), but with a wide range of sizes. Very large PAH's may be indistinguishable from the smallest dust particles. Studies of the effectiveness of these particles in absorbing both ultraviolet and infrared light indicate at least two distinct types of dust particles: one is primarily carbon and may be the extension of the PAH's to larger structures, like graphite; the other is composed of silicates, minerals of silicon and oxygen. Silicate particles are similar to what we would call dust on Earth, while the carbon particles are probably more similar to what we would call smoke or soot. Both are produced in the material flowing outward from old stars, but the different histories of the parent stars determine which kind of dust they produce. Supernovae may produce yet other kinds of dust. All these different sources of dust are mixed in interstellar space, and when they become relatively concentrated into a "cloud," they protect molecules from ultraviolet light.

We still must understand the formation of the molecules. Dust also plays a role here. Consider the simplest and most common interstellar molecule, H_2, which we discussed above in terms of two hydrogen atoms coming together to form a molecule. We argued that H_2 would tend to form because it has a lower potential energy than the sum of the energies of the two separate atoms, releasing energy as it forms. However, the newly formed molecule must get rid of this released energy or it will break apart again. For reactions that produce more than one particle, the energy can be released as the kinetic energy of the particles moving away from each other. Since only one particle is produced in the reaction ($H + H = H_2$), the H_2 must get rid of its energy either by colliding with another particle or by emitting a photon. In interstellar space, the density of particles is so low that the H_2 will fall apart before it would run into another particle, leaving photon emission as the only possibility. However, H_2 is very slow to emit a photon, a consequence of the fact that it is composed of two identical atoms. The net result is that H_2 cannot actually form in space by the collision of two gas-phase atoms. Instead, the H_2 forms on the surface of dust grains. This process, called surface catalysis, is similar to what catalytic converters do in car exhaust systems. Catalysis allows a reaction to occur much more rapidly or easily than it otherwise would. The formation of H_2 occurs when two H atoms, both of which have struck a dust particle, meet and combine. The energy released can be used both to eject the newly formed molecule from the surface and to add some energy to the vibration of the atoms in the dust particle.

The formation of more complex molecules from H_2 involves a different kind of problem. The temperatures of most parts of molecular clouds are so low that the typical collision lacks the energy needed to get over the activation energy barrier. Therefore, interstellar chemistry must proceed by reactions that lack activation energy barriers. Suppose one of the molecules we want to react is missing an electron; this type of molecule is called a molecular ion, indicated by a + superscript at the end of its chemical formula (e.g., HCO^+, one of the molecules in the table in Appendix 2). Since the activation energy barrier is caused by the initial repulsion of the electrons of the two original atoms or molecules, it is not present if one of them has a net positive charge. Instead, that positive charge will actually attract the other particle, and the reaction can proceed rapidly.

How do molecules become ionized? If they are near the surface of the cloud, they may be ionized by ultraviolet light, but molecules formed there will be short-lived. Deeper in the cloud, where ultraviolet light is very weak, molecules are occasionally ionized by cosmic rays, fast moving nuclei, usually protons, that are probably created in supernovae explosions. Cosmic rays are rare, but when one passes by, it can ionize many atoms or molecules. A molecular ion can then react quickly with other atoms or molecules, building up a more complex molecular ion, until it combines with an electron to form a neutral molecule. The sequence of reactions results in a more complex neutral molecule than we began with. As an example, consider what happens to a molecule of H_2, ionized by a cosmic ray:

$$H_2^+ + H_2 \rightarrow H_3^+ + H$$
$$H_3^+ + CO \rightarrow HCO^+ + H_2$$

The HCO^+ can then undergo further reactions to build up still more complex molecules. The detection of HCO^+ in interstellar space proves that these ion-molecule reactions are occurring. While the details will differ, we will encounter this pattern again when we consider the chemistry of the early Earth, where an input of energy may have driven simple molecules to become reactive (like the molecular ions) and form more complex molecules.

For some time, we thought that this was the whole story: H_2 formed on dust grains and the other molecules formed in ion-molecule reactions. However, detailed calculations were unable to explain very large molecules in this way. Furthermore, if H atoms can run into dust particles and stick, even for a little while, the more complex molecules should also collide with dust particles and stick much more firmly. In this case, the gas phase molecules, which were detected by observing the radio waves emitted when they spin, would be steadily depleted as they stick to dust particles. Once on the surface, they could no longer spin, but their atoms could still vibrate, emitting or absorbing infrared light. Infrared observations do indicate that some molecules are stuck on dust particles, in the densest parts of molecular clouds. For example, water molecules stuck on dust particles absorb infrared light at a wavelength of 3×10^{-6} m; the detection of absorption at this wavelength showed that the dust particles in some regions have a mantle of solid water (in other words, ice). More recently, characteristic infrared absorption features indicated that ammonia (NH_3), methane (CH_4), and methanol (CH_3OH) are also frozen on the dust. Dust grains appear to have icy mixtures on their surfaces, containing all the basic four bioelements, H, O, N, and C.

Could further reactions occur among these ices? Some calculations indicate that they should, and there is some evidence from infrared observations of other, more complex, molecules with bonds between carbon and hydrogen, as well as a molecule with a triple bond between carbon and nitrogen. Mayo Greenberg, working with a group in The Netherlands, studied what happens to ices in the laboratory when they are exposed to ultraviolet light, heating, etc., as may occur occasionally in interstellar space. Considerable complex chemistry can occur, including the production of a complex component he referred to as "yellow stuff." This stuff has some similarities to products of experiments designed to study the origin of life. Whether these same reactions occur in interstellar space is somewhat controversial, but the possibilities are intriguing.

Fred Hoyle and Chandra Wickramasinghe have taken these ideas still farther and suggested that life actually originated on these dust particles. Indeed they later argued that new life occasionally reaches the Earth even today, in the form of viruses from cometary debris, occasioning the outbreaks of new epidemics. These ideas are so far-fetched they have found little support among scientists. Other workers are however taking serious note of the complexity that seems to exist in interstellar space and considering its possible role in the origin of life on Earth. As we shall see, new stars and, apparently, planets form in the parts of molecular clouds where complex molecules exist, so a role for interstellar molecules may not be so far-fetched. We will return to this topic when we discuss the origin of life.

B. Molecular Clouds

The places where interstellar molecules are found are called molecular clouds, analogous to clouds in the Earth's atmosphere. They consist primarily of molecular hydrogen (~93%) and atomic helium (~6%). The remaining 1% (actually a bit less) consists of the heavy elements, mostly in the form of molecules or dust particles. The most common molecule after H_2 is carbon monoxide (CO). Because H_2 is so poor at emitting photons, most of it is not directly detectable, leaving CO as the main tracer, telling us where molecular clouds exist. CO also works like a thermometer, with the intensity of the radio waves emitted by CO molecules telling us the temperature in the cloud. Typical temperatures determined from CO are about 10 degrees above absolute zero (10 K).

Using other molecules, like H_2CO, HCN, or CS, we can deduce the density in the clouds. We have been using the concept of density without a careful definition, but now we want to be more precise. By density, we mean the number of particles (in this case, molecules) in a standard volume. In the units we are using, this standard volume is a cubic centimeter, written cm^3, to indicate that it is a cube, one cm on a side, about the size of a sugar cube. Imagine counting all the molecules in a box this size. If the box were filled with air, we would have to count about 10^{19} molecules! If we filled the box with water, we would count still more, about 10^{24}. We express this by saying the density, denoted by n; is 10^{24} cm^{-3}, the superscript of -3 on cm indicating that we are referring to the number *per* cubic centimeter. Compared to these numbers, interstellar space has a very low density; an average value would be about 1 cm^{-3}. That's right, ONE molecule in our sugar cube sized box! The parts of space between molecular clouds have even lower densities, while the densities of molecular clouds begin at about 10^2 cm^{-3} and are still higher in places. Note however, that even the densest known part of a molecular cloud is much less dense than air, or even the best vacuum on Earth. Dense is clearly a relative term!

The average density in a star like the Sun is similar to that of water, $n \sim 10^{24}$ cm^{-3}, many orders of magnitude denser than molecular clouds. Yet, we will see that new stars form out of molecular clouds. How can such rarefied clouds make the much denser stars? First of all, the clouds are much bigger than stars, so even a low density adds up in the large volume to a large mass. Sizes of molecular clouds range from about 1 ly (remember that this stands for a light year) up to about 300 ly or more. Consequently, they contain a lot of mass, from 1 $M_\odot$ to 10^6 $M_\odot$. In comparison, stars range in mass from 0.1 $M_\odot$ to about 100 $M_\odot$, so the largest molecular clouds contain enough mass to make many stars, and even the smallest clouds could at least make the lower mass stars. Indeed, some small clouds appear to be making low-mass stars, while most large clouds are making many stars with a wide range of masses. Before discussing the process of star formation in greater detail, we will make a rough estimate of R_*, the first factor in the Drake equation.

C. Star Formation

1. Estimate of R_*

To calculate the star formation rate, we could study current star formation, but our studies are too preliminary to calculate an accurate rate in that way. Instead, we calculate R_* by dividing the total number of stars in the Galaxy by the time that the Galaxy has had to form them. No one has counted all the stars in the Galaxy; besides the fact that this task would be very tedious, interstellar dust blocks our view of most of them. Instead, we use basic laws of physics to compute how much **mass** lies within the Sun's orbit, from knowing the distance to the center and the velocity of the Sun. The physics of orbital motion tells us that the kinetic energy of an orbiting body is half its gravitational potential energy. If we apply this rule to the case of the Sun in orbit around the center of the Galaxy, we get the following equation:

$(1/2)\ M_\odot\ Mg/Rg = (1/2) M_\odot\ v^2$,

where $M_\odot$ is the mass of the Sun, M_g is the mass of the Galaxy inside the Sun's orbit, R_g is the distance of the Sun from the center of the Galaxy, v is the velocity of the Sun, and G is the universal gravitation constant. By canceling factors on both sides and doing a little algebra, we can get an equation for M_g:

$M_g = R_g v^2/G$

The Sun orbits the center of the Galaxy at a velocity of about 250 km s^{-1}, and the current best estimate of the distance to the center is about 25,000 ly. After we put all these numbers in consistent units, they imply that the mass inside the Sun's orbit is about $1.0 \times 10^{11}\ M_\odot$. Some of this mass is interstellar matter, not stars. On the other hand, there are some stars outside the Sun's orbit, so a reasonable estimate for the mass of the stars in our Galaxy is about $1.6 \times 10^{11}\ M_\odot$. To get the **number** of stars, we divide this mass by the mass of an average star. Because there are many more low-mass stars than high-mass stars, the average mass is about 0.4 $M_\odot$. Dividing $1.6 \times 10^{11}\ M_\odot$ by 0.4 $M_\odot$ gives us an estimate for the total number of stars in our Galaxy of 4×10^{11} (400 billion).

To get the star formation **rate**, we need to divide the number of stars by the lifetime of the Galaxy. From studies of old stars, we estimate that the age of the Galaxy is close to 10 billion years (10^{10} yr). From these numbers, we would get

$$R_* = 4 \times 10^{11} \text{ stars}/\ 10^{10} \text{ years} = 40 \text{ stars/year}$$

There are two assumptions hidden in this calculation. One is that we have accounted for all the stars ever formed. Since we know that some stars have died in the process of creating the heavy elements, this assumption is incorrect. Correction for stars that have died would increase our estimate of R_*, but the correction would be small because the stars that have died were relatively massive, and there are few massive stars, compared to low mass stars. The second assumption is that star formation has occurred at a steady rate. Some models of our Galaxy's history indicate that star formation should have been much more rapid at early times, when there was more gas available for star formation. In this case, the R_* calculated above would be an average rate, but the actual rate would have varied. Lately, studies have tended to indicate a star formation rate that does not decline much with time, while other data have even suggested that star formation occurs in bursts. Considering all the unknown factors, any estimate of R_* between about 5 and 50 stars per year is reasonable.

2. Aspects of Current Star Formation

We discussed the process of star formation briefly in discussing the formation of the first generation stars (Chapter 2). Recall the basic process: gravity causes a mass of gas to collapse, increasing the density and temperature until nuclear reactions begin. We will now consider this process in more detail and note some likely differences between current star formation and the formation of first generation stars. Remember that first generation stars formed from a gas containing only hydrogen and helium, whereas current star formation occurs in molecular clouds with many molecules and dust particles. Most kinds of molecules are good at emitting radio and infrared light, as are the dust particles. Because this light escapes the cloud, carrying away energy, molecular clouds are very cold, as previously noted. The clouds that formed the first generation stars would have been warmer. The mass of gas required for gravity to cause collapse depends on the temperature. Warmer clouds would require a larger mass to collapse, thus forming more massive stars. These considerations may explain the fact that we do not observe any first generation stars. Formed from warmer clouds, first generation stars were fairly massive. Since massive stars have lifetimes much less than the age of our Galaxy, they would all have died by now.

How do we know that stars form in molecular clouds? First, we observe young stars in proximity to molecular clouds. Stars still forming in the clouds are usually hidden from view because the dust in the cloud scatters and absorbs their visible light. However, infrared light is more effective in penetrating the dust, and recent advances in infrared technology have allowed us to detect many stars buried deep in the clouds. These studies have shown that stars often form in large clusters of hundreds of stars, located in relatively dense parts of clouds ($n > 10^4$ cm^{-3}), rather than forming throughout the cloud. Theoretically, collapse should occur faster at higher densities, so this observation makes some sense. Since most star formation appears to occur in clusters in large molecular clouds, it is most likely that our Sun formed in such an environment. On the other hand, some stars, particularly low-mass stars, form in isolation in a small, dense region of a larger cloud, or in a small cloud, and these more isolated stars are easier to study in detail. Note that we generally think of stars above about 8 $M_\odot$ as "massive," while those below 8 $M_\odot$ are thought of as "low-mass" stars; with this criterion, our Sun is a "low-mass" star.

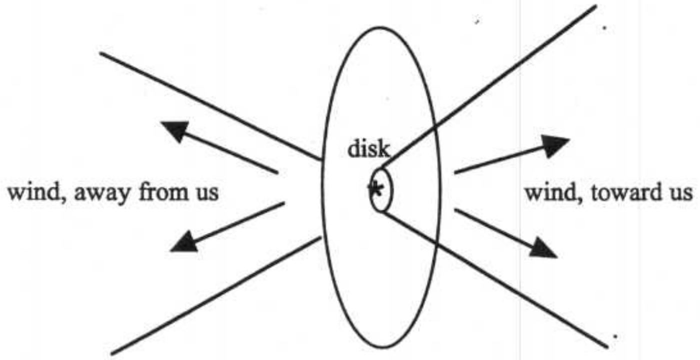
Frank Shu and his colleagues have developed a detailed theory for the formation of isolated low-mass stars. In this theory, the early stages of the process are quite gradual, with the gas slowly becoming denser. A true collapse is prevented by a combination of pressure and magnetic fields. The density slowly builds up in the center of what will become the star, until finally gravity overwhelms the forces resisting the collapse and the gas begins to collapse. The collapse begins close to the center and only later do the outer parts begin to fall in, earning this theory the name, "inside-out collapse." Observations of a few small molecular clouds have found very good agreement with the detailed predictions of the theory, but other observations suggest a collapse scenario differing in the details.

As the collapse proceeds, emission of radio and infrared photons by the molecules and dust keep the temperature low until the density is about 10^{11} cm^{-3}. At that density, the photons begin to be absorbed before escaping and the temperature begins to rise. The resulting rise in pressure slows the collapse of the inner parts, which then form a slowly contracting core, while the outer parts continue to fall in on this core. The core accumulates matter from the infall, while shrinking in size and getting hotter and denser. At some time, nuclear reactions begin, and a main-sequence star begins its life. This process takes quite a while in low-mass stars, so there is a long period in which we have a protostar, an object that is not yet burning nuclear fuel. For stars more massive than about 8 $M_\odot$, the protostar phase is shorter than the time it takes all the gas to fall in, so we expect massive protostars to be hidden, except to infrared techniques, and this expectation is generally borne out by observations.

3. Winds from Young Stars

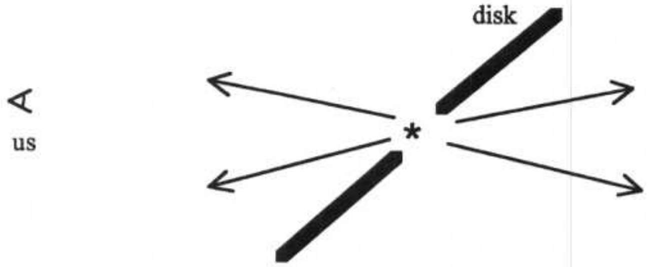
Low-mass protostars can sometimes be seen with visible light; they have been known for many years by the name T Tauri stars, after the name of the first one to be studied. These stars have long been known to be ejecting matter, called a wind, and radio studies of their more deeply buried younger brothers indicate that their winds are even stronger. How do stars form by ejecting matter? What is going on? It is suspected that the winds are the protostars' solution to a problem caused by rotation of the original cloud or portion of the cloud that collapsed to form them. If this original cloud was rotating, even a tiny bit, the rotation will be amplified in the collapse. The rotation is amplified because of a general principle, called **conservation of angular momentum**. This principle implies that a rotating object that contracts will spin faster. A familiar example is the figure skater in a spin, who can spin faster by pulling her arms in. Since the original cloud is many orders of magnitude larger than the final star, an initial small rotation will also be amplified by many orders of magnitude during the collapse. The resulting rapid rotation would prevent further collapse unless angular momentum can be carried away. The winds can carry off large amounts of angular momentum.

The winds from the young, deeply buried stars have swept up the surrounding gas, accelerating it to substantial velocities, so that we observe a Doppler shift of this gas with respect to the rest of the cloud. The first such object to be studied was a source called L1551 IRS5, meaning the 5th infrared source (IRS) in the molecular cloud L1551 (the 1551st entry in a list of clouds compiled by Beverly Lynds). The swept-up gas was clearly distributed in two lobes, one coming towards us and one, on the opposite side of the infrared source, going away from us. The discoverers of this gas suggested that this "bipolar" appearance was caused by the wind being channeled by a *disk* of material around the star.



4. Disks around Young Stars

Recently, other evidence has accumulated for disks around young stars. Studies of dust emission in the infrared and at wavelengths near 1 mm both indicated that a great deal of dust exists around T Tauri stars. If the dust were distributed spherically around the stars, it would absorb so much of their visible light that we would not be able to see them. Since we do see them, it seems that the dust must lie in a disk; then, as long as the disk is somewhat tilted so that we do not have to look through the entire disk to the star, we could still see the star. Observations of a number of young stars using arrays of radio telescopes have shown that the emission near 1 mm arises from a region very close to the star, about the size expected for a disk. In some cases, disk shape has been seen directly with these techniques.



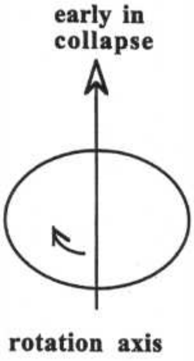
Also, the Hubble Space Telescope has seen disks of dust around some young stars. Disks are clearly quite common around young low-mass stars. There is even evidence for disks around stars forming in clusters, and around some massive stars, suggesting that disks may accompany most forming stars. As we shall see, these disks provide a natural place for planets to form. Studies of the infrared emission indicate that the dust in the disks lasts for a few million years. It may have been ejected from the system, or it may have fallen into the forming star. A more exciting possibility is that it has formed planets. The infrared emission from planets is too faint for current techniques to detect. In any case, it seems that planet formation must be quite far advanced within a few million years after the beginning of star formation if it is to succeed.

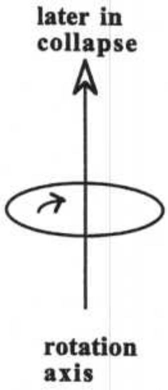
D. Planet Formation

1. Collapse with Rotation

Motivated partly by the growing evidence for disks and partly by the recognition that conservation of angular momentum will play a role in collapse, Shu and collaborators modified their inside-out collapse model to include rotation. In these models, the rotation causes an initially spherical cloud (or part of a cloud) to become flattened. Flattening occurs because matter collapsing along the rotation axis can fall straight in, but matter collapsing perpendicular to the rotation axis is slowed by centrifugal forces. These forces become greater as the material moves in because the conservation of angular momentum causes the matter to spin faster as it approaches the center of the collapsing region.

The flattening becomes greater the closer the matter gets to the star, until a true disk is formed. In this picture, most of the collapsing matter actually falls onto the disk, which then slowly feeds matter to the protostar, while retaining most of the angular momentum not carried away by the wind. To tie up all the loose ends, we would like to understand what causes the wind. We suspect that it originates from the rotation together with magnetic fields, but several models for the detailed mechanism are competing at the moment.





With sufficient rotation, we may expect the collapsing cloud to break into pieces and form a pair of stars (a binary star system). It was unclear whether a binary star would have disks as well. In fact, many young binary stars do have disks, either around each star, or around the pair of stars. In some cases, both kinds of disks exist in binary systems. However, the companion star will clear out any disk material that lies near its orbital path around the other star. Thus, it now seems likely that binary stars may form planets. However, the habitability of these planets may be affected by binarity (section F).

We shall see in section E that many facts about our own solar system can be understood in models where the planets formed from a rotating disk. The success of these models indicates that disks are likely locations for the formation of planets. Consequently, the ubiquity of disks suggests that planetary systems are common.

2. Evidence Regarding f_p

At this point, we are ready to consider the second factor in the Drake equation, f_p. The apparent ubiquity of disks, together with the evidence that our planetary system formed from a disk (section E below), suggested that planets are common, but one had to assume that planets actually formed in the disks. In 1995 and 1996, planets around other stars were detected convincingly for the first time, as we shall discuss below. By summer of 2002, more than 100 stars were known to have planets and 11 were known to have multiple planets. The numbers are growing rapidly as more searches yield results. After describing some of the techniques for finding planets, we will return to the question of the value of f_p.

So far, there is no **direct** detection of a planet around another star. This fact becomes less surprising if we consider the difficulties of detecting planets around other stars. Planets in our solar system are visible because they reflect light from the Sun. Since they are far from the Sun, the light from the planet is very much less than the light from the Sun. Nonetheless, we could still see planets around nearby stars except that their light is simply swamped by the light from the star. Because the distance from us to the star is so much greater than the distance from the star to its planet, the angle between the star and the planet is small. Even with the Hubble Space telescope, planets cannot be seen against the glare of their stars. Searches at longer wavelengths could be more successful. Since planets are much cooler than stars, they will emit most of their light at infrared wavelengths; searches at these wavelengths would be less troubled by emission from the star, but current techniques are not capable of detecting planets in the infrared either. Plans are being made, both in the U. S. and Europe, to launch an advanced space telescope capable of directly detecting planets around nearby stars and even studying their atmospheres for chemical signatures of life. Such a telescope is very challenging and launch before about 2015 is unlikely.

While the future may hold the possibility of direct detection of planets around other stars, current efforts are focused on indirect methods. The idea behind most searches is that a planet's gravitational pull will cause the star to wobble slightly as the planet orbits. While we usually think of planets orbiting around the Sun, it is more correct to think of both the Sun and the planets orbiting around the center of mass of the solar system. In our solar system, Jupiter's gravitational effects dominate those of the other planets, so effectively the Sun and Jupiter orbit around their center of gravity.

If other planetary systems are similar, we could attempt to detect this wobble. If the planet's orbit lies in the plane of the sky, the star's position could be observed to change relative to more distant stars. Since the accurate measurement of stellar positions is called astrometry, we will call this the astrometric technique. This technique has been applied to some nearby stars, and some researchers have claimed to detect planets, but their results could not be confirmed by other studies. The expected wobble is so small that it stretches the limits of what can be observed. The astrometric technique will be used by a space mission called GAIA, to be launched by the European Space Agency around 2010; it should be able to detect Jupiter-mass planets around stars out to about 600 ly.

If the planet does not orbit in the plane of the sky, the star will also wobble back and forth toward and away from us. We could never measure the distance to the star accurately enough to detect changes in its distance, but we can measure velocities very accurately. The Doppler effect associated with the star's motion toward or away from us will cause a narrow absorption line in the star's atmosphere to shift back and forth in wavelength. If these shifts fit a regular pattern, they would indicate the presence of an orbiting object. This spectroscopic technique is the one used in most of the current searches. Like the astrometric technique, the spectroscopic search for planets is pushing the limits of what is possible. The expected velocities are only a few meters per second, similar to that of human motions, whereas stars can have bubbling motions in their atmospheres as fast as 1000 meters per second. In this case, only a repeated pattern, with a period characteristic of an orbiting planet, would be convincing evidence of a planet. Consequently, detection of a planet with an orbit like that of Jupiter with either of these methods requires at least a decade of data. Surprisingly, the spectroscopic searches have turned up many planets; the detected planets are mostly large planets in orbits that are much closer to their stars than that of Jupiter. The large masses give a detectable Doppler shift and the close orbits mean that many orbits occur in a short time, rather than the decades that were expected. We shall discuss the significance of the detections at the end of this section.

The unexpected success of the spectroscopic technique has spurred scientists to plan or begin searches with other techniques. One idea is to look for planets that orbit in such a way that they pass between the star and us. As the planet **transits** the star, the star's light is dimmed by a tiny amount. While the odds are against such an orbit, persistent monitoring of many stars could find the small percentage (about 0.5%) that are lined up correctly. A space mission is planned by NASA for launch around 2006 to monitor 100,000 stars for four years; it should be able to detect even Earth-like planets. One such transiting planet is already known from ground-based studies; the Hubble space telescope observed a transit in 2001 that even allowed some study of gas in the planet's atmosphere absorbing the star light.

An even more exotic effect may be seen in monitoring studies. If a star with its planet passes directly in front of a background, more distant star, the light from the background star can be amplified by gravitational lensing. In essence, the gravity of the foreground star bends the path of light, and the star can act as a lens. Albert Einstein predicted this effect, and it has been observed. If the star has a planet, the planet can also act as a gravitational lens. The effect will be brightening of the background star with a characteristic variation with time that allows this effect to be distinguished from other causes of light variations. Several surveys for this effect are underway. As with transits, the chance of a good alignment is small, so many stars must be monitored.

Both astrometric and spectroscopic techniques have been used for years to search for companion stars. These studies have shown that about 2/3 of all stars have stellar companions. Until recently, it was believed that binary stars would not have planetary systems, limiting f_p to less than 1/3, but the evidence of both companions and disks in young stars undermines this argument.

As the sensitivity of the spectroscopic technique improved, many astronomers expected to find an increasing number of smaller mass companions, since there are more low-mass stars than high-mass stars. However, there seemed to be a distinct cut-off in stellar companions around .07 $M_\odot$, which is also the lowest mass object that can begin nuclear reactions. Objects less massive than 0.07 $M_\odot$, but more massive than about 0.01 $M_\odot$, or 10 times Jupiter's mass, are called brown dwarfs. The names reflect the fact that these objects will be dim because they lack nuclear reactions. They emit only the energy released by their slow gravitational contraction, so they gradually cool and emit less and less energy. Most of their energy is emitted at infrared wavelengths and searches at these wavelengths confirm that few brown dwarfs exist as companions to other stars. Most theories of star formation predicted that objects as low in mass as brown dwarfs could form, so it was surprising that so few likely brown dwarf companions were found. This lack of objects between 0.07 $M_\odot$ and 13 Jupiter masses was dubbed the "brown dwarf desert." The limit of 13 Jupiter masses is a crude divider between planets and brown dwarfs, but it is somewhat controversial. The brown dwarf desert suggests that planets and stars form by very different processes and intermediate mass objects rarely form in disks around stars.

The scarcity of brown dwarfs as companions to stars led astronomers to believe that brown dwarfs themselves were very rare. In another surprise, many **free-floating** brown dwarfs were found, beginning in the late 1990s and some of these appear to have disks! Could brown dwarfs have their own planets? At this point, no one knows. Since brown dwarfs are now known to be quite common, this possibility could increase the number of planets, though the habitability of these planets is unclear.

The current situation is extremely interesting. First, there is growing evidence that disks are common. If these disks form planets, we might predict a value of f_p close to 1/3 (if binary stars are excluded) or even 1 (if binary stars can also have planets). Second, brown dwarfs are now known to be common, but not in orbit around stars. The distinction between planets and brown dwarfs is not entirely clear, with different astronomers supporting different definitions. There may even be free-floating planets, though none are clearly known, as of 2002. Third, many planets have now been detected. Let us review briefly the history of planet detections, noting the nature of the planets, before we use the results to help us estimate f_p.

In 1995, two astronomers in Switzerland announced detection of a planet around a star called 51 Pegasus, and other astronomers soon confirmed it. This planet is similar in mass to Jupiter, but it orbits very close (0.05 AU) to its star. As we shall see below, this location for a massive planet was totally unexpected, given the properties of our solar system. Consequently, it was a little unclear what this detection meant for the larger question of f_p. Then in early 1996, two more giant planets were detected, one around a star called 70 Virginis and another around a star called 47 Ursa Majoris. The one around 70 Virginis has a mass of about 8 Jupiter masses and an orbit around 0.4 AU from its star. Subsequently, a second planet was discovered around 47 Ursa Majoris, one with a mass of 2.4 M_J and one with a mass of 0.56 M_J, where M_J is one Jupiter mass. As of summer 2002, we know of over 100 planets and eleven systems of multiple planets. The least massive planet around a normal star has a mass of 0.16 M_J and the nearest star known to have a planet is Gliese 876, at a distance of 15 ly. This star is much less massive than most of the stars known to have planets, demonstrating that stars as small as 1/3 the mass of the Sun can have planets. For most of the planet detections, the mass value is a minimum value, based on assuming that the planet's orbit is aligned to give the maximum Doppler shift. If the orbit is tilted, the masses are actually higher. In about four cases, the orbital plane is known and the masses are certain.

The planets found so far are rather different from what we see in our solar system. First, most are large, similar to Jupiter, and orbit closer to their star than Jupiter does. Second, many have quite eccentric orbits, which means that their distance from the star changes quite a lot during the orbit. As we shall see in the next section, the orbits of most planets in our solar system are nearly circular (low eccentricity). The planets detected so far are probably only the tip of the iceberg. Massive planets close to the parent star are the easiest to detect with the indirect techniques, so we are now seeing only the most obvious ones. As the searches extend over longer periods, they should find planets at larger distances from their stars. However, the large eccentricities are not so easily explained; it may be that the very circular orbits in our solar system are not the most common situation. In 2002, planet searchers announced several planets with masses and orbits similar to that of Jupiter. These discoveries probably herald the detection of many more planets that are less massive and farther from their stars.

Before leaving the subject, we should note that planets have also been detected around a pulsar. Pulsars are remnants of stars that have passed through a red giant phase. During this phase, planets in the orbits that are deduced would have spiraled in to the star and been destroyed. If planets now exist around the pulsars, they must have formed after the star had ceased to be a main-sequence star, but then the star would not be suitable for life to exist on its planets. Thus, planets around pulsars are unlikely to be directly relevant to extraterrestrial life, but if planets can form around a pulsar, it seems even more likely that they could form around young stars.

Now let us try to put together all these facts to make an informed guess about the value of f_p or at least set a limit to it. Since the searches so far can only find fairly massive planets orbiting close to their stars, they cannot set an upper limit; there could be small planets around all the stars studied so far that would not be detected. Because a large fraction of young stars are surrounded by disks, we could have a value as large as 1 for f_p. What about a lower limit? We could say that the fraction of stars studied for planets that actually have planets would be a lower limit. This seemingly simple calculation is complicated by another property of the stars that have turned out to have planets. On average, they have more heavy elements than do stars without planets. For stars with heavy element abundances similar to or greater than that of the Sun, the fraction with planets is 0.03 to 0.04. The fraction drops to 0.01 for stars with abundances about half those of the Sun.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

This manuscript was supplied camera-ready by the author.

Please visit our web site at www.pearsoncustom.com

ISBN 0–536–72314–1

BA 996477

PEARSON CUSTOM PUBLISHING
75 Arlington Street, Suite 300, Boston, MA 02116
A Pearson Education Company