Chapter 6:

Inconstant Stars

Juliet: O! swear not by the moon, the inconstant moon,
That monthly changes in her circled orb,
Lest that thy love prove likewise variable.

Romeo and Juliet, Act II. Scene II, William Shakespeare (1564–1616), *The Oxford Shakespeare*, 1914

Galileo and Bruno contended that the heavens change with time just like the base Earth and the inconstant Moon. The appearance of comets and "new stars" were two examples. In the *Dialogue*, Galileo forcefully stated that Aristotle did not have this new information when he postulated that the heavens never change. Halley confirmed that the heavens do change when he detected the proper motion of stars (Figure 1).

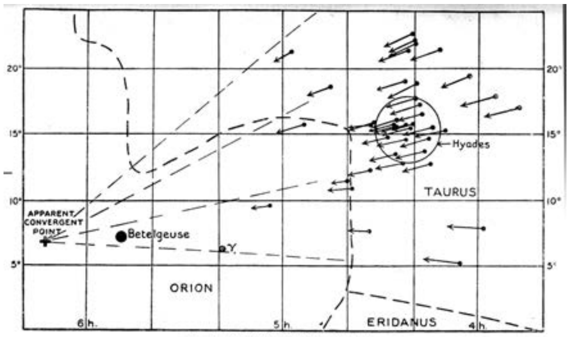

Fig. 4. TAURUS MOVING CLUSTER

Figure 1: Studies of the proper motions of stars revealed that star clusters are real groups. The stars in the open Taurus cluster move in essentially parallel paths relative to the Sun. Astronomers see their proper motion again the background of distant stars in 2-D, like railroad tracks on a painting. The convergence point of the stars corresponds to the vanishing point of parallel lines on a painting. The Sun was once part of an open cluster that has dispersed over the last 4.5 billion years. From Doig's 1927 book *Stellar Astronomy*.

The Physics of Stars

Finding that stars move around proved to be a lot simpler than figuring out how they shine and how they change over cosmic time. Bruno contended that the stars exchanged light with each other. This perpetual motion did not make much sense then and became nonsense once conservation of energy became understood. Changeable stars became the center of the science of astrophysics in the middle 1900s. To cut to the key results, the finite lifetimes of stars limit the duration of habitability of their planets. In turn, new stars and their planets form from dregs expelled from dying stars.

Primer on spectra and nuclear physics. Astronomers needed to figure out of what stars are made to figure out how they work. Quantum mechanics provided insight on both these topics. This highly complex field of mathematical physics has qualitative straightforward implications to astronomy. I use some technical terms for those already familiar with physics and those wishing to access more information.

We observe light arriving from stars. This light is a form of electromagnetic radiation. You are familiar with some of its other forms, probably radio waves and X-rays. Chapter 5 discussed gamma rays from nuclear decay. All electromagnetic radiation, including visible light, has a wavelength, analogous to the waves seen on the surface of a lake. A prism separates visible light by wavelength into a rainbow. The red end has a longer wavelength than the violet end.

You can tell the relative brightness of sunlight at the visible wavelengths by projecting the rainbow on a white sheet of paper. This measurement is a simple example of a spectrum (plural, spectra). Astronomers once used film to record spectra of the Sun and the other stars. They now use sensitive equipment similar to the innards of digital cameras. This equipment takes advantage of quantum mechanics by capturing individual photons (see below) and turning their energy into electricity.

I have already discussed the energy levels of electrons around atoms and molecules in Chapter 4 and the use of spectra to measure cosmic velocities in Chapter 5. The electrons around an atom can only have certain energies. I consider light arriving at an atom to show how these energy levels of produce observable features in the spectra of stars.

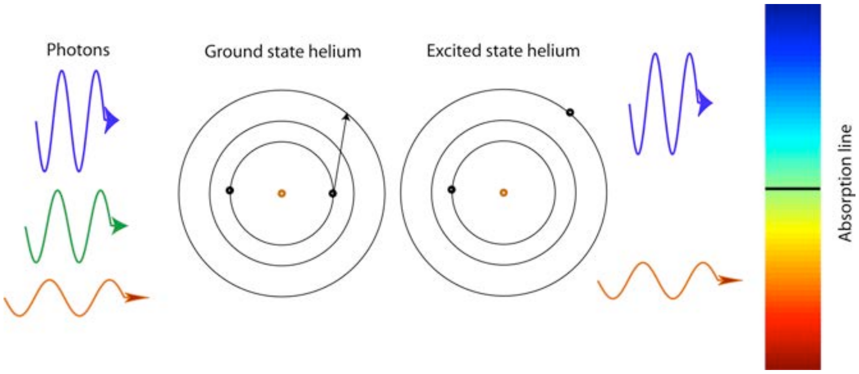

Figure 2: Photons pass through helium at room temperature. The helium is in its ground state with both electrons in its inner shell. Photons with the right energy to put an electron in an outer shell are absorbed putting helium atoms in an excited state. Many lines exist as the electron may be lifted into anyone of many shells. An actual spectral line of helium is yellow which does not plot well on a printed page, nor on a computer screen.

I start with a thought-experiment jar filled with helium atoms (Figure 2). The inner electron shell of helium has room for 2 electrons. At room temperature, the 2 available electrons are in this shell and the atom is in its lowest energy (ground) state. I shine white light through the jar. On the atomic scale, light behaves as particles called photons each with a specific quantum energy. A key experimental result is that energy of a photon is inversely proportional to its wavelength. Violet photons thus have more energy than red photons. The white light consists of all of the wavelengths in the rainbow.

The light passing through the jar interacts with the helium atoms. An atom either absorbs all or none of the energy in a photon. It is most likely to absorb when the quantum energy matches the energy to boost an electron from its ground state to some higher (excited) energy level. This process acts as a filter, which removes the wavelengths of light that match energy level differences. We can observe dark absorption lines in the spectrum of light that passed through the jar (if it is large enough). These dark lines were not present in the original white light.

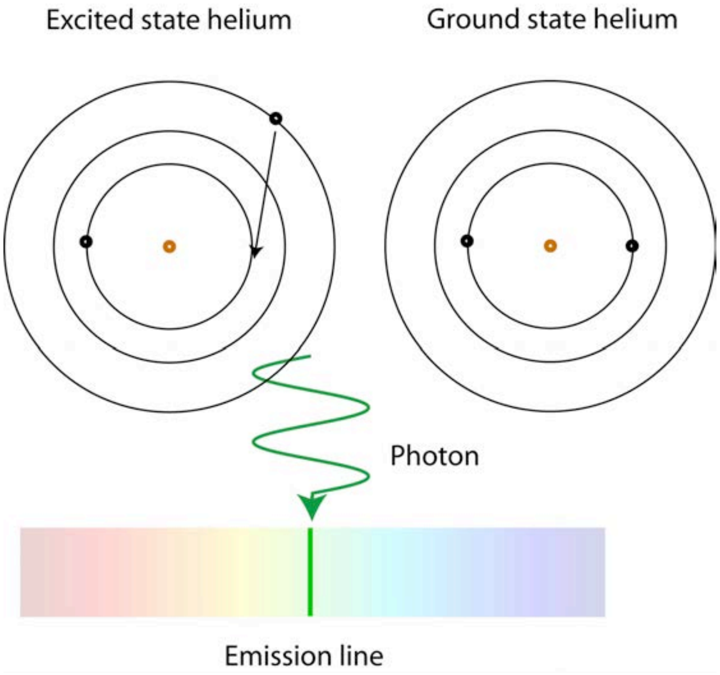

Figure 3: A helium atom in a hot gas is in its excited state with one of its electrons in an outer shell. The atom emits a photon with the right energy to "fall" into the inner shell. This leaves the atom in its ground state. Many lines exist as the electron may start in any one of many shells and fall to a shell that is not the ground state. Actual helium has a strong spectral line in the yellow.

I now fill the jar with hot atoms with a temperature of a few thousand Kelvin (Figure 3). As this is a thought experiment, I do not worry about practical ways to do this. Many of the electrons are in excited energy levels. An atom with an electron in an excited state may emit a photon of light with the right energy for the electron to jump to a lower energy state. We observe this effect as a bright emission lines in the spectra.

Scientists in the late 1800s were familiar with spectral lines even though they did not understand the underlying physics. The energy levels and possible jumps are different for each atom or molecule. The pattern of lines is like a bar code on a soup can that allows the atom or molecule to be identified. This feature provided a routine method of analyzing chemical composition that is still extensively used.

Actual stars have both emission and absorption lines in their spectra. Starting in the late 1800s, astronomers measured the intensity of the lines in the spectra of the Sun and compared them with known gases heated in the laboratory. They found many elements known to occur on the Earth and that most of the Sun is hydrogen. They found strong lines of an unknown element, which they named helium after the Greek name for the Sun. Subsequent study quickly showed that it occurs as a trace inert gas in the Earth's atmosphere. Comparison of the solar spectra with that of other stars provided belated confirmation that the Sun is in fact a star.

The quantum basis of nuclear physics is similar to that of chemistry and the spectra of visible light, except that the energies from reactions are much greater by factors on the order of a million. I present a phenomenological theory analogous to electron shells discussed in Chapter 4. Each nucleus contains protons each with a positive charge equal and opposite to that of an electron. Except for hydrogen with is a single proton, all nuclei contain neutrons. The positive electrical charges of the protons repel each other. The nucleus, however, does not fly apart. Forces, called "strong" by physicists, hold it together.

As in chemistry filled shells result in low available energy and stability. Proton and neutron shells fill separately. The “magic” numbers for filled shells are 2, 8, 20, 28, 50, 82, and 126. Atoms with filled neutron and proton shells are particularly stable. These include helium-4 with 2 neutrons and 2 protons, oxygen-16 with 8 protons and 8 neutrons, and calcium-40 with 20 protons and 20 neutrons. Stable atoms heavier than calcium-40 have more neutrons than protons. Calcium-48 with 20 protons and 28 neutrons and lead-208 with 82 protons and 126 neutrons have two filled shells.

In general, atoms with even numbers of both protons and neutrons are the most stable. Atoms with even protons and odd neutrons come next. Atoms with odd neutrons and even protons follow. Atoms with odd protons and odd neutrons are the least stable. They can decay to an even-even element by emitting an electron (beta particle), which has the net effect of decreasing the neutron number by one and increasing the proton number by 1. They can also emit a position (the anti-particle that annihilates an electron) or capture an electron from the inner shell and take it into the nucleus. Both processes decrease the proton number by 1 and increase the neutron number by 1.

The long-lived isotope potassium-40 decays in two ways (Figure 4). It has 19 protons and 21 electrons. It decays to calcium-40 with 20 protons and 20 neutrons or to argon-40 with 18 protons and 22 electrons. The half-life of potassium-40 is about 1.3 billion years. The argon comprising 1% of our air formed this way from potassium within the Earth over geological time. Geologists routinely use the decay to argon to date rocks and occasionally study its decay to calcium. (Calcium 40 is very abundant in common rocks so it is very hard to detect a little more of it from potassium decay.)

The potential energy barrier discussion in Chapter 4 for chemical reactions carries through to nuclear reactions with the modification that the energies are about a factor of a million higher. The total mass and energy of a closed system are conserved, rather than mass and energy separately. The conversion factor is the well-known formula $E=mc^2$ where c is the speed of light. A small change in mass lets out lots of energy because the speed of light is large. In fact, precise measurements show that chemical and nuclear reactions are coupled to some extent. For example, the decay rate from electron capture depends very slightly on whether an atom is in a chemical compound or pure element. One could detect a minute mass change in a chemical reaction with sufficiently precise equipment.

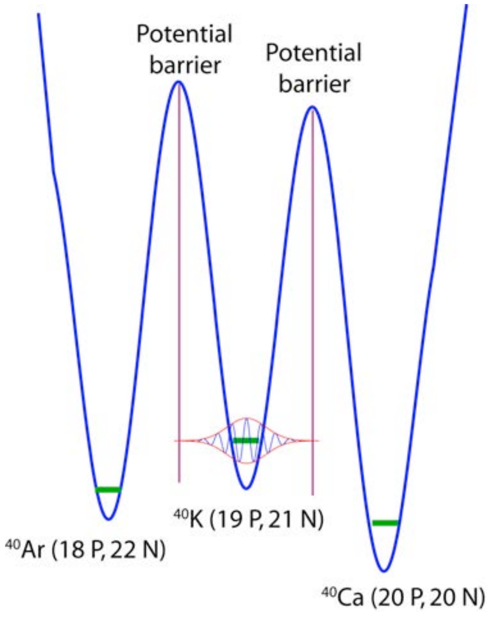

Figure 4: A quantum potential energy surface schematically represents the decay of ^{40}K. The potassium nucleus is in its ground state deep in a potential well. There are no accessible excited states. The state (represented by left and right on the diagram) of the nucleus behaves like a wave in a tub having no definite position at any one time. Very rarely (on a billion year time scale) the state ends up on the other side of one of the potential barriers. The nucleus then tunnels to one of the low energy nuclei, ^{40}Ar and ^{40}Ca. The reaction releases energy when this happens.

Returning to potassium-40, I use standard chemical notation ^{40}K for this atom so that the reader may become accustomed to it. Our ^{40}K atom has been around for geological time and is in its lowest energy state. There are excited energy states, but the thermal energy at room conditions or even the conditions in the deep interior of the Earth (a few 1000 K and pressures millions of times that of the Earth's atmosphere) is too feeble to put the atom into one.

A quantum mechanical effect called "tunneling" allows the atom to eventually decay. Nuclear particles, as do photons, behave partly as waves distributed in space and partly like particles. This hypothesis leads to well-defined but heavy equations. Physicists spend much time solving the mathematical wave formalism so they can compare its predictions with observations. To date, quantum theory has passed these tests.

The actual arrangement of particles in a nucleus is quite complicated. I modify the 2-D analogy of a ball on a ramp from Chapter 4 to simplify discussion. A "standing" water wave sloshing back and forth in a bathtub provides some analogy. The quantum wave oscillates the tub (or potential well) in Figure 4. It can have only certain energy levels, represented by the horizontal lines, which correspond the average height of our water wave when it reaches the side of the tub. The position of the crest of the water wave corresponds to the particle-like behavior of the quantum wave.

Unlike a water wave, the height of the quantum wave on the side of the trough varies statistically about is mean value. Very rarely, the height of the wave is large and the position of quantum wave crest is on the other side of the energy barrier. The particle-like behavior of the crest causes the atom to act as if the whole wave is on the low energy side. Radioactive decay then releases particles from the nucleus that move rapidly away carrying energy with them. This process has no macroscopic analog. It would be like all the water in your sloshing bathtub suddenly going on the floor when just a little bit of it spills over the side of the tub. Quantitative treatment of tunneling involves heavy mathematics.

Physicists routinely observe the quantum energy levels of nuclei. For example, cosmic rays impinging on the Earth's surface collide with stable nuclei putting them into excited states. The affected nucleus may decay to another element say by emitting a beta particle. It may also return to its ground state or an immediate energy state by emitting a high-energy photon called a gamma ray. As with spectral lines of visible light, the energy matches that of the energy jump. The spectrum of emitted gamma rays lets chemists identify an element. Chemists routinely use gamma ray spectra to analyze the composition of materials. Planetary scientists use it to determine the composition of the surfaces of the Moon and Mars from orbit. The Earth's atmosphere is an effective gamma ray absorber, precluding use of this satellite method to analyze the Earth's surface.

Absorption lines also exist for gamma rays. The photon that exactly matches the energy jump from the ground to an excited state is preferentially captured. The precise energy of an atom in a solid depends on its chemical (that is electron) configuration. Chemists routinely use this nuclear-chemical coupling, called the Mössbauer effect, to analyze the structure of solids containing iron. The isotope ^{57}Fe has an energy jump suitable for this precise work.

Composition and energy source of the Sun. By the 1920s, it was clear to astronomers that the Earth and hence the Sun are over a billion years old. This indicated that gravitational energy from the Sun's accretion is inadequate heat source. Lord Kelvin correctly showed that it could keep the Sun going for only about 20 million years in his argument for a relatively young Earth. Unlike the case of the Earth's heat flow, ordinary radioactivity did not offer any help. There is too little uranium, potassium, and thorium in the Sun to supply its vast energy output. Significant amounts would have been obvious in the spectra even in 1900.

These studies of the spectra of the Sun revealed its composition. It is mostly hydrogen with some helium. Only about 2% of the atoms are heavier elements. This fact presented the first key to figuring out how stars work. The second key came from radioactivity and the theory of general relativity.

The change of mass in a candidate nuclear reaction was large enough to detect with early 1900's technology. Four hydrogen atoms are more massive than one helium atom. There is enough hydrogen in the Sun for this reaction to power it for billions of years. Postulating that this reaction occurs, however, is different than showing that it occurs.

Our instantaneous view of space. An unavoidable sampling bias delayed progress. Astronomers are stuck with observing the light that now falls on the Earth and on near-Earth orbital telescopes. They have had this equipment for only an instant of cosmic time. They can see lots of stars, but cannot detect any long-term change in the vast majority of them.

To give an analogous thought experiment, aliens land a probe that took movies of a public square on the Earth for a few minutes and transmitted back to their home planet. They see there are humans of various ages, dogs, and pigeons. It is not obvious to them that babies become children, children become adults, and adults age to become the elderly. They might think it occurs the other way around or that men become women or pigeons become dogs.

Until about 1950, astronomers had this predicament. They suspected that stars gradually cool and become less luminous as they use up their gravitational and nuclear energy. That is, our Sun was once like the more luminous stars and will eventually be among the least luminous. This natural idea, which sets the plot for H. G. Wells' *War of the Worlds*, proved to be grossly incorrect. As in all of science, intuition is often an unreliable master, but a nice guide for formulating a testable hypothesis.

How the Sun works. Progress occurred after W.W. II. Scientists could achieve the energy levels at the center of the Sun with atomic accelerators. The hydrogen bomb was based on these studies; it worked on the first try. Physicists figured out how hydrogen becomes helium at the center of stars.

The temperature at the center of our Sun is about 15.6 million K. The electrons and protons are a dense (162 times water) gas. The pressure is about 248 billion times that of the Earth’s atmosphere. Perhaps surprisingly, solar physicists have direct constraints. The properties of sound (that is pressure or seismic) waves in the Sun’s interior provide much information. Spectral lines and the Doppler effect allow them to measure the velocity of the gas everywhere on the side of the Sun facing the Earth. These measurements are equivalent to placing seismographs (which measure the (shaking) velocity of the Earth’s surface) everywhere on that hemisphere of the Sun. With some heavy mathematics, the scientists determine the density and temperature throughout the Sun.

Pressure is the weight per unit area of an imaginary column of material above a place inside the Sun. The pressure in a stable star and in our atmosphere is the equal and opposite force that balances the effect of gravity. Weight is density (mass per volume) times the acceleration of gravity. It is straightforward to compute the acceleration of gravity with depth within the Sun from Newton's law once density is known. One then computes pressure by extending an imaginary column down from the surface where the pressure is small. The procedure is analogous to that for computing pressure within the Earth's atmosphere in the next chapter.

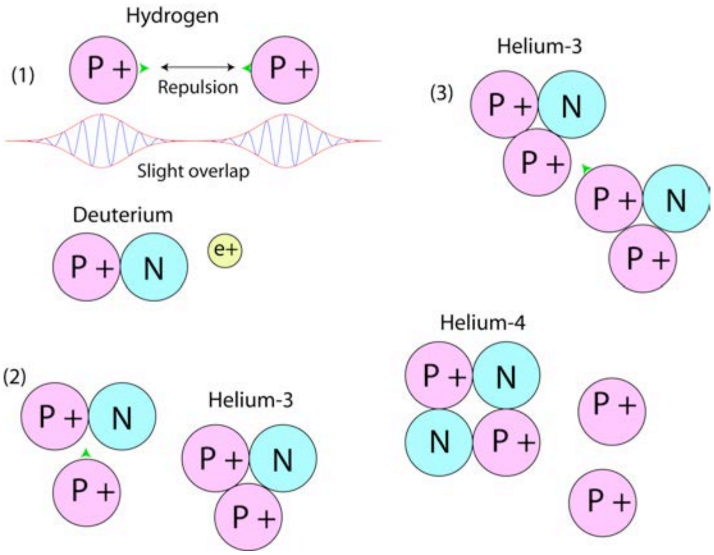

Figure 5: The conversion of protons to helium nuclei takes several steps at the center of the Sun. (1) Protons collide. Typically the like electrical charges repel and no reaction occurs. Very rarely the quantum waves for the protons cause the particles to behave as if they were much closer together. The protons then react to form a deuterium nucleus and a positron (e^+). (2) The deuterium readily reacts with another proton to form a ^{3}He nucleus. (3) Two ^{3}He nuclei collide and react to form a ^{4}He nucleus and 2 protons. The net effect is the change 4 hydrogen atoms into one helium-4 atom.

Now back to the Sun. In the most common fusion mechanism, protons collide with each other (Figure 5). Their positive charges repel. This generally keeps them from getting close enough to react. Rarely pairs of exceptionally energetic protons collide. This alone does not get them close enough. In very rare cases of these collisions, tunneling causes the “wave crests” of the protons to be very close. They react forming deuterium (hydrogen with one proton and one neutron) and a positron that immediately annihilates an electron. The deuterium atom does not last long. It collides with a proton becoming ^{3}He with one neutron and 2 protons. Two ^{3}He atoms then collide. This yields one ^{4}He atom and 2 protons. This process occurs slowly in the Sun as the available protons last billions of years. The hydrogen at the center of our Sun is already half used up. There is enough for the Sun to shine another 5 or 6 billion years.

The net energy output per volume (say energy per meter cubed of gas per second) by nuclear reactions in the Sun is millions of times less than that by chemical energy output per volume in a candle flame. This gradual rate of conversion of mass to energy acts over the huge volume of the center of the Sun, providing its vast energy output. (The interior of the Sun as already noted is vastly hotter than a candle flame.) Physicists cannot observe these nuclear sluggish reactions at sun-center conditions in the laboratory. Astronomers obtain direct information by detecting neutrinos, small nearly massless particles produced by fusion reactions in the Sun. Laboratory physicists observe at somewhat higher collision energies and use quantum mechanics to extrapolate to sunlike conditions. Overall this work confirms that they have a reasonable understanding of fusion within the Sun.

Fusion of hydrogen near the center of the Sun is self-stabilizing. If (slightly) more fusion than usual occurs, the heat from the reaction increases the temperature. The gas expands (like hot air in a balloon) and becomes less dense. The collisions are more energetic because of the higher temperature but rarer because the gas is less dense. The density effect dominates and the reaction rate slows down cooling that part of the Sun. Conversely, if fusion is usually slow, the gas cools and compacts. This increases the reaction rate and heats the gas. This feedback keeps hydrogen-burning stars like the Sun from suddenly exploding.

Fusion within other stars. Moving to other stars, the temperature at the center of stars increases with their mass. Collisions are more energetic and more frequent. Additional fusion mechanisms speed the conversion of hydrogen to helium. The carbon-nitrogen-oxygen cycle occurs within stars larger than the Sun. ^{12}C (6 protons and 6 neutrons) captures a proton and becomes ^{13}N (7 protons and 6 neutrons). This atom emits a positron and becomes ^{13}C (6 protons and 7 neutrons). Capture of another proton yields ^{14}N (7 protons and 7 neutrons). This atom captures another proton to become ^{15}O (8 protons and 7 neutrons). It then emits a positron to become ^{15}N (7 protons and 8 neutrons). The capture of another proton does not yield ^{16}O (8 protons and 8 neutrons). Rather it yields ^{12}C and a ^{4}He atom (2 protons and 2 neutron). The net effect is to make one ^{4}He atom from 4 protons just like the common mode of fusion in the Sun. The ^{12}C atom is not consumed. It is a catalyst, providing yet another analogy between nuclear physics and chemistry.

Overall, the nuclear energy production and hence the light energy emitted by a star increase with about the fourth power of its mass. This relationship explains the surface physics of stars. Hydrogen burning stars are in a (quasi) steady state with heat loss balancing heat generation. A massive star radiates more light than a small star. It does this partly by being larger and thus having more surface area. Its surface is hotter causing more light to be emitted per surface area.

This temperature effect is huge, the amount of energy released from a surface per time per area scales with the fourth power of the absolute temperature. This is a result from quantum mechanics. Its mathematical derivation uses the observation, the energy of a typical photon scales linearly with temperature. The photon's wavelength thus scales inversely with temperature. The number of photons that "fit" in a volume is the inverse of the cube of their wavelength or proportional to the cube of their energy (or the cube of the temperature). The total energy of photons in a volume scales to their number times their energy, which is the fourth power of temperature. Small cool stars that radiate mainly in the red and infrared (wavelengths somewhat shorter than red) are much less luminous than massive hot stars that radiate in the violet and the ultraviolet (wavelengths somewhat shorter than violet). By the way, space telescopes observe spectral lines in both the ultraviolet and infrared providing further constraints on the composition of stars.

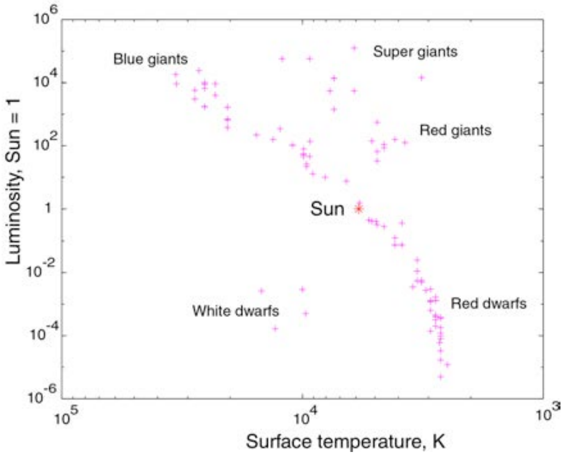

Figure 6: Astronomers Hertzsprung and Russell plotted stellar luminosity plotted as a function of surface temperature. Note log-log scale and that temperature increases to the left. The Sun is part of the main sequence between red dwarfs and blue giants. These stars burn hydrogen to helium. Sun-sized stars evolve into red giants that burn helium. They become white dwarfs when these cease to burn. Blue giants evolve into super giants that then explode as supernovae.

These inferences explain an observation from spectra studies (Figure 6). If one plots luminosity (total energy emission) as a function of surface temperature (given by the dominant wavelength of the emitted light), most of the stars fall in a narrow band called the *main sequence*. These are hydrogen-burning stars. (Figure 6) Other classes of stars exist. There are luminous red stars called red giants and red supergiants. Their diameters measure in AU. There are small hot white stars with low luminosity, called white dwarfs. These uncommon classes of stars figure in the formation and ultimate fate of habitable planets.

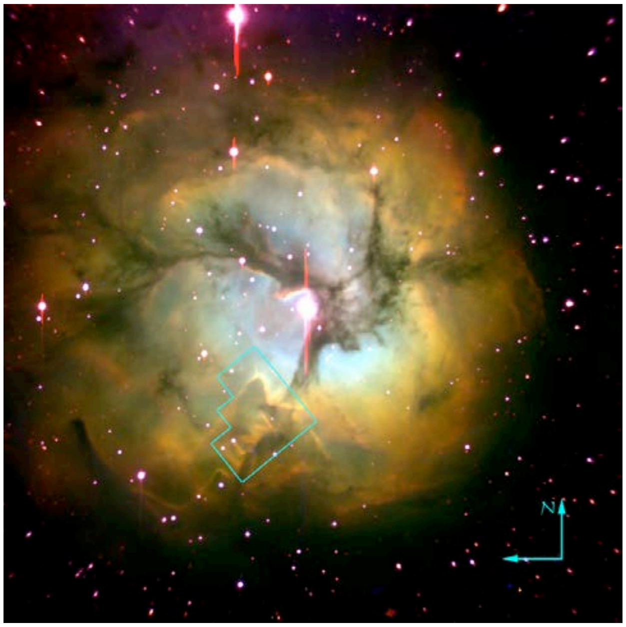

Figure 7: Astronomers observe modern star-forming regions, here the Trifid Nebula. Light from young bright stars is dispersing the gas and dust and exposing very young stars. There are such stars near the top of the blue box. Star formation will end here once the gas and the dust are dispersed. The bright area is somewhat larger than a parsec. Hubble space photo. http://antwrp.gsfc.nasa.gov/apod/ap040618.html

The life history of stars. Stars are now forming from gas clouds near (for astronomers) to the Earth (Figure 7). I follow the evolution of such a cloud. Initially the cloud is cool, about 10 degrees above absolute zero. It begins to collapse under its own gravitation. Some regions collapse faster than others. The collapsing regions rotate conserving angular momentum, like an ice skater pulling in her arms. In a few million years, some of the collapsing regions are dense enough to be optically thick. The trapped gravitational energy heats them up. Soon after, some of the dense regions collapse into rapidly rotating stars. Hydrogen begins to burn to helium at their centers.

I first follow the fate of a star that is 20 times the mass of the Sun. Its intense ultraviolet light heats the surroundings. From our simple scaling, this star has energy for 1/8,000 the life of the Sun or 1 million years. (A more precise calculation yields 7 million years.) It exhausts its interior hydrogen while much of the interstellar gas cloud is still collapsing.

It does not go quietly. Once hydrogen is exhausted, there is no fusion heat source to maintain the high temperature of its center. The center collapses gravitationally become more dense and hotter. Helium-4 is initially inert. If two ^{4}He atoms join, they quickly (within 10^{-16} seconds) fly apart. At temperatures over 100 million K, 3 ^{4}He atoms can join to make ^{12}C.

Once the temperature is high enough to make ^{12}C, a series of exhaustions of nuclear "fuel" and gravitational collapses to higher temperatures occurs. Helium forms carbon. Carbon forms oxygen ^{16}O. Oxygen forms silicon ^{28}Si (14 protons and 14 neutrons). This renewed energy source is huge. The star becomes a supergiant. The intense pressure from the light of the star expels its outer regions. This leaves the inside core (with no hydrogen) exposed. The other elements lighter silicon form during this process.

The star center quickly becomes hot enough that silicon fuses to make iron and nickel. The lowest energy atom is ^{60}Ni (28 protons and 32 neutrons). Additional fusion to make still heavier atoms consumes energy. However, by this time the center of the star is small and dense enough that gravitational energy dominates over nuclear energy. The pressure at center of the star is high enough that electrons combine with protons to form neutrons. Unlike electrons and protons, neutrons do not repel each other until they are packed to the density like that within an atomic nucleus. Within seconds, the core of the star collapses into a few kilometer-wide region composed of neutrons. This process releases vast amounts of neutrinos. Physicists have detected a few of these neutrinos, confirming that the process does occur.

Figure 8: Hot gas diverges from the 1680 Supernova Cassiopeia A. A similar supernova occurred before the Solar System accreted. Our supernova material contained short-lived radioactive isotopes. It locally shocked the gas triggering the Sun's collapse. The field of view is a few parsecs across. Hubble space photo. http://antwrp.gsfc.nasa.gov/apod/ap030830.html

Normally, neutrinos do not interact much with ordinary matter. (Most of those produced in our Sun escape unaffected.) But the interior of our star is quite dense and the high-energy neutrinos collide with the material around the center of the star, ejecting it. Rebound of the shock wave from the initial collapse aids the ejection. The conflagration reaches the surface about an hour later. Over this time, elements heavier than iron form in the neutron-rich environment. The star is a supernova shining with a billion times the intensity of the Sun.

The expelled gas spreads in a shock wave through the gas cloud (Figure 8). This triggers further collapses in some places and disrupts the cloud in others. The debris enriches the cloud in elements heavier than helium. Some of these atoms have short half-lives.

Supernovas and bright massive stars heat our gas cloud. This often expels enough gas that the cluster of stars is no longer gravitationally bound. The stars disperse slowly in the galaxy, each going its own way (Figure 1). In other cases, a somewhat gravitationally bound "open" cluster persists after the gas disperses.

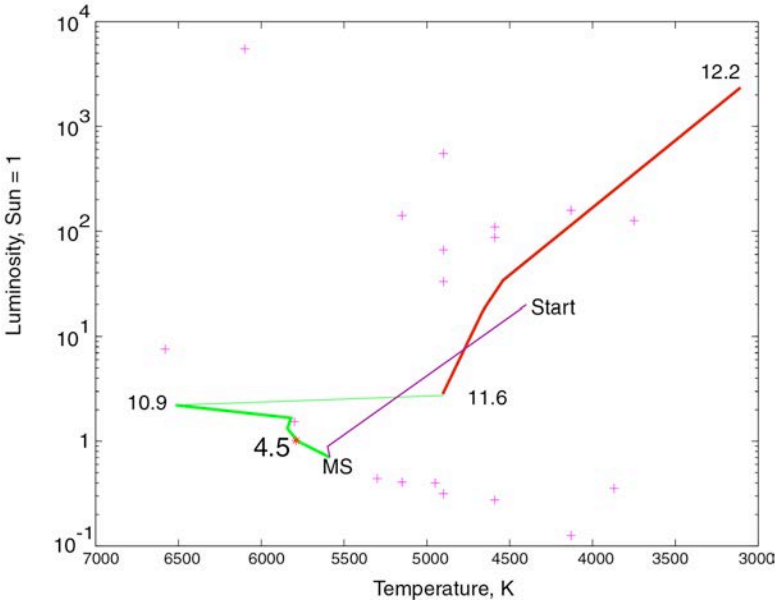

Figure 9: Astrophysicists have predicted the Sun's evolution the Hertzsprung-Russell diagram. Note that temperature is now linearly scaled. Gravitational energy makes the Sun 20 times more luminous than now during its collapse. The Sun joins the main sequence and starts to burn hydrogen. It gradually becomes more luminous. It is now about 4.5 billion years old. It will expand and become redder between 10.9 and 11.6 billion years after its formation. It then becomes more luminous as its interior contracts and becomes hotter, while its surface expands and becomes cooler. The curve ends at 12.2 billion years where helium ignites in the deep interior. After that the evolution is complicated by episodes of unstable helium ignition.

A sun-sized star has a more protracted history (Figure 9). First it collapses gravitationally and becomes a star. Initially it rotates rapidly. Magnetic fields couple the rotating star to its solar wind. This transfers angular momentum to the wind and slows the rotation down. Gravitational energy is a powerful and, as Lord Kelvin suspected, shortterm energy source. A sun-size star is initially 10 times as luminous as the Sun. After a few million years (basically the time computed by Kelvin), the star settles down and joins the main sequence. It gradually burns hydrogen to helium. This causes the interior of the star to become more dense and hotter. Hydrogen burns at an increasingly faster rate. The luminosity of our Sun has increased by 30% over its lifetime.

After ten billion years our star exhausts the hydrogen at its center. Like the massive star, it does not go quietly. It contracts until helium can burn. This process is somewhat unstable. The outer regions of the star expand to around an AU. It is a red giant. Much gas gets expelled at times of energy surges. Eventually, only the hot core remains. It has no more nuclear material to burn. It becomes a white dwarf and cools slowly over billions of years. Depending on its initial size and the amount of gas expelled, it may end up composed mainly of helium or of carbon and oxygen.

Minor amounts of elements heavier than oxygen form in proton-rich and neutron-rich regions of red giants. Geochemists have been detected within pre-solar silicon carbide (grinding powder) grains in meteorites. They have also detected elements produced by supernovas soon before our solar system formed. The element abundances in the solar system are those expected from a mixture of original gas cloud material with supernova and red giant debris.

Note that, even though ^{60}Ni is the lowest energy nucleus, iron is a more abundant product of supernovae. Nickel atoms absorb gamma rays more readily than do iron atoms, putting them in excited states and making them vulnerable to further nuclear reactions. This explains why iron is more abundant than Ni in the solar system. For example, the Earth contains about 30% iron by mass but only 3% Ni. All this supports the hypothesis that our sun formed in a region similar to modern gas clouds and open clusters.

Making planets

After the Big Bang, the universe had only hydrogen, helium, and a slight amount of lithium. The first stars formed from this mixture. The elements in rocks, like silicon, magnesium, and oxygen did not exist. Neither did the elements in ices, like oxygen and carbon. Since these elements did not exist yet, it is impossible for very old stars in our galaxy to have icy or rocky planets.

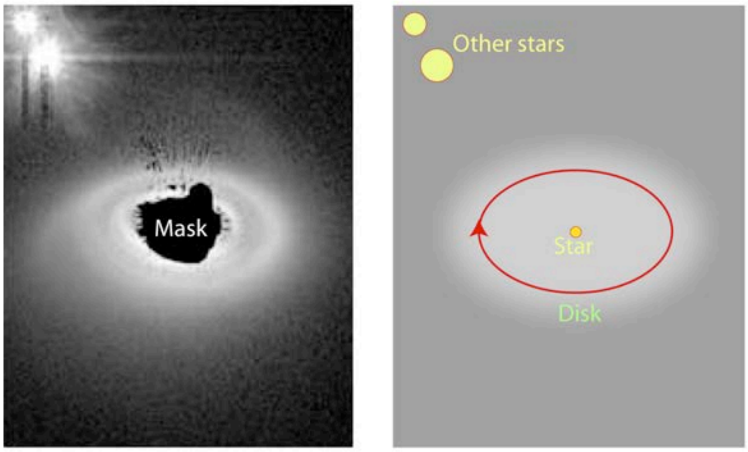

Figure 10: Astronomers can see clear signs that planets are now forming, in the protoplanetary dust and gas disk around star HD 141569A. The disk is 1200 AU across. Our Sun had a smaller disk. The black area in the center is a mask that blots out the bright light of the star so that the faint disk was visible. The disk rotates clockwise.
Hubble space photo. http://hubblesite.org/newscenter/newsdesk/archive/releases/2003/02/image/b

Rock and ice forming elements built up in the galaxy over time as young massive stars and older red giants supplied light elements and supernovas supplied both light and heavy elements. (Astronomers call all elements heavier than lithium metals. This usage is a vestige of the fact that metal atoms in the usual chemical sense are easy to detect from spectra.) The fraction of metal atoms varies from place to place within our galaxy. It is easier to form earthlike rocky planets around some stars than others.

Solar nebula. Just then how do planets form? The modern theory involves the rotating gas cloud that collapses to form a star (Figure 11). Early geologists were familiar with this hypothesis. Buckland discussed it in his 1837 book. Late in the 19th and early twentieth century, some scientists contended that the planets formed in the near collision of our Sun with another star. Since stars rarely come close in the vastness of space, this appealed to those who thought the Earth is special. The physics do not work out though. A near miss does not leave planets in its wake.

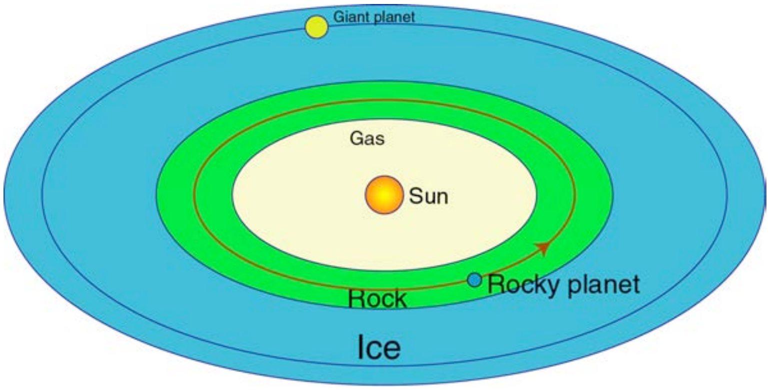

Figure 11: The protoplanetary dust and gas disk orbits the Sun at a time it was more luminous than at the present. The region near the Sun is so hot that only gas is present. Rock and iron metal condense farther out forming rocky planets like the Earth. Still farther out, it is cool enough for ice to condense. This leads to the formation of icy objects that accrete gas becoming giant planets. The gas orbits more slowly than the planets. The effects of stellar winds and UV light from nearby bright stars limit the outer edge of the disk. Viewed obliquely with the top farther away, not to scale.

A few years ago, astrophysicists had only our solar system as an example. They thought they understood that the formation of the planets occurred at about their present distance from the Sun. They understood the basic physics, which applies reasonably well to our solar system.

Our solar nebula was composed of elements similar to our Sun. For convenience, it contained 98% helium and hydrogen, 1.5% of material like water and carbon dioxide that form ices at low temperatures, and 0.5% of material that forms rock and iron metal. The Earth and other terrestrial planets formed from this half percent. It is obvious that it was hot near the Sun and cool further out. The contracting Sun was a lot brighter than the modern one. It was thus very hot at one AU. Rock and iron condensed outward of a fraction of an AU of the Sun, icy material condensed in the cool outer regions beyond 5 AU. Hydrogen and helium remained a gas. This arrangement yielded rocky terrestrial planets near the Sun.

Farther away, the collapse of icy objects contributed to the collapse of the surrounding gas in the outer solar system. Jupiter and Saturn contain much nebula gas. Uranus and Neptune are mostly icy and rocky material but have enough gas to be giant planets. Pluto is a small icy object that never collided with a larger planet.

Astrophysicists understood the basic physics of the collapse of (ice or rock) dust into planets. Gas rotates slower than solid material in Keplerian orbits because its pressure partly resists its fall into the Sun (Figure 11). Gas drag causes the rock orbits to become circular and to collapse to the plane of the rotating solar nebula. The solid particles become dense enough to collapse under their own gravitation in kilometer to 100's of kilometer diameter objects called planetesimals (Figure 12). There is a minimum size of object that can collapse for a given mass per area (surface density) of dust on the disk. The velocities imparted by the gravity of the collapsing region must exceed the random velocities of the dust. There is also a maximum size of the instability. The outer part of the nebula revolves more slowly than the inner part. A slowly collapsing instability will be sheared before it can collapse. No collapse can occur if the minimum size is greater than the maximum size. This situation prevails for gas, which cannot collapse unless a massive icy or rocky object is already present (its pressure resists compaction).

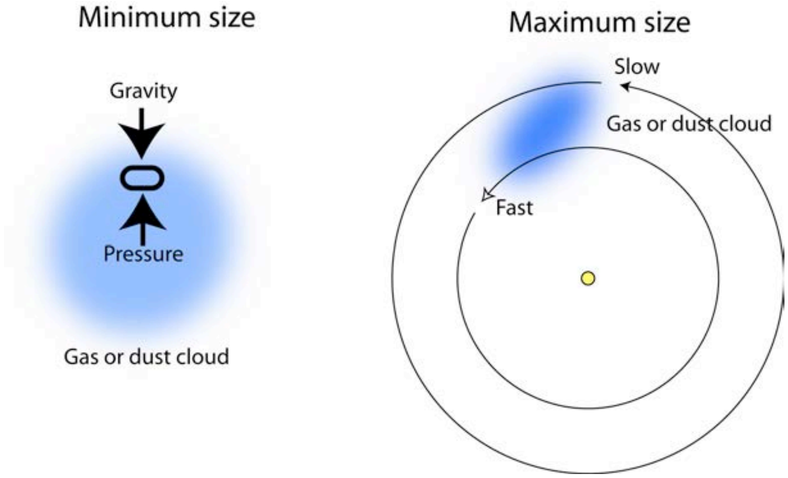

Figure 12: A gas cloud collapses when the force from its gravitation generates pressure sufficient to unstable compress it. The cloud must exceed a minimum size depending on its temperature and the mass per surface area of the disk for this to happen. The outer part of a gas cloud revolved more slowly about the star than the inner part. The effect shears out the gas cloud before it can collapse. This sets a maximum size of a collapsing cloud. Dust clouds behave similarly with random motions in the disk having the effect of molecular motions and pressure in a gas.

Making planets from rocks. Once planetesimals are present, modeling of the formation of terrestrial planets becomes conceptually simple, but complicated. The bodies collide with each other eventually forming planets. The gravity of the largest planetesimals attracts smaller ones causing the process to become unstable. Formation of giant planets requires two steps. Massive icy objects grow until they can trigger gas collapse. Jupiter and Saturn accreted large amounts of gas. Uranus and Neptune accreted modest amounts.

Planet formation is complicated, involving turbulence in the gas, the tendency of dust to stick mechanically, and even gravitational effects and ultraviolet light of nearby stars in the nascent cluster. In retrospect, the 1990 astrophysical theory lacked predictability. Astrophysicists still cannot even hindcast our own solar system.

Detected extrasolar planets. This idea that all planets formed about where they sit collapsed when astronomers, Geoff Marcy, Paul Butler, Michel Mayor, and Didier Queloz began to detect planets around other stars. Their method is simple in concept. Planets revolve around the center of mass in their solar system. (See Figure 13.) The star is very massive so the center is typically within it. The star moves back and forth around the center of mass and hence flees and approaches the Earth. (This method observes only the motion along the line of sight.) For Jupiter-sized planets, the star velocities are meters per second, like those of a healthy jogger. The astronomers had to develop precise equipment before they could start looking.

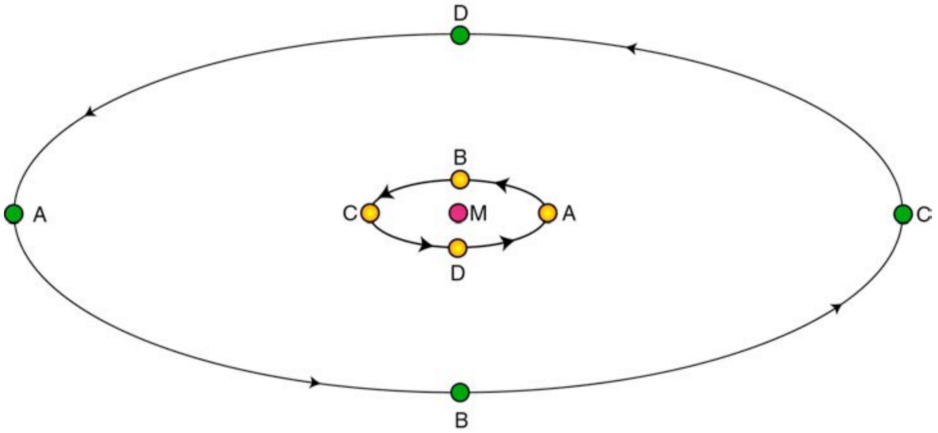

Figure 13: Careful observations let astronomers detect planets around other stars. The planet and the star orbit around the center of mass M. A, B, C, and D denote positions of the planet and star around the orbit. Viewed obliquely from the Earth the circular orbits are ellipses. Astronomers observe the motion along the line of sight. The star moves toward the Earth through BCD and away from the Earth through DAB. Astronomers attempt to resolve the elliptical path of the star against the background of more distant stars. If the system is lined up so that we see it edge on, the planet transits the disk of the star dimming the light that we see. The situation with elliptical orbits and several planets is more complicated but analogous.

In 2006, astronomers had detected planets around 5% of the stars examined, a total of over 100 out of over 2000. These are all giant planets. The mass of a rocky (terrestrial) planet is too small to have a then measurable effect with their equipment. The detected giant planets are typically near their stars, often much closer than an AU. This is a sampling bias in that close-in planets produce the largest velocities. The year of close-in planets is short so it is easy to resolve that the star moves back and forth over a reasonable observation period. Improved detection technology and continued effort has increased the fraction of stars with found planets along with decreasing the minimal detectable size (in 2008 to 5 Earth masses) and increasing the maximum detectable year.

Still, the result shocked astrophysicists. They did not realize that giant planets spiral in toward their star after they form. Again the physics are simple, but extremely complicated to model (Figure 14). There are two mechanisms. (1) A giant planet cleans up the gas in the nebula near its orbit. Pressure variations in the gas cause some of it to flow into the void. This occurs irregularly at the edge of a donut-shaped region surrounding the planet’s orbit. Gravitational interactions set up waves in the gas, coupling the orbit of the planet to the gas. As the gas is revolving slower than the planet, the effect is like drag. The planet spirals inward, especially if the nebula becomes denser inward. The process ends either when the planet collides with its star, reaches a gas free region near the star, or the nebula disperses. (2) Gravitational interactions between the planets remaining at that time may expel some of them from their solar systems, cause others to collide with their star, or scatter others to enter highly elliptical orbits. Astronomers have detected numerous giant planets in highly elliptical orbits. Some of these evolve in close circular orbits that may be in the opposite direct that the star rotates.

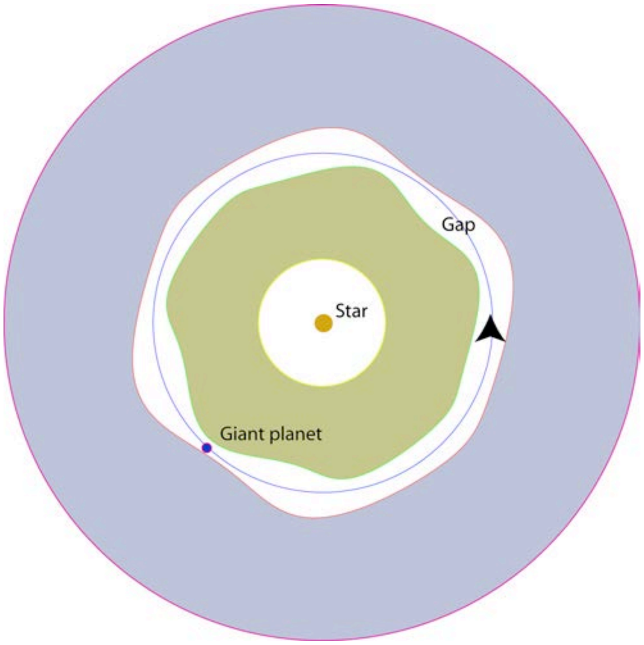

Figure 14: Schematic diagram of giant planet within nebula around star, not to scale. The giant planet captures gas and dust near its orbit, producing a gap in the nebula. The gas inside and outside the gap tends to flow into it. Gravitational attraction between the gas and the planet sets up waves that couple the orbit of the planet to the orbit of the gas.

There are subtle hints of this process in our solar system. Jupiter is enriched in the rare gas argon relative to solar abundance. This indicates that argon condensed as an ice, which could happen only in cold regions much farther from the Sun than Jupiter's present position. That is, Jupiter spiraled in. Uranus and Neptune appear to have spiraled out. Astronomers routinely detect 100's of kilometer diameter objects and small planets beyond the orbit of Neptune. These "Kuiper belt" objects, however, have little total mass. This dearth may indicate the effects (including intense ultraviolet light) of nearby stars forming with the Sun that kept the diameter of the nebula in check. The Kuiper belt objects have orbits that indicate they moved outward along with Neptune.

Other ways to look for planets. We would like to detect terrestrial planets around other stars and giant planets at Jupiter-like distances. As Bruno guessed, this is not an easy task. In addition to line of sight accelerations, three other methods seem practical. A fourth gives some information.

First, we can use the motion of the star perpendicular to our line of sight (Figure 13). The star revolves around the center of mass and sweeps an elliptical path in the sky. (A circle viewed at an oblique angle is an ellipse. You can see that with a sausage slice. See primer on ellipses in Chapter 3.) This method is most sensitive to giant planets that are well away from their star. It takes them several Earth years to complete their orbits. One needs to observe for over 10 years to separate the orbital motion from the unknown proper motion of the star. So far this method initially worked only with already discovered planets.

Second one can wait for planets to transit their star, as long as the planet's orbit is lined up right. Astronomers have observed the transit of planets already detected by line-of-sight velocities. A close-in planet is most likely to transit and the star dims slightly during the transit. Transits have the advantage that the spectrum of the star is slightly affected by light that passes through the planet's atmosphere. This allows astronomers to determine the composition and the temperature of the planetary atmosphere. It also teels if the orbital motion and rotation of the star have the same sense. Well-equipped amateurs detect giant planets this way. The sensitive space-based Kepler telescopes has detected 1000's of planets and planet candidates in transit, including terrestrial planets. Statistics of the geometry of transit indicate that stars typically have planets.

Third, one can try to see the reflected or emitted light of a planet directly. The Earth, Venus, and Jupiter are about one to one-tenth billionth as bright as the Sun in visible reflected light. This is plenty bright enough to see with modern telescopes at the distance of nearby stars. The difficulty is, as Bruno suspected, the glare of the star. One way around this is to take advantage of the fact that light is a wave. Instruments called interferometers are built to null out the light of a star in order to see planets. This works best in the thermal infrared where the emitted light of the planet is a millionth that of the star, rather than a billionth with visible light. This method provides spectra of the planet, particularly on the existence of ozone and methane in the atmosphere (See the next chapter for their significance to life.) Astronomers have observed emitted light from already known extrasolar planets.

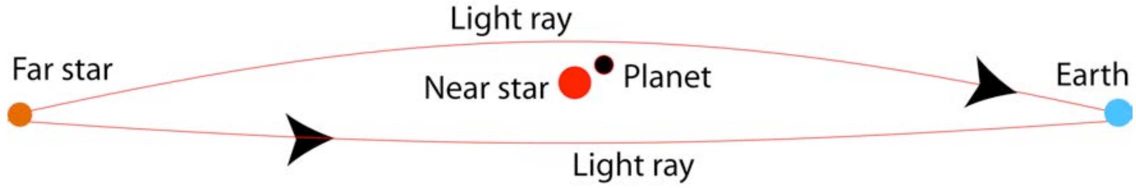

Figure 15: Schematic diagram of a nearer star acting as a gravitational lens for light from a distant star. The distant star appears to brighten when the alignment produces the lens. The presence of a planet produces additional focusing of light waves that pass near it. The amount of bending is grossly exaggerated in the diagram.

Finally, one can use the general relativistic effect of gravitational lensing (Figure 15). The light from a distant star brightens when a nearer star is directly in line with the path of its light to the Earth. This situation rarely occurs for an individual star. Stars fill only a minute proportion of the sky. Otherwise the sky would be as bright as the disk of the Sun. See Obler's paradox in Chapter 5. There are, however, a vast number of stars between the Earth and the center of our galaxy. Automated telescopes check lots of stars periodically and wait for transient changes in star brightness. The fluctuation in brightness is simple if the nearer object is an isolated star. The gravity of a planet around the nearer star causes a more complex but interpretable signal. Planets have been detected in this manner. The method has the major disadvantage that the nearby star will not likely pass in front of another distant star in reasonable future.

Bottom line: The Sun also rises

The Sun is a typical star in the middle of the range of masses and luminosity. It contains the elements that dying stars expel and hydrogen and helium from the Big Bang. The Earth and other solar system objects that we have sampled have the elements that we expect to condense from a nebula of solar composition. Biologically important elements (including hydrogen, carbon, nitrogen, oxygen, sodium, magnesium, phosphorus, sulfur, chlorine, potassium, calcium, and iron) are well represented in the accreting material and in rocky objects like the Earth and meteorites. There is nothing special in this regard about either the Earth or the Sun.

Astrophysics tells us that our Sun will become a red giant in 5 to 6 billion years. Then the Earth will be incinerated if not consumed altogether. Life has taken a long time in evolving to its present state. What are our prospects of other earthlike planets? Where should we look?

Astrophysical theory is some help. The star-forming gas must contain "metals" to form rocky planets but the relationship of gas composition to rocky planet size is unknown. The first stars in our galaxy formed from metal-free gas. It took time - billions of years - for supernovas to build up metals in our galaxy, but local metal-rich regions exist in the oldest galaxies. About 10% of the stars have metal content close to the Sun. Many of these formed before the Sun. Fortunately for our search, the 10% includes many neighboring stars that we can actually study.

The other hard exception involves the time for planets to form and life to begin. We need not look at bright massive stars that meet a violent end within millions of years after they form. These stars are rare anyway so little living room is lost.

Other situations are unpromising for rocky planets but not exclusionary. Multiple star systems often preclude stable terrestrial planet orbits. The exceptions are when the stars are much closer than an AU or much more apart than an AU. Astronomers in fact have detected such planets. The 1000s solar systems with giant planets near the star are also unpromising. Again there are no stable orbits for rocky planets near the giant planets. The exceptions include icy and rocky moons of giant planets.

Going to the other end, faint red stars last almost forever. (None have yet entered the red giant stage in the observable part of the universe.) The smallest planets so far detected orbit red stars. This again is a sampling bias. The small mass of red stars makes it easier to see the feeble gravitational tug of planets. A similar bias occurs in transit as the diameter of red stars is small. I address the habitability of their planets after I discuss climate and evolution.

Our search for other Earths has just started. As in the time of Bruno, astronomers lacked technology that would detect our Earth around nearby sunlike stars. They can barely detect a Jupiter-sized planet at Jupiter's distance from its star. This changes as space-based observatories come in place. In practice, this search will continue to be controlled by what we can detect, not by theory. Whether we like it or not, it's rather like a drunk looking for lost car keys only in the well-lit areas.

Notes

I obtained the history of the Sun from an article "Our Sun .3. Present And Future" by I. J. Sackmann, A. I. Boothroyd, and K. E. Kraemer in Astrophysical Journal, November 20, 1993, volume 418, number.1, part 1, pages 457-468. This article provides a good summary for those with some background in physics.