15 Exoplanets: Habitability and Characterization

Until recently, the study of planetary atmospheres was largely confined to the planets within our Solar System plus the one moon (Titan) that has a dense atmosphere. But, since 1991, thousands of planets have been identified orbiting stars other than our own. Lists of exoplanets are currently maintained on the *Extrasolar Planets Encyclopedia* (http://exoplanet.eu/), NASA's *Exoplanet Archive* (http://exoplanetarchive.ipac.caltech.edu/) and the *Exoplanets Data Explorer* (http://www.exoplanets.org/). At the time of writing (2016), over 3500 exoplanets have been detected using a variety of methods that we discuss in Sec. 15.2. Furthermore, NASA's *Kepler* telescope mission has reported a few thousand additional planetary "candidates" (unconfirmed exoplanets), most of which are probably real. Of these detected exoplanets, a handful that are smaller than 1.5 Earth radii, and probably rocky, are at the right distance from their stars to have conditions suitable for life (Batalha, 2014; Batalha *et al.*, 2013; Borucki *et al.*, 2011). Describing the different planetary detection techniques and summarizing the data collected so far would require a book in itself (which would immediately be out-of-date), so we shall not attempt to do that. Rather, we will focus on general issues related to planetary habitability and the detection of life, both of which build on concepts discussed earlier within this book.

For general discussions of exoplanets, several good reviews are available. Kasting (2010) covers some of the material discussed here, but at a lower level. Tutorial books at approximately the upper undergraduate level include Scharf (2009) for exoplanet astrobiology and Haswell (2010) on transiting exoplanets; both have great clarity. Seager (2010), Perryman (2014), and Winn and Fabrycky (2015) review exoplanet detection and characterization at the researcher level.

15.1 The Circumstellar Habitable Zone

We focus our attention on Earth-like planets and on the possibility of remotely detecting life. We begin by defining what life is and how we might look for it. As we shall see, life may need to be defined differently for an astronomer using a telescope than for a biologist looking through a microscope or using other *in situ* techniques.

15.1.1 Requirements for Life: the Importance of Liquid Water

Biologists have offered various definitions of life, none of them entirely satisfying (Benner, 2010; Tirard *et al.*, 2010). One is that given by Gerald Joyce, following a suggestion by Carl Sagan: "Life is a self-sustained chemical system capable of undergoing Darwinian evolution" (Joyce, 1994). This definition is useful for laboratory scientists: If they make a self-replicating, mutating chemical system, they will hopefully recognize it, regardless of its construction. The same applies for astrobiologists who look for life on Mars or Titan, although efforts to date have focused on looking for signs of metabolism rather than a genome, e.g., the *Viking* mission to Mars (Sec. 12.1.3). We have not yet seen evidence for life in the Solar System, but as exploration becomes more sophisticated, we could possibly identify life – even exotic life – if it exists.

Astronomers are not so fortunate because they have no hope of actually visiting exoplanets in the foreseeable future. Instead, they must use remote detection. One way to find life involves taking spectra of exoplanet atmospheres and looking for gases such as O_2 or CH_4 that might be produced by life. Other, more general, life-detection criteria have been proposed, e.g., finding extreme thermodynamic disequilibrium (Lederberg, 1965; Lovelock,

1965), which we discuss later (Sec. 15.4). But such criteria may be difficult to interpret if they involve chemical signatures that are alien from those we find on Earth. For this reason, it makes sense to concentrate initially on planets that could support life similar to that found on Earth. A key requirement of Earth-like life is the availability of liquid water. It will be much easier to believe that some possible biosignature is indeed evidence for life if the planet on which it is observed supports liquid water. Of course, the most credible remote biosignature of all would be a radio or visible light transmission from an extraterrestrial civilization. But if such signals exist, the well-known Drake Equation implies that they should be rarer than the atmospheric signatures of biospheres (e.g., Vakoch and Dowd, 2015), which is our focus here.

Exoplanets with hidden subsurface life are not practical candidates for remote life detection. For example, Mars is dry and lifeless at its surface, but might harbor liquid water at depth and, at best, a meager subsurface biosphere. But even though Mars is our planetary neighbor, a contemporary debate has erupted about whether trace levels of methane are actually present in its atmosphere, irrespective of whether such methane is biogenic or abiotic (see Sec. 12.2.1). Detecting similar miniscule signals from an exoplanet atmosphere is beyond the realm of feasibility for the foreseeable future. For life to influence a planet's atmosphere in a way that is readily detectable from interstellar distances, it needs to exist on the planet's surface. For life as we know it, this means that liquid water must also be present at the surface. Hence, we focus our interest on planets that could support surface liquid water. These planets orbit in the region that we now call the *circumstellar habitable zone*.

15.1.2 Historical Treatment of the Habitable Zone

The idea of a habitable zone has a long history. William Whewell was the first to note, in a book about extraterrestrial life, how Earth's orbit is in a *temperate zone* between a "central torrid zone" and external "frigid zone" (Whewell, 1853). A century later, the astronomer Harlow Shapley (1953) defined a *liquid water belt* as the region in a planetary system where liquid water could exist at a planet's surface. Also at that time, Strughold (1953, 1955) defined an analogous *ecosphere* around the Sun. The astronomer Su-Shu Huang (1959, 1960) then identified a variety of issues bearing on habitability. He pointed out that binary or multiple star systems are less likely to harbor habitable planets than single stars because the planetary orbits would be unstable in most cases. He also concluded, perhaps correctly, that stars that are similar in mass to our Sun are the most likely to have habitable planets. But his most lasting contribution, probably, was to coin the term *habitable zone* as a synonym for Shapley's liquid water belt.

Shortly after this, in *Habitable Planets for Man*, Dole (1964) considered a more focused question: How many nearby stars might harbor planets suitable for human colonization? His anthropocentric conditions for habitability included mean annual temperatures of 0–30 °C over 10% of a planet's surface, an O_2-rich atmosphere, and a surface gravity less than 1.5 times that of Earth, so that humans could walk upright. Dole's climate models were crude: black planets with no greenhouse effect and 45% cloud cover. Nevertheless, Dole made a number of valid points, reiterating Huang's concerns about orbital stability in binary star systems, and pointing out the problem of tidal locking of planets orbiting red dwarfs, which we discuss later.

None of the early researchers attempted to define the habitable zone using a realistic climate model until Hart (1978, 1979). In retrospect, Hart's climate models were flawed, but his papers were influential and provoked further scientific interest. In particular, Hart defined a *continuously habitable zone*, or CHZ, for short, as the region around a star where a planet could support liquid water for some specified period, usually the star's main sequence lifetime. The conventional *habitable zone*, then, can be abbreviated as "HZ," and is defined at a single instant in time. The HZ must move outwards with time because main sequence stars brighten as they age (Sagan and Mullen, 1972), as indicated in Fig. 15.1. Suppose the Sun's initial HZ at time t_0 covered the range of distances shown in the figure, and that by some later time t_1 it had moved further out, as indicated. Then the CHZ is represented by the overlap between the two regions.

Hart's approach to calculating the boundaries of the HZ was ambitious. He performed time-dependent calculations for planetary atmospheres and climates. His models included a simplified greenhouse effect, along with numerous physical processes, including outgassing of volcanic CO_2, reactions of CO_2 with surface minerals, the presence of reduced greenhouse gases (CH_4 and NH_3) early in Earth's history, organic carbon burial, the rise of O_2, changes in solar luminosity, and ice albedo feedback.

A key flaw in Hart's climate model was that he underestimated the CO_2 greenhouse effect, making the outer edge of his HZ too small. Volcanic CO_2 could build up in a planet's atmosphere if the planet became entirely ice-covered, but a planet could not deglaciate (Levenson, 2015). Consequently, Hart determined that the outer edge

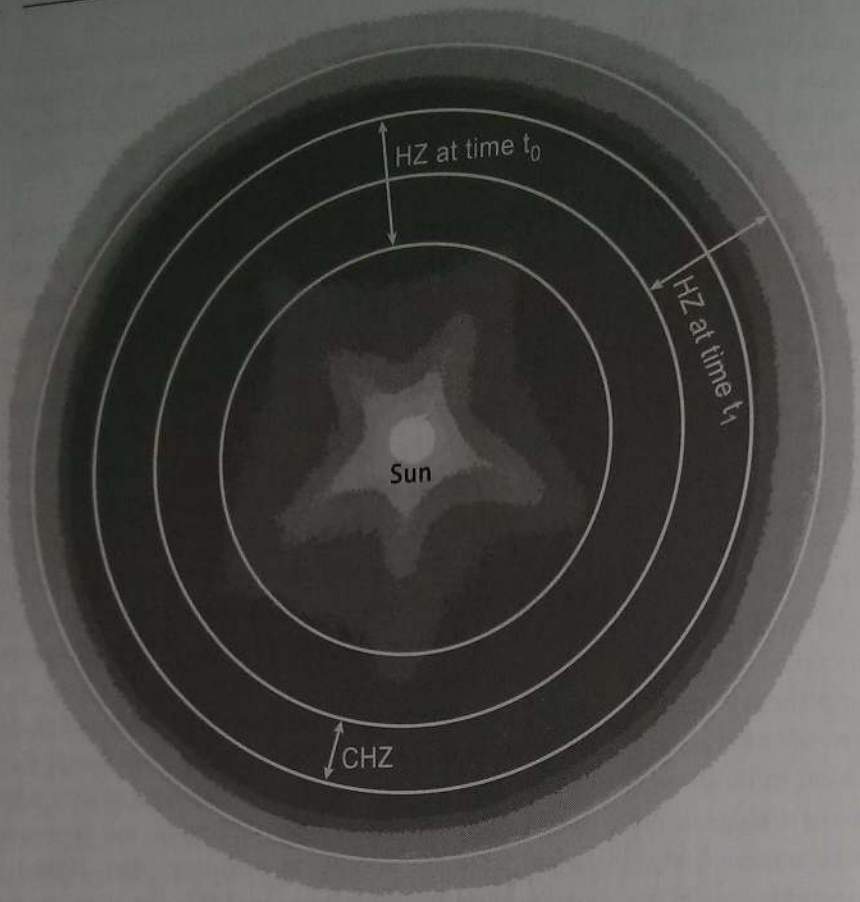

Figure 15.1 Diagram illustrating the habitable zone and continuously habitable zone (HZ), which shifts between times t_0 and t_1, and the continuously habitable zone (CHZ) around a star. (From Kasting (2010). Reproduced with permission of Princeton University Press. Copyright 2012.)

of the HZ for our own Solar System was at 1.01 AU. If the Earth had formed farther from the Sun than this, it would have remained perpetually frozen. We have already argued, though, that a Snowball Earth planet *should* be able to deglaciate (Sec. 11.10). Hart also found that a runaway greenhouse would have occurred if the Earth had formed inside of 0.95 AU. This prediction turns out to have been reasonable, as the best current estimate for this threshold based on 3-D climate modeling is at that same distance (see next section). In any case, the overall conclusion of Hart's (1978) paper was that the 4.6 b.y. continuously habitable zone around the Sun was quite narrow, 0.95-1.01 AU. Then, in his 1979 paper, Hart argued that the CHZ around other main sequence stars was even narrower, or nonexistent. The bottom line was that habitable planets were extremely rare. Indeed, pessimists who believed Hart's papers might have concluded that Earth was the only one in the galaxy.[1]

[1] Hart also repeatedly argued that humans are special and that advanced life is absent from the rest of the Universe (Hart, 1975, 1982). These views fit rather snugly with the results of his habitable zone models.

15.1.3 Modern Limits on the Habitable Zone Around the Sun

Estimating the inner and outer boundaries of the habitable zone for our own Sun is closely related to the problem of understanding long-term climate evolution on Venus and Mars, respectively. In Sec. 13.4, we discussed Venus' susceptibility to a runaway greenhouse, which sets a hard inner edge to the HZ, while in Sec 12.5.2, we considered how a CO_2–H_2O greenhouse can fail to warm early Mars, which is an issue pertinent to the outer edge of the HZ. Given these earlier discussions, we summarize the ideas here.

The inner edge of the HZ is determined by when a planet develops a wet stratosphere and loses its water through photodissociation, followed by escape of hydrogen to space (Kasting *et al.*, 1993b; Kopparapu *et al.*, 2013). In the model of Kopparapu *et al.* (2013), this happens when the solar flux became more than ~2% higher than its present value at Earth or, equivalently, at a distance of ~0.99 AU. This is the *water-loss limit* of Kasting *et al.* (1993b). The *runaway greenhouse limit* in the Kopparapu *et al.* model occurs at ~0.97 AU, corresponding to an effective solar flux, S_{eff}, of ~1.05

(The effective solar flux is defined as $S_{eff} = S/S_0$, where S is the solar flux at some distance r from the Sun and S_0 is the solar flux at Earth's orbital distance of 1 AU.) As pointed out in Sec. 13.4.5, a 3-D calculation by Leconte *et al.* (2013) does *not* predict a wet stratosphere because their tropopause is extremely cold. The inner edge of the HZ in this model is determined by the last stable climate simulation, which occurs at S_{eff} = 1.1, or 0.95 AU.[2] Henceforth, we use 0.95 AU as the conservative estimate for the inner edge of the habitable zone, as this 3-D calculation is arguably better than the 1-D model estimates. By accident, this limit agrees with that calculated over 20 years or so earlier by Kasting *et al.* (Kasting, 1988; Kasting *et al.*, 1993b).

Following this same line of reasoning, the outer edge of the HZ can be estimated by modeling the effect of slowly sliding the Earth out toward Mars' orbit. As this happens, the climate should get colder, silicate weathering should slow down, and volcanic CO_2 should build up in the planet's atmosphere. This negative feedback is sufficient to prevent global glaciation, provided that the solar flux is above a critical value. However, at some distance, CO_2 clouds begin to form in the planet's atmosphere (the *first condensation limit*), and at some greater distance CO_2 condensation becomes so extensive that further greenhouse warming is prohibited (the *maximum greenhouse limit*). The relevance of the first condensation limit depends on whether CO_2 clouds can warm a climate or not, which recent models don't favor (see Sec. 12.5.2.2). The *maximum greenhouse* limit is a firmer but more generous limit for the outer edge of a CO_2–H_2O greenhouse. In the Kasting *et al.* (1993b) model, this limit was reached at 1.67 AU, or $S_{eff} = 0.36$, whereas Kopparapu *et al.* (2013) place it at 1.69 AU, or $S_{eff} = 0.35$. Both estimates are well outside of Mars' orbital distance of 1.52 AU. So, as pointed out in Ch. 12, Mars might well be habitable today if it was able to recycle its atmospheric CO_2.

These estimates for the boundaries of the HZ were calculated for cloud-free atmospheres and fully saturated tropospheres and may thus be too conservative. In Leconte *et al.* (2013), the unsaturated troposphere is a stabilizing influence, but cloud feedback is *positive*, so these effects tend to cancel. Cloud feedback is notoriously difficult to calculate, and other models predict that it is *negative* near the inner edge of the HZ (Wolf and Toon, 2014). Thus, Wolf and Toon predict that the inner edge is at 0.93 AU, or even closer, although their model does not account for water vapor as a major constituent near the inner edge of the HZ, unlike Leconte *et al.* (2013). The outer edge of the HZ could be much farther out if there are other greenhouse gases, e.g., H_2, that provide additional warming, which we discuss further below.

The uncertainty on the outer limit of the HZ may not matter too much because even the conservative estimate for the HZ width is fairly broad. If the outer edge is near 1.7 AU and the inner edge is near 0.95 AU, then the width of the HZ is ~0.75 AU. Our own Solar System contains four terrestrial planets between 0.4 AU and 1.5 AU, and the mean spacing between them is ~0.35 AU. This suggests that two of them ought to be in the HZ, which is exactly what we observe, because both Earth and Mars are in it, according to Kopparapu *et al.* (2013). If rocky planets in other planetary systems are spaced as they are in our Solar System, then the chance that at least one of them will be in the habitable zone is reasonable. As discussed further below, our own Solar System is dynamically "packed," meaning that it contains as many dynamically stable planets as it can. Originally, there were many more, smaller planets, and they continued to collide with each other until they reached a configuration in which collisions became infrequent on Gyr timescales, producing a stable, packed system. But it remains to be seen if such packing holds true for other planetary systems.

We can also use climate modeling results to derive an estimate for the width of the 4.6-billion-year (b.y.) continuously habitable zone (CHZ) around the Sun. The Sun is brighter today than it has been in the past. Hence, the inner edge of the CHZ is the same as the inner edge of the modern HZ: 0.95 AU. The outer edge, though, must be closer in because at 4.6 Ga the Sun was only about 70% as bright as it is today. Hence, a planet would need to have been closer to the Sun by a factor of $0.7^{1/2} = 0.84$ in order to receive the same flux that it does at present. If the outer edge of the modern HZ is at 1.7 AU, then at 4.6 Ga it should have been at $1.7 \times 0.84 \cong 1.4$ AU. Our CHZ is thus relatively wide, ~0.4 AU, which may be compared with the value of 0.06 AU in Hart's model. Once again, this illustrates the importance of the carbonate–silicate cycle and the negative feedback it provides on atmospheric CO_2 and climate. Without this feedback, habitable planets might indeed be rare.

A caveat should be added here: The CO_2–climate feedback on Earth depends on the presence of exposed continents. A "water world" with little or no continental area would not be subject to this same feedback

[2] Whether or not a moist greenhouse atmosphere can occur is currently a matter of contention. A recent 3-D climate simulation by Wolf and Toon (2015) suggests that it will. Their stratosphere is not quite as cold as that of Leconte *et al.* (2013), and they are able to extend their calculations to higher surface temperatures. A 1-D study by Kasting *et al.* (2015) supports the Wolf and Toon result.

(Abbot *et al.*, 2012). As discussed in Sec. 11.4, CO_2 would likely be removed by weathering of the seafloor on such a planet. The dependence of this process on surface temperature is uncertain. Coogan and Gillis (2013) argue that carbonatized seafloor shows that it provided a stabilizing negative feedback on high atmospheric CO_2 concentration on the Late Mesozoic Earth.

Other authors have proposed modifications to the inner edge of the HZ. Abe *et al.* (2011) used a 3-D climate model to show that hot, rocky planets with small water endowments and low obliquities might remain habitable in their polar regions, because the lack of a large ocean would reduce the positive feedback of water vapor on climate. They called such planets "*Dune* planets" after Frank Herbert's eponymous novel, which describes a desert planet with small habitable regions near its poles. Their estimated inner edge is at 0.77 AU, or $S_{eff} \cong 1.7$. But the water on such *Dune* planets might combine chemically with the surface, so it is unclear if they would remain habitable (Kasting *et al.*, 2014). Similarly, Zsom *et al.* (2013) estimated 0.5 AU for the inner edge using 1-D calculations of low-relative-humidity planets. But surface energy balance suggests that liquid water would quickly evaporate under such conditions (Kasting *et al.*, 2014).

As mentioned earlier, the outer edge of the HZ can expand if greenhouse gases other than H_2O and CO_2 are considered. Of the various candidates, H_2 is very effective because of its broad infrared absorption spectrum (Sec. 2.5.6) and low condensation temperature. As discussed in Sec. 12.5.2.2, Ramirez *et al.* (2014a) invoked several percent of H_2 to produce a warm climate on early Mars, which would otherwise have resided outside of the conventional HZ. Other authors have considered much larger amounts of H_2. Following Pierrehumbert and Gaidos (2011), Seager (2013) suggested that the outer edge of the HZ around our Sun might be as far out as 10 AU for a 3-Earth-mass planet with a 40-bar H_2 atmosphere that was captured during accretion. Stevenson (1999) had demonstrated earlier that H_2-rich rocky planets could maintain liquid water on their surfaces even if they were wandering freely in interstellar space. In that case, geothermal heat is sufficient to keep a planet warm if its H_2-rich atmosphere is thick enough.

The debate about HZ boundaries has practical consequences. Kasting *et al.* (2014) argue that a space telescope to look for Earth-like planets should be designed using a conservative definition of the HZ. That's because the frequency of Earth-size planets around stars, $\eta_\oplus$ (said as "eta sub Earth" or "eta Earth"), depends on the assumed HZ width. If $\eta_\oplus$ is high, then fewer stars need to be searched and the telescope can be made smaller. If $\eta_\oplus$ is low, then a larger telescope is needed. (Note that there are varying definitions of $\eta_\oplus$ in the literature that have different choices for the meaning of "Earth-like" or size of the HZ, which we discuss in Sec. 15.2.3.) Because both *Dune*-like planets and H_2-rich super-Earths are speculative, and they may or may not exist in reality, wider HZ limits for these objects should probably *not* be used when designing an Earth-finding telescope, lest the instrument be undersized. Once such a telescope has been launched, however, then broadening one's definition of the HZ makes sense, as one would not want to overlook any potentially habitable planets.

15.1.4 Empirical Estimates of Habitable Zone Boundaries

Theoretical calculations, regardless of how sophisticated they might be, must always be viewed skeptically, as nature is often subtler than our imagination. This statement applies to estimates of habitable zone boundaries, as well. Fortunately, our Solar System provides some *empirical* estimates for these boundaries. As discussed in Ch. 13 and above, Venus appears to have lost its water through either a runaway or moist greenhouse effect. The young cratering age of Venus' surface ranges 0.35–1.5 Ga, based on uncertainties in modeling, with a typical value near ~0.7 Ga (Korycansky and Zahnle, 2005). Consequently, Venus had probably already lost its water prior to 0.7 Ga, at which time the Sun was about 6% dimmer than it is today, using eq. (11.3) (Gough, 1981). Venus' semi-major axis is 0.723 AU, so the effective solar flux at that distance is $S_{eff} = (1/0.723)^2 = 1.91$. Taking into account the lower solar luminosity at 0.7 Ga, the "recent Venus" empirical estimate for the HZ inner edge is $S_{eff} = 1.91 \cdot (0.94) = 1.8$, or 0.75 AU. Note that this is slightly higher than the *Dune*-planet limit of $S_{eff} = 1.7$ estimate by Abe *et al.* (2011).

At the other end of the habitability spectrum, Mars appears (to some, at least) to have been habitable back around 3.8 Ga during the heavy bombardment period (Sec. 12.4). Mars' semi-major axis is 1.52 AU, and the solar flux at that time was about 25% lower than today by eq. (11.3), yielding $S_{eff} = (1/1.52)^2 \cdot 0.75 = 0.32$, which corresponds to a distance of 1.76 AU. This is just slightly lower than the maximum greenhouse limit of $S_{eff} = 0.35$ from the Kopparapu *et al.* (2013) model, which is why warming early Mars is a challenge (see Sec. 12.5.2). Thus, the empirical and theoretical limits on the outer edge of the HZ are in fairly close agreement. Only if large

amounts of H_2 are present in a planet's atmosphere should it remain habitable beyond about 1.7 AU.

15.1.5 Habitable Zones Around Other Main Sequence Stars

The same types of climate calculations that are described above can be done for planets orbiting other types of stars. (The stellar classification scheme was described in Sec. 2.1.2.) The calculations change in two ways, however. Most obviously, the stellar flux is much higher around bright blue stars and much lower around dim red ones; hence, the habitable zone must move either out or in depending on the stellar type. But it is not just the star's luminosity that changes. The spectral distribution of its radiation changes as well, because of the change in the star's surface temperature. This is illustrated in Fig. 15.2 (a), which shows the distribution of radiation for a planet orbiting an F2 star, a G2 star (the Sun), and a K2 star. The radiation from an F star is relatively bluer, while that from a K star is redder. These color shifts affect the climate calculations because blue light is more easily reflected from a planet due to increased Rayleigh scattering, whereas red and near-infrared radiation is scattered less and is also partly absorbed by the planet's atmosphere. Ice is also more reflective at shorter wavelengths (Warren *et al.*, 2002). Hence, according to 1-D climate simulations, the HZ around a blue star is slightly closer in than one would expect based solely on its luminosity, while the HZ for a red star is slightly farther out. (See below for how these predictions can change in 3-D.)

A graphical summary of HZ calculations is shown in Fig. 15.3(a). Here, the horizontal axis represents the distance from the star, and the vertical axis is the star's mass, relative to that of our Sun. The habitable zone is the strip running from the lower left-hand part of the diagram to the upper right. Also shown are the eight planets of our Solar System. Because the HZ moves outwards with time at different rates for stars of different masses, one has to choose a particular time in the star's lifetime in order to make a plot like this. In Fig. 15.3(a), the HZ has been plotted at the time when each star first enters the main sequence.

Figure 15.3(a) shows various relationships. First, the habitable zone lies farther out for more massive stars and closer in for less massive ones. That is to be expected, as a planet must receive roughly the same amount of starlight as does Earth in order to be habitable. Second, Fig. 15.3 (a) demonstrates a point made earlier: The HZ in our own Solar System is relatively wide, compared to the spacing between the planets. The orbital distance is shown on a log scale because the planets in our own Solar System are spaced logarithmically. This is "geometric" spacing, because each planet's orbital distance is larger than that of its inner neighbor by an average factor of ~1.7. Of course, the asteroid belt occupies the gap between Mars and Jupiter, where it appears as if a planet should exist. This observation is nothing new. The relationship between planetary orbital distances was noticed a long time ago and is referred to as *Bode's Law* or the *Titius–Bode Law*.

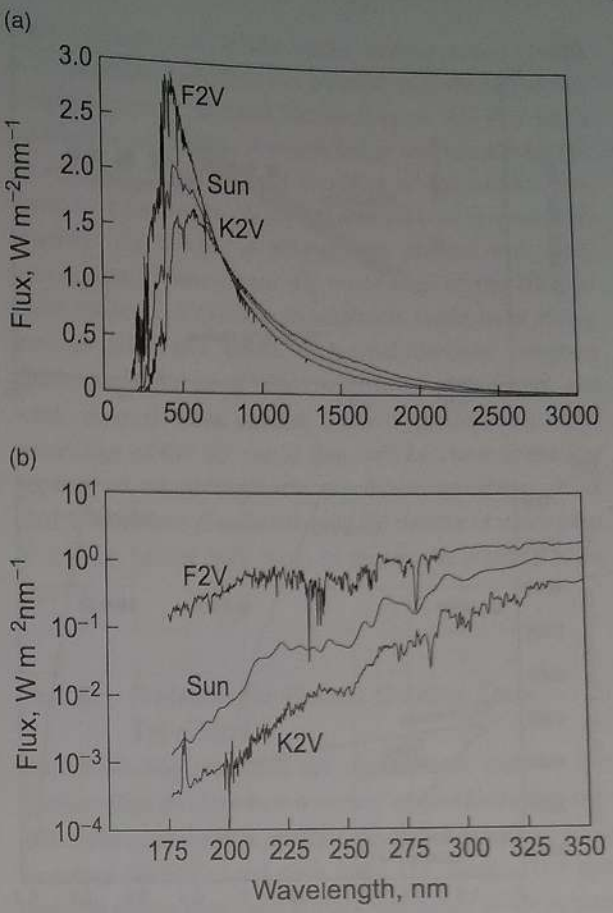

Figure 15.2 Incident stellar flux distribution for a planet orbiting an F2 star, a G2 star (the Sun), and a K2 star. The planet is assumed to receive the same total amount of sunlight as the present Earth. Panel (a) shows the entire wavelength spectrum; panel (b) shows the far ultraviolet portion of the spectrum. (From Kasting, (2010). Reproduced with permission of Princeton University Press; originally from Segura *et al.* (2003). Reproduced with permission from Mary Ann Liebert, Inc. Copyright 2003.)

Bode's Law has frequently been dismissed as nothing more than a simple empirical fit, but a Bode's Law spacing of planetary orbits can arise naturally from the right initial mass distribution within the solar nebula in order for orbits to be stable (Laskar, 2000). More specifically, separate power-law fits for the four inner and four

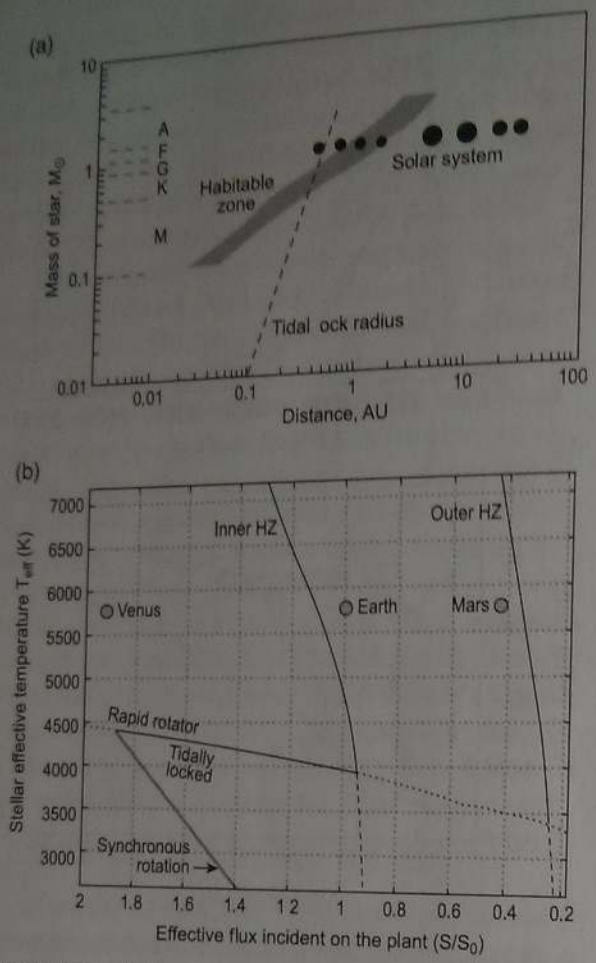

Figure 15.3 (a) Diagram illustrating the extent of the habitable zone around different types of stars. The vertical scale represents the mass of the star in units of the mass of our Sun, $M_\odot$, and the figure is drawn for the time when the stars have just reached the main sequence. The eight planets of our own Solar System are shown. The dashed line shows the orbital distance at, and below which, a planet experiences tidal locking of its rotation rate. (b) Diagram showing the habitable zone (HZ) in units of stellar flux S relative to today's solar flux, S_0. The runaway greenhouse "inner HZ" (IHZ) limit is shown for an Earth-mass planet, corresponding to 3-D climate model results (Leconte *et al.*, 2013). For cool stars ($T_{eff} < 4500$ K), the INZ jumps inward along the "synchronous rotation" line within the tidal locking limit (dotted curve) because 3-D models suggest a high planetary albedo develops on the dayside from thick water clouds at the subsolar point (Yang *et al.*, 2013). The outer HZ is the *maximum greenhouse* limit. (Adapted from Kopparapu *et al.* (2014). Reproduced with permission. Copyright 2014, American Astronomical Society.)

outer planets describe the orbits fairly accurately (Laskar, 2000). In any case, Bode's Law may have a physical basis. At least for the giant planets, it has been argued that you cannot pack planets more closely together without making their orbits unstable (Chambers *et al.*, 1996; Gladman, 1993). Whether or not *terrestrial* planets should be spaced geometrically is debated, however. Resonances with giant planets may influence terrestrial planet spacing more than mutual interactions between the terrestrial planets. This question should be answered when we are able to observe other inner planetary systems.

The issue of planetary spacing becomes important in considering stars that are significantly different from the Sun. As can be seen from Fig. 15.3(a), the HZ is roughly constant in width when plotted against the logarithm of orbital distance. In actual distance units, the HZ around an M star is quite narrow compared to the Sun's HZ, whereas the HZ around an F star is quite large. This caused Huang (1959) to conclude that the chance of finding habitable planets around M stars was small. But such a conclusion seems premature (or it may be correct for other reasons – see Sec. 15.1.5.2 below). When expressed in terms of log distance, there is just as much habitable space around an M star as there is around a G star. Whether one can populate this space with Earth-like planets remains to be seen.

Although Fig. 15.3(a) depicts the conventional habitable zone, it represents that zone only at the time a star enters the main sequence and it is only valid for rapidly rotating, 1-Earth-mass planets. A better way of defining habitable zones is in terms of stellar flux derived from 1-D or 3-D climate models. In particular, it makes much more sense to delineate the HZ boundaries in terms of stellar fluxes than in terms of planetary effective temperatures, which some astronomers have done. Calculating an effective temperature for a planet requires that one assume a value for its Bond albedo (eqs. (2.15) and (2.20)). But the albedo of a planet depends both on the composition of its atmosphere, which should be quite different near the two habitable zone boundaries, and on the stellar spectrum. It is not safe to assume that a planet's albedo is near Earth's value of ~0.3 (Table 2.2). Thus, defining the habitable zone in terms of stellar flux is better.

Kopparapu *et al.* (2013, 2014) have derived a useful parameterization from 1-D climate modeling results. This parameterization is now being continually updated, including using results from 3-D models.[3] A fit was performed as follows: Theoretical stellar spectra (non-blackbody) were used. Then, HZ boundaries for our Solar System were defined in terms of effective solar flux, $S_{eff} = S/S_0$. Correction factors for stars of other types are expressed in terms of their effective radiating temperature, T_*.

[3] Currently, a stellar flux HZ calculator is available here: depts.washington.edu/naivpl/content/hz-calculator

$$S_{eff} = S_{eff}^{0} + aT_* + bT_*^2 + cT_*^3 + dT_*^4 \quad (15.1)$$

Here, S_{eff}^{0} is the HZ boundary for our Sun. The "recent Venus" and "early Mars" limits are estimated by making the fluxes proportional to those for water loss and the maximum greenhouse, respectively. The corresponding orbital distances can then be obtained by applying the inverse square law, where L is the stellar luminosity.

$$r = 1\,\mathrm{AU} \cdot \left(\frac{L/L_S}{S_{eff}}\right)^{0.5} \quad (15.2)$$

Some results are obvious: The habitable zone moves inward for M stars and outward for F stars because of their vastly different luminosities (Fig. 15.3(a)). But there are also subtle differences. When expressed in terms of S_{eff}, the boundaries shift inward by an additional 10%–30% for F stars because their relatively bluer radiation is more effectively scattered (and more poorly absorbed) by a planet's atmosphere, thereby raising the planet's albedo. Just the opposite happens for M stars: the HZ boundaries shift outward because the planet's albedo is lowered (Fig. 15.3(b)).

Three-dimensional (3-D) climate models generate alternative boundaries. Leconte *et al.* (2013) considered rapidly rotating planets orbiting a Sun-like star, as discussed previously. Yang *et al.* (2013) have used a different 3-D climate model (the NCAR CCSM3) to look at synchronously rotating planets around late K and M stars. As discussed further below, these are planets on which tidal forces are sufficiently strong to slow the planet's rotation so that one side permanently faces the star as the Moon does to the Earth. Yang *et al.* find that the sunlit side of such planets should be permanently cloud-covered, greatly increasing the planet's albedo, and allowing $S_{eff} \cong 1.85$ for the inner edge of the HZ. Cloud cover is enhanced on slowly rotating planets where long daytime illumination and a weak Coriolis force promote strong convergence in the substellar area (Yang *et al.*, 2014a). So, even planets inside the *Dune* planet limit could remain habitable if they are tidally locked. These limits derived from 3-D models are incorporated in the version of the habitable zone shown in Fig. 15.3(b). The inner edge depends on planet mass, for reasons discussed in Sec. 13.4.2 (it will move inwards with bigger mass), and it depends as well on the planet's assumed rotation rate.

Tidal locking is also affected by atmospheric tides, which were described in Sec. 4.4.3. Some planets within the HZs of M and late K stars should *not* rotate synchronously because their rotation is affected by atmospheric thermal tides (Cunha *et al.*, 2015; Laskar and Correia, 2004; Leconte *et al.*, 2015). Stellar heating creates a bulge in atmospheres, which is not aligned with the sub-stellar point because of thermal inertia (Fig. 4.31). The star's gravitation acts on the pressure bulge and accelerates the atmosphere; then frictional coupling to the surface produces a torque on the planet that can prevent synchronous rotation. The torque is stronger on planets with thick atmospheres. Planets near the outer edge of the HZ that build up dense, CO_2-rich atmospheres could have strong thermal tides and rotate non-synchronously, whereas planets near the inner edge with thin atmospheres and weak thermal tides should rotate synchronously. The inner edge of the HZ could thus still be close to the star because of cloud feedback, as shown by Yang *et al.* (2013), but some planets towards the middle or outer parts of K- and M-star HZs may, in theory, be rotating more rapidly.

15.1.5.1 *Problems for Planets Orbiting Early-Type Stars*

Planets orbiting stars that are significantly more or less massive than the Sun face a variety of problems that may affect their habitability. Early-type stars have short main sequence lifetimes, and emit large UV fluxes. The first problem makes the O and B stars, and many of the A stars as well, not very interesting for astrobiology. The F stars, though, are a different matter. The main sequence lifetime of an F0 star is about 2 billion years. This is enough time for life to originate and evolve, as we know from our own planet's history. It may not be enough time to develop complex, multicellular life, based on our experience here on Earth (e.g., Catling (2005)), but that is a separate issue.

High stellar UV may appear detrimental to life, but it is not necessarily so. The problem is illustrated in Fig. 15.2(b). A planet at 1-AU equivalent distance around an F star would receive around four times as much UV radiation as does Earth at wavelengths less than 315 nm (see Table 15.1). This wavelength region includes biologically damaging UVB (290–320 nm) and UVC radiation (100–290 nm). Relative fluxes at 250 nm, where absorption by DNA peaks, are even higher – a factor of 10 or more. For a planet like the early Earth, that lacks an ozone layer, this could lead to higher rates of mutation for near-surface life. To astronomers, this appeared to be a serious problem (Sagan, 1973) but not to biologists, because organisms can avoid UV damage by forming mats, such as the stromatolites mentioned in Chapters 9 and 10 (Margulis *et al.*, 1976; Rambler and Margulis, 1980). If a mat-forming strategy worked on Earth, perhaps it would work on F-star planets as well.

Table 15.1 Relative values of UV flux (< 315 nm), ozone, and DNA dose rate on Earth-like planets around different stars

Stellar type	Incident UV flux	Ozone column depth	Surface UV flux	Relative dose rate*
G2 (Sun)	1	1	1	1
K2	0.26	0.79	0.43	0.5
F2	3.7	1.87	0.68	0.38

* Dose rate for DNA damage.
All values from Segura *et al.* (2003).

Furthermore, for F-star planets with O_2-rich atmospheres, a thick ozone layer should develop and provide UV shielding (Segura *et al.*, 2003). So, if life on such planets could make it through the anoxic–oxic transition, the stellar UV should cease to be an issue.

15.1.5.2 *Problems for Planets Orbiting Late-Type Stars*

Planets orbiting late-type stars face a completely different set of problems related to habitability. One issue that was recognized very early (Dole, 1964) and was discussed by Kasting *et al.* (1993b),[4] is the tidal locking problem (see the dashed curves in Figs. 15.3(a) and (b)). The HZ for an M star, and for a late K star as well, lies within the *tidal locking radius* of the star. This problem was discussed already in the previous section because it affects the HZ boundaries. But it can also affect a planet's climate. The potential danger with a tidally locked planet is that the planet's atmosphere and oceans could freeze out to form a giant ice cap on the dark side, thereby rendering the entire planet uninhabitable. Fortunately, this particular problem can be circumvented in several ways. One of these is illustrated by the planet Mercury in our own Solar System. As can be seen from Fig. 15.3(a), Mercury is within the Sun's tidal locking radius, yet it does not rotate synchronously. Instead, it spins three times on its axis for every two times it orbits the Sun because it is in a *spin–orbit resonance*. This probably happened because Mercury's mass distribution is slightly non-spherical as a consequence of violent, large impacts that occurred during its formation process, and because Mercury's orbit is highly eccentric ($e \cong 0.21$). Consequently, the energy associated with Mercury's rotation is lowest when its long axis (the one associated with the lowest moment of inertia) is aligned with its radius vector at perihelion. Other close-in planets in highly eccentric orbits may also avoid synchronous rotation in a similar manner.

Another way out of the tidal locking problem is atmospheric or oceanic advection of heat. If a synchronously rotating planet has an atmosphere with at least 30 mbar of CO_2 – about 100 times the amount in Earth's present atmosphere – 3-D climate simulations show that the atmosphere transports sufficient heat from the dayside to the nightside to prevent the atmosphere from freezing out (Joshi, 2003; Joshi *et al.*, 1997). High CO_2 concentrations facilitate such heat transfer by lengthening the time required to radiate heat off to space. Similarly, if a planet has a deep ocean, like Earth, then ocean currents can also carry heat from the dayside to the nightside (Edson *et al.*, 2011; Merlis and Schneider, 2011). Pierrehumbert (2011) also identifies an alternative *eyeball Earth* state in which the front side of the planet is warm while the back side is cold and frozen. So, from a climatic standpoint, M-stars should not be excluded as candidates for harboring habitable planets.

Several other potential problems for M-star planets exist, though, and these may be more serious than the tidal locking issue. One problem concerns the ability of planets to retain an atmosphere. M stars have much more flare activity than does our Sun, and they have correspondingly enhanced stellar winds. Furthermore, tidally locked planets around all but the least massive M stars should be rotating fairly slowly, 10–100 days, and hence might be unable to generate strong magnetic fields. In such a situation, a planet's atmosphere could conceivably be stripped away by the intense stellar wind (Lammer *et al.*, 2007). Other problems with M-star planets could result from differences in the nebular environment in which they form (Lissauer, 2007). The habitable zone around an M star is very close in; hence, the amount of material available to be swept up by a growing planet is relatively small, and so the planets that form there may be significantly less massive than Earth. Furthermore,

[4] Dobrovolskis (2009, p. 9) has pointed out that the tidal despinning rate from equation (8) of Peale (1977) is a factor of two too fast, while the corresponding timescale from his equation (9) is a factor of two too short. Thus, the tidal locking distance should be closer by a factor of $0.5^{1/6} \sim 0.8909$, and the coefficient in equation (10) of Kasting *et al.* (1993b) should be 0.024 instead of 0.027.

because the orbital times at these distances are short, accretion of planets should occur rapidly, giving the nebula less time to cool. This means that terrestrial planets, as they form, would be further removed from the *snow line* – the distance where icy planetesimals can form (Sec. 6.2.1). Delivery of water could also be inhibited by the apparent dearth of Jovian-size planets orbiting M stars (Bonfils *et al.*, 2013), which should result in reduced radial mixing of water-rich planetesimals from outer planetary systems. The tight orbits within M-star habitable zones also mean that the planetesimal relative velocities would have been high, and so impacts would have been even more violent than those that formed the Earth. Such energetic impacts may have swept away, or eroded, more atmosphere than they delivered (Catling and Zahnle, 2013; Melosh and Vickery, 1989).

The biggest problem for M-star habitability, though, is probably that the pre-main sequence luminosity of M stars is high and can cause a runaway greenhouse on M-star planets (Luger and Barnes, 2015; Ramirez and Kaltenegger, 2014). The luminosity, of course, comes from gravitational energy released by the collapsing protostellar cloud. The Sun was about twice as bright during its pre-main sequence phase as it was later on (Baraffe *et al.*, 1998; Baraffe *et al.*, 2002) (illustrated in Fig. 6.5). The corresponding ratio for an M star is 20–180, with the higher values applying to the later (smaller) M stars. The formation time for an M star is also slow – a few hundred million years, as compared with less than 50 million years for the Sun (Luger and Barnes, 2015). At the same time, the expected accretion time for a planet within the HZ of an M star is shorter than for our Sun because the orbital periods are shorter. Earth is thought to have accreted in 10–100 million years, whereas the corresponding time for a planet accreting in situ in the HZ of an M star is ~a few million years (Lissauer, 2007). Thus, whereas the latter part of Earth's accretion occurred *after* the Sun had reached the main sequence, the same is not true for M-star planets. A planet forming *in situ* within the HZ of an M star would likely suffer a runaway greenhouse and lose its water at that time. If this water was not later replenished, then the planet might never be habitable. The best bet for creating a habitable M-star planet might thus be to form the planet well outside of the HZ, beyond the ice-line of the stellar nebula where it would accrete lots of volatiles, then migrate it into the HZ (Luger *et al.*, 2015). If the planet starts with an overabundance of water, perhaps it could lose a large percentage of it and still retain enough water to be habitable. Alternatively, an M-star *Dune* planet might avoid a runaway.

Although none of the above points are necessarily show-stoppers, M stars appear to have more issues working against their habitability than planets around F, G, and early K stars. Eventually, exoplanet data will help determine the validity of this idea. Certainly, over the past 50 years, the suggestion that one should focus primarily on F, G, and K stars has been made consistently (Dole, 1964; Huang, 1960; Kasting *et al.*, 1993b). Fortunately, ~20% of stars in the solar neighborhood have spectral classifications between F0 and K5, so this limitation is not too restrictive. On the other hand, most of the closest stars are M stars, and the small sizes and even smaller luminosities of these stars also make it easier to search for and characterize Earth-like planets using transit observations, as discussed further below. So, for the next few years, M-star exoplanets are likely to remain high-priority targets.

15.1.5.3 *Limit Cycling in the Outer Habitable Zone*

While planets in the outer parts of a HZ may not be able to maintain stable, warm climates, they may instead oscillate between short-lived warm periods and longer periods of global glaciation (Haqq-Misra *et al.*, 2016; Kadoya and Tajika, 2014; Menou, 2015). This behavior is referred to as *limit cycling*. On such planets, CO_2 is consumed by weathering during the warm periods faster than it can be resupplied by volcanism. The resulting, time-dependent behavior depends on the planet's volcanic outgassing rate of CO_2, the availability of water to form snow and ice, and on the type of host star. Planets around F and early-G stars are more prone to limit cycling than are planets around K and M stars because the stellar radiation is shifted towards the blue, where the albedo of water ice is higher (Joshi and Haberle, 2012; Shields *et al.*, 2013; Warren *et al.*, 2002). This type of climate behavior may or may not pose problems for simple (unicellular) life, as we know that simple life was able to make it through repeated Snowball Earth episodes on Earth (Sec. 11.10.5). But complex, multicellular life – in particular, animal life on the continents – would be challenged (Haqq-Misra *et al.*, 2016). Humans are a subset of such life, of course, and so this may imply that other forms of land-based, intelligent life are limited to the inner region of the HZ around hotter, bluer stars. Earth is in the inner part of the Sun's HZ (Fig. 15.3), and perhaps we owe our own existence partly to this fortuitous circumstance.

15.1.6 Other Concepts of the Habitable Zone

There is obviously more to habitability than the habitable zone because planetary properties matter. An extension of the habitable zone concept is to take the properties of a

planet into account (Franck *et al.*, 2000a, b; Franck *et al.*, 1999). For example, different continental growth rates lead to differences in CO_2 uptake by silicate weathering, and so the greenhouse effect varies in a time-dependent manner in the calculations of Franck *et al.* Thus, the width of the HZ varies with time for different assumed planetary properties. Similarly, the length of time that a planet can remain habitable depends on its distance from its parent star. For an Earth-like planet orbiting a star like the Sun, Franck *et al.* calculate an optimal orbital distance of 1.08 AU. To use these types of calculations, it is necessary to have lots of information about the planet being considered. But if we already know that much about the planet, then we shouldn't need to predict whether or not it is in the HZ. So, it is not clear that these extensions of the HZ concept are practically useful.

Another extension of the habitable zone concept is for planets with significant tidal heating (Barnes *et al.*, 2009; Barnes *et al.*, 2013; Barnes *et al.*, 2010; Jackson *et al.*, 2008a; Jackson *et al.*, 2008b; Jackson *et al.*, 2010). The conventional (insolation-determined) HZ for late-K and M stars lies within their tidal locking radius. Planets that are beyond this orbital distance but for which tidal drag is strong have an internal heat source that can add to the stellar insolation and widen the HZ. Such planets must be on eccentric orbits to generate this tidal heating, and their eccentricities must be forced by other planets within the system; otherwise, their orbits would rapidly circularize. This factor should be included when studying planets orbiting such late-type stars.

15.1.7 The Galactic Habitable Zone

There may be a region around the center of our galaxy that is optimal for finding habitable planetary systems, which is analogous to a stellar habitable zone. The name given to this concept is the *galactic habitable zone*, or GHZ (Gonzalez *et al.*, 2001; Ward, 2000). The idea can be elaborated to include time as well as space (Lineweaver *et al.*, 2004), but, in essence, the concept is that not all stars in the Milky Way are equally likely to harbor habitable planets. Planets orbiting stars that are too close to the center of the galaxy could have their orbits perturbed by close stellar encounters, and they are also more likely to experience catastrophic events such as nearby supernovae and gamma ray bursts. Stars too far out towards the rim of the Milky Way spiral galaxy are less metal-rich than the Sun and may be less likely to be accompanied by rocky planets. (Recall that a "metal" to an astronomer is any element heavier than hydrogen and helium.) Similarly, stars that form too early in the history of the galaxy are likely to be metal-poor, because not enough hydrogen and helium will have been reprocessed through stars to form the heavy elements. In contrast, our Sun formed at 4.6 Ga and is located 27 200±1100 light years from the galactic center (Gillessen *et al.*, 2009) – about half of the radius of the Milky Way, whose stellar disc is considered to be 100 000–130 000 light years in diameter (e.g., Schneider, 2014).

The spatial GHZ concept is challenged by recent simulations showing that stars migrate within the galaxy as a result of scattering off of spiral arms (Roskar *et al.*, 2008). Thus, habitable planets could theoretically exist anywhere. In any case, this concept is more relevant to the distant future of galactic exploration than it is to the near-term search for habitable planets. For practical reasons, the exoplanet systems that we hope to study within the foreseeable future are all located relatively nearby, within ~50 light years. The host stars appear to have roughly the same metal content as the Sun, provided that one restricts the comparison to F–G–K stars (Boone *et al.*, 2006). (Some early work suggested otherwise, but it included M stars, which tend to be older on average, and hence less metal-rich (Gonzalez, 1999).) Nearby stars are also subject to the same background level of supernovae and gamma ray bursts that the Solar System experiences. So, even if the potentially habitable area of the galaxy is indeed limited, this should have little effect on the probability of finding habitable planets in our stellar neighborhood.

15.2 Finding Planets Around Other Stars

Before examining exoplanets to see if they harbor life, we must find suitable exoplanets. Here we give an overview of exoplanet detection methods. More technical detail is given in the review by Wright and Gaudi (2013).

Astronomers have been searching for exoplanets for a long time. In the 1960s to 1980s, Peter van de Kamp thought he had detected a wobble in the position of Barnard's star as it moved along its track, i.e., deviations of the star's *proper motion*, which is movement relative to the Sun. Barnard's star is an M star six light years away, which is the closest individual star after the triple α-Centauri system. At the end of a series of papers, van de Kamp concluded that the perturbations were the effect of two gas giants orbiting the star with periods of 12 and 20 years (van de Kamp, 1969, 1975; van de Kamp, 1982). The data were photographic plates recorded between 1938 and 1981. However, subsequently systematic errors were found, including shifts when the telescope lens was adjusted. Further observations, using a variety of

methods, show no evidence for large planets orbiting Barnard's star (Benedict *et al.*, 1999; Choi *et al.*, 2013; Dieterich *et al.*, 2012; Kurster *et al.*, 2003).

15.2.1 The Astrometric Method

Van de Kamp's technique was an example of the *astrometric method*. This finds exoplanets by looking for perturbations in the motion of a star, which occurs because the star moves around a common center of mass with the exoplanets. Although the modern astrometric method uses CCDs (charge-coupled devices) with high spatial resolution, ground-based astrometry still has limited accuracy – generally not enough to find planets.[5] The problem is that the Earth's atmosphere *scintillates*, or twinkles, because of the disruption of the wavefronts coming from stars by atmospheric turbulence. Some of this interference can be removed by *adaptive optics*, which are techniques to sense and correct the distortions caused by turbulence. However, it remains challenging to measure stellar positions accurately enough to find exoplanets.

An obvious solution to atmospheric effects is to do astrometry from space. ESA's Gaia Mission, which has operated since 2014, measures stellar positions to an accuracy enough to detect Jovian plantes orbiting Sun-like stars. NASA's proposed Space Interferometry Mission, SIM (later "SIM Lite"), would have done even better. The estimated error in a single angular measurement from SIM would have been ~1 μas (microarcsecond) (Unwin *et al.*, 2008). Earth's motion around the Sun causes the Sun's position to wobble by ~0.3 μas, as viewed from a distance of 10 pc. This follows from the definition of a parsec, and from the ratio of Earth's mass to the Sun's mass, $M_\oplus/M_\odot \simeq 3\times10^{-6}$. One parsec is the distance at which the Earth's orbital radius, 1 AU, subtends an angle of 1 arcsec. Viewed from 10 pc, Earth's orbital radius would subtend one-tenth of this angle, or 0.1 arcsec. If we ignore the effects of the other planets, Earth and the Sun would orbit at distances $r_\oplus$ and $r_\odot$ from a common barycenter defined by

$$M_\oplus r_\oplus = M_\odot r_\odot \tag{15.3}$$

As Earth orbits the Sun, the Sun moves by a distance $r_\odot = r_\oplus \cdot (M_\oplus/M_\odot) = 3\times10^{-6}$ AU, which would subtend an angle of 0.3 μas, as viewed from 10 pc. Random errors decrease as $1/\sqrt{N}$, where N is the number of measurements. Hence, within ~100 measurements, SIM Lite should have been capable of detecting an Earth-like planet. Unfortunately, SIM Lite was cancelled in 2010. We will have to find the Earth-like planets using some other method.

[5] By 2013, the *Extrasolar Planets Encyclopedia* website reported just one planet candidate identified by PHASES – The Palomar High-precision Astrometric Search for Exoplanet Systems.

15.2.2 The Radial Velocity Method

The planet-finding technique that has proved most successful from ground-based telescopes is the *radial velocity* (*RV*) or *Doppler method* (e.g., Mayor *et al.*, 2014). Instead of measuring the position of the star on the sky, one measures its motion back and forth in the line of sight by looking at the Doppler shifts of multiple spectral absorption lines. The Doppler shift, $\Delta\lambda$, of a line centered at wavelength λ is given by $\Delta\lambda/\lambda \cong v_S/c$, where v_S is the star's radial velocity and c is the speed of light. (See Beaugé *et al.* (2008) for a technical review of RV theory.)

If we consider the Earth moving around the Sun, ignore its eccentricity (which is small), and ignore the other planets, as we did above, conservation of momentum ensures that

$$M_\oplus v_\oplus = M_\odot v_\odot \tag{15.4}$$

The average velocity of the Earth around the Sun, $v_\oplus$, is about 30 km s^{-1}, or 3×10^4 m s^{-1}; hence, the Sun's maximum radial velocity, assuming an edge-on view of the system, is: $v_\odot = v_\oplus \cdot (M_\oplus/M_\odot) \cong 0.1$ m s^{-1}, or 10 cm s^{-1}. Current RV precision around quiet, bright, Sun-like stars is ~1 m s^{-1}. So, with sufficient measurements it is possible, in principle, to find an Earth-mass planet around a Sun-like star using this technique. But, in practice, it currently takes a very bright star (e.g., α-Centauri), along with lots of time on a very big telescope.

The RV technique works best for massive planets in tight orbits around their parent stars. (By contrast, astrometry works best for massive planets at large distances, because their lever arm is longer.) As such, the first exoplanets to be discovered by the RV method were so-called *hot Jupiters* – which are loosely defined as Jovian gas giant planets with orbital radii less than roughly 0.1 AU (e.g. Perryman, 2014, p. 103).[6] The first such planet to be discovered was 51 Peg b (Mayor and Queloz, 1995). The notation follows that of binary star systems, where the companion (in this case a planet) is designated by the star's name, followed by "b". 51 Peg b has a minimum

[6] Exoplanet descriptions such as "hot Jupiter," "very hot Jupiter," "mini-Neptune," "sub-Neptune," "Super-Earth," and so on, currently have different definitions in the literature, which is still evolving. In this chapter, we use definitions that are close to the current consensus.

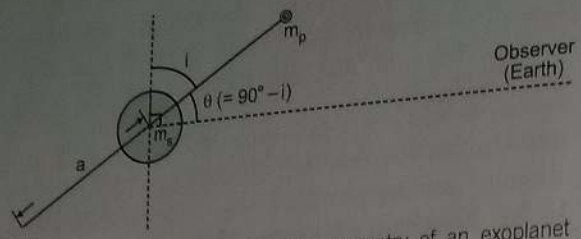

Figure 15.4 Diagram illustrating the geometry of an exoplanet observation. Here, i is the inclination of the planet's orbit with respect to the plane of the sky, and θ (= 90° – i) is the angle of the planet's orbit with respect to the observer on Earth. The star's mass is m_s and the planets mass is m_p. The planet orbits the star at a semi-major axis of a. (From Kasting (2010). Reproduced with permission of Princeton University Press. Copyright 2012.)

mass of 0.44 Jupiter masses and orbits its (Sun-like) star in just over 4 days.

With the RV technique, one cannot measure the planet's mass, m_p, directly because one does not generally know the inclination, i, of the planet's orbit with respect to the plane of the sky (see Fig. 15.4). So, the reported mass is actually $m_p \sin i$. The mass of an ensemble of such RV planets can be estimated statistically from

$$\langle \sin i \rangle = \frac{\int_0^{\pi/2} \sin i \cdot \sin i di}{\int_0^{\pi/2} \sin i di} = \frac{\frac{1}{2}\int_0^{\pi/2} (1 - \cos 2i) di}{1} = \frac{\pi}{4} \cong 0.785 \tag{15.5}$$

Thus, the average exoplanet detected by the RV method is heavier than its measured mass by a factor of 1/0.785 ≅ 1.27. However, applying this factor to any individual planet is risky.

The RV method has been astounding successful, with almost 700 detected exoplanets by late 2016. As the time series lengthens, planets with longer orbital periods emerge from the data. New instruments are also aiming for precision that would allow detection of Earth-mass planets, in principle. For example, the Eschelle SPectrograph for Rocky Exoplanet and Stable Spectroscopic Observations (ESPRESSO) on the European Southern Observatory's Very Large Telescope array aims for a precision of RV variations of a few centimeters per second, compared to Earth's 9 cm s^{-1} effect on the Sun (Pepe *et al.*, 2014). The practical limit of RV measurements may depend on inherent noise levels of the stars themselves. Stars like the Sun have vertical motions on their surfaces of the order of several meters per second, and they also have sunspots that rotate with the star, along with irregular flare activity. So, even with essentially perfect spectral calibration techniques (laser combs) that are currently being developed, there may well be a mass limit below which planets cannot be detected.

The most famous RV detection is a 1.3 (minimum) Earth mass planet in the HZ of the nearest star, Proxima Centauri (Anglada-Escude *et al.*, 2016). This planet could conceivably be imaged using 30–40 m ground-based telescopes.

15.2.3 The Transit Method and Results from NASA's *Kepler* Mission

Another exoplanet detection technique that has been extremely successful is the *transit method* (Cameron, 2016). When a planet *transits* (passes in front) of its parent star, as seen from Earth, it blocks out some starlight, and can be detected. To give an example from our own Solar System, the diameters of the Sun, Jupiter, and Earth are approximately in the ratio 100:10:1. The projected area of a planet is just πr_p^2, where r_p is the planet's radius. Hence, if Jupiter or Earth were to pass in front of the Sun, as viewed from a great distance, the Sun's brightness would diminish by 1% and 0.01%, respectively. A 1% change in stellar brightness can be readily detected using a ground-based telescope. Indeed, the first such planet detected, the hot Jupiter planet HD209458b, was found from a 1.6% dip in starlight using a small "backyard" telescope equipped with an accurate CCD-based photometer (Charbonneau *et al.*, 2000). The light curve from his measurements is reproduced in Fig. 15.5(a), along with later, precise measurements from the *Hubble Space Telescope* (Brown *et al.*, 2001) (Fig. 15.5(b)).

Charbonneau realized that he could measure such a transit because the planet itself had already been detected by radial velocity searches. It was one of about ten such hot Jupiter planets that were known at the time. The probability that such a planet will transit is approximately equal to the radius of the star divided by the semi-major axis of the planet's orbit. This probability can be computed by integrating the inclination angle, i, weighted by $\sin i$, over the range of angles for which a transit occurs. Equivalently, if we let $\theta = 90° - i$ (see Fig. 15.6) and integrate over angles from face-on, $\theta = 0$, to some actual angle $\theta = \theta_0$, then the probability, P, that a transit will occur is given by this integral divided by an integral over all angles to $\theta = 90°$, i.e.

$$P = \frac{\int_0^{\theta_0} \cos\theta d\theta}{\int_0^{\pi/2} \cos\theta d\theta} = \sin\theta_0 = \frac{R_S}{a} \tag{15.6}$$

Here, R_S is the radius of the star and a is the semi-major axis of the planet. The calculation is more complicated if

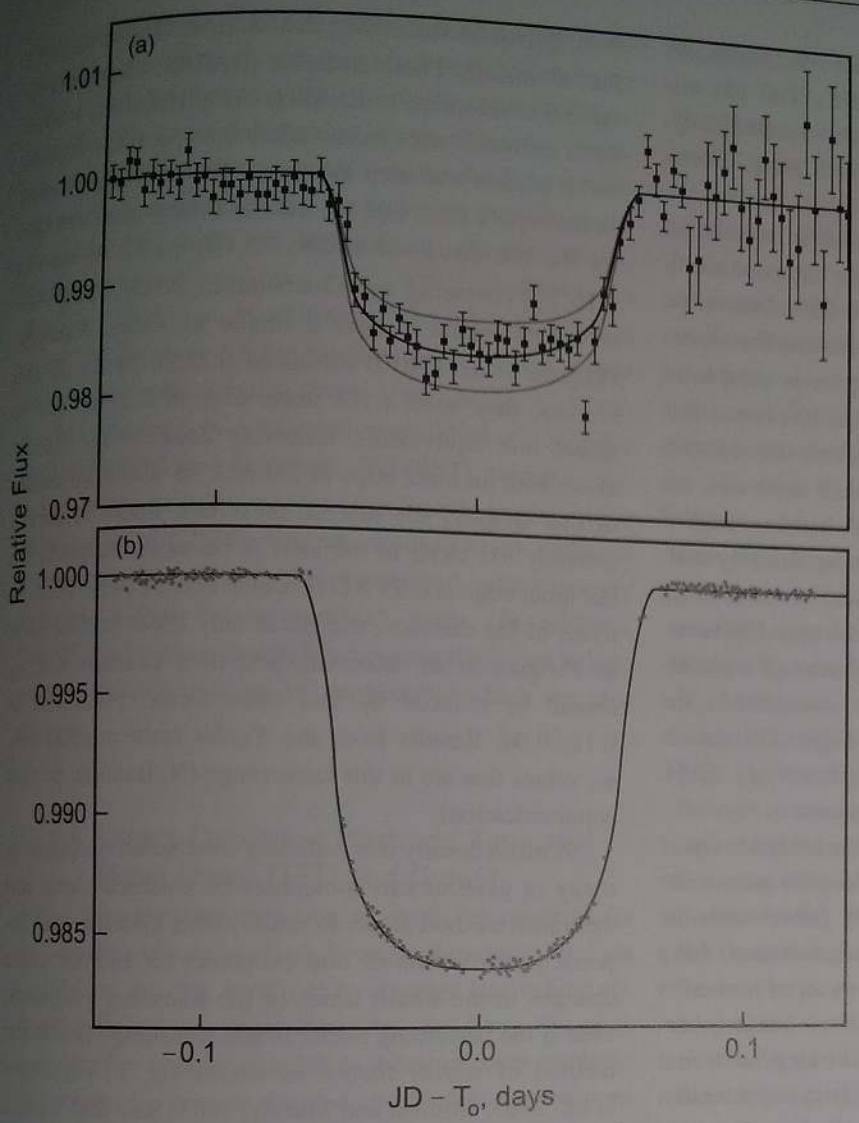

Figure 15.5 (a) Light curve for the star HD 209458. (From Charbonneau *et al.*, 2000.) (b) The same light curve measured a few months later using the *Hubble Space Telescope*. (From Brown *et al.*, 2001). (From Kasting (2010). Reproduced with permission of Princeton University Press. Copyright 2012.)

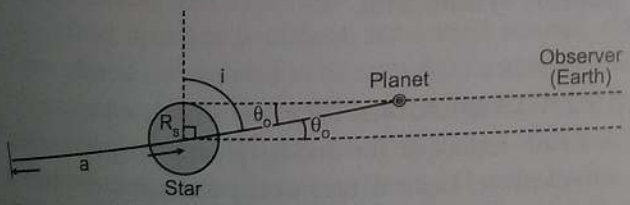

Figure 15.6 Diagram illustrating the geometry of a transiting planet relative to an observer on Earth. Here a is the semi-major axis of the planet's orbit, i is the orbit's inclination, and R_s is the stellar radius. Here θ (= 90° − i) is the angle of the planet's orbit with respect to the observer on Earth. (From Kasting (2010). Reproduced with permission of Princeton University Press. Copyright 2012.)

one allows for eccentric orbits, but the results do not change much unless the eccentricity is very high. Now, the radius of the Sun is about 7×10^5 km, while the semi-major axis of a typical hot Jupiter is about 0.05 AU $\cong$ 0.05 · (1.5×10^8 km) = 7.5×10^6 km. Hence, the probability that a hot Jupiter will transit is ~10%. So, Charbonneau knew that if he observed 10 hot Jupiters each for long enough to see a transit, the odds suggested that he would see one.

Seeing Earth-like planets by the transit method from ground-based telescopes is more difficult. An Earth-like planet passing in front of a Sun-like star would only block one part in 10^4 of the star's light. Because of scintillation, this signal is too small to detect. The chances of seeing an Earth-like planet transit are also much smaller. For Earth around the Sun, the probability is $R_S/(1\text{ AU}) \cong 7\times10^5$ km/ 1.5×10^8 km = 5×10^{-3}, or 0.5%. Hence, one would need to observe ~200 such systems in order to find one transiting Earth. Consequently, ground-based surveys, such as *MEarth*, have focused on the easier task of searching for

short-orbit planets around nearby M-dwarfs, which are smaller, dimmer stars (Berta *et al.*, 2012). This has discovered excellent candidates for follow-up detailed study, such as a $1.2R_\oplus$ Super-Venus 12 parsecs away (Berta-Thompson *et al.*, 2015).

Fortunately, there is a way to observe transits of Earth-size planets. NASA's *Kepler* space telescope monitored the brightness of more than 150 000 stars in the region of the Cygnus and Lyra constellations for almost four years, starting in June, 2009. *Kepler* is capable of measuring stellar brightness to one part in 10^5; hence, it is able to find Earth-sized planets, particularly around stars that are smaller than the Sun. (G stars, it turns out, are more difficult to study than anticipated, because most of them are noisier (have more spots) than the Sun.) By mid-2015, the analysis of *Kepler* data had revealed over 3900 unconfirmed planet candidates and over 1000 confirmed planets orbiting more than 2000 stars. If a planetary size range of $0.75–2.5R_\oplus$ is considered, the occurrence rate of planets per star for *Kepler* GK stars is 0.77 with an range of 0.3–1.9 (Christopher *et al.*, 2015). Basically, GK stars on average have planets.

Some *Kepler* planets orbit within the habitable zones of their parent stars, allowing statistical inferences to be extrapolated to the rest of the galaxy (albeit with the caution of residual sample bias even after attempts have been made to correct for bias). A parameter of interest is $\eta_\oplus$, which has slightly varying definitions but is generally described as *the frequency of Earth-size planets in a circumstellar habitable zone*. One needs to pay attention to what authors mean by "Earth-size" and which boundaries they choose for the HZ in order to compare different estimates of $\eta_\oplus$. For example, assumptions of "Earth-size" include 0.5–1.4 $R_\oplus$ (Kopparapu *et al.*, 2013), 0.5–2 $R_\oplus$ (Silburt *et al.*, 2015) or 1–2 $R_\oplus$ (Petigura *et al.*, 2013). Observations where radius and mass are both available, so that mean density can be deduced, suggest that the upper size limit for rocky planets is $1.5–1.75R_\oplus$ (Lopez and Fortney, 2014; Rogers, 2015; Weiss and Marcy, 2014). Meanwhile, 0.5 $R_\oplus$ (the size of Mars) is a reasonable lower limit for an Earth-size planet, as planets much smaller than that would have trouble holding onto their atmospheres. So, to be conservative, we suggest 0.5–1.5 $R_\oplus$ as the appropriate size range to define $\eta_\oplus$. The range of spectral types of star is also another parameter because one defines $\eta_\oplus$ for a particular stellar class.

Interestingly, $\eta_\oplus$ is perhaps not greatly different for G, K, or M stars. Early estimates indicated that $\eta_\oplus$ for M stars is between 0.4 and 0.6 (Gaidos, 2013; Kopparapu *et al.*, 2013). Bonfils *et al.* (2013) also estimated $\eta_\oplus = 0.41^{+0.54}_{-0.13}$ for M stars using data from the High Accuracy Radial velocity Planet Searcher (HARPS) instrument on the 3.6 m telescope at La Silla, Chile. However, both of these estimates were made before the upper size limit on rocky planets had been estimated. When 1.5 $R_\oplus$ is taken as that upper limit and we use the conservative estimate for the HZ discussed in Sec. 15.1.3, $\eta_\oplus$ for M stars is $0.16^{+0.17}_{-0.07}$ (Dressing and Charbonneau, 2015). More massive stars appear to have a similar $\eta_\oplus$ value. Although Petigura *et al.* (2013) estimated 0.22 ± 0.08 for K and G stars, they used a HZ inner edge of 0.5 AU (or its stellar flux equivalent), following Zsom *et al.* (2013), along with an outer edge of 2.0 AU. As discussed earlier in this chapter, the first of these two limits is almost certainly too close to the star. A better estimate for the HZ inner edge is 0.95 AU (Leconte *et al.*, 2013). Thus, in terms of log distance, the HZ is only about half as wide as Petigura *et al.* assumed, and their estimate for $\eta_\oplus$ should be reduced by this same factor, putting it at 0.11 ± 0.04. Results from the *Kepler* team suggest GK $\eta_\oplus$ values that are in this same range (N. Batalha, private communication).

A mean density that indicates whether an exoplanet is rocky or gaseous can sometimes be obtained using just the transit method alone. In multi-planet systems, gravitational perturbations of one exoplanet on another cause changes in the transit times of the transiting exoplanet, even if the perturbing planet is non-transiting. This is the method of *transit timing variations* (or TTVs) (Agol *et al.*, 2005; Holman and Murray, 2005; Sam and Yoram, 2014) (covered as Ch. 7 of Haswell (2010)). Such mean density information is not only critical for planet mass–radius relationships but also our understanding of how planetary systems form.

15.2.4 Gravitational Microlensing

A fourth technique for finding planets is *gravitational microlensing*. Light is bent as it passes around a star, a planet, or a galaxy. Equivalently, spacetime is curved by massive objects, and light follows a straight line in curved space. This general relativistic effect is easiest to discern over large distances; hence, *gravitational lensing* has been used for some time to study distant galaxies and clusters of galaxies. Only more recently has this technique been applied to stars and planets.

Microlensing is a regime of gravitational lensing where the additional light bent forward to an observer brightens the source but without multiple images or distortion of the source image that happens in the two other cases of lensing, called *strong* and *weak* lensing.

respectively. In microlensing, when a planet orbits a lens star, the background star brightens more than once because of the lens star and its planet. The technique provides no direct information about planetary atmospheres or nearby planets, so we will not dwell on it here. Space-based microlensing can, however, provide statistics, and potentially an independent estimate of $\eta_\oplus$. Surveys suggest every star has $1.6^{+0.72}_{-0.89}$ planets in the $5M_\oplus$ to $10M_J$ range orbiting at 0.5-10 AU (Cassan *et al.*, 2012); also ~1.8 Jupiter-mass planets per star exist towards the galactic bulge (Sumi *et al.*, 2011).

Microlensing is a planned component of NASA's Wide Field Infrared Telescope (WFIRST), which is currently a top priority mission for launch in the 2020s (Spergel *et al.*, 2015). So, we should obtain statistical information about Earth-like planets and outer planetary systems within the next 10–15 years. In addition, WFIRST's baseline design has a coronograph (see below) to allow direct imaging of exoplanets, which we now discuss.

15.2.5 Direct Detection Methods: *Terrestrial Planet Finder* (TPF) and *Darwin*

All of the exoplanet detection methods discussed so far are termed *indirect methods* because they look for light emitted by the star itself (or by another star behind it). Astrobiologists are most interested in *direct detection methods* for finding exoplanets because such methods can potentially provide detailed spectra that may be used to characterize their atmospheres and surfaces. (The transit method can yield planetary spectra, as well, as discussed in Sec. 15.3.1.) As its name suggests, direct detection means looking directly for light from the planet itself.

Direct detection is difficult for several reasons. For one, planets are assumed to be embedded within a cloud of exozodiacal dust, analogous to the zodiacal dust in our own Solar System, which comes from the collisions of asteroids and comets. Resolving the planet in the midst of this background requires a large telescope (> 4 m for a dust cloud similar to our own). Furthermore, the planet being observed is very close to a much brighter star, and so one needs to be able to separate the planet's image from that of the star and to block out any diffracted light from the star that might otherwise obscure the planet. This can also place lower limits on telescope size, as discussed further below.

Direct detection has already been done using both large telescopes on the ground and the *Hubble Space Telescope*. So far the imaged planets are all massive planets, Jupiter-sized or larger, with semi-major axes ranging from tens to hundreds of AU (e.g. Delorme *et al.*, 2013; Kuzuhara *et al.*, 2013; Lagrange *et al.*, 2010; Macintosh *et al.*, 2015; Rameau *et al.*, 2013). To see planets, the telescope must have a *coronagraph*, an instrument that can block out the light from the star and retain the light from the planets around it. The term "coronagraph" comes from attachments that were originally used to block out the light from the Sun in order to expose the tenuous corona surrounding it. Alternatively, the telescope could be flown in combination with an external occulter, or *starshade*, that blocks out the star's light before it enters the telescope (see below).

Finding Earth-like planets by direct imaging will require either large (≥4 m) telescopes in space or extremely large telescopes on the ground. Several 30 m-plus ground-based telescopes are being considered, however (Kasper *et al.*, 2010; Matsuo and Tamura, 2010). These may or may not be able to image planets like Earth, depending on how well they are able to remove the effects of the atmosphere. To search for spectral biosignatures (e.g., O_2, O_3, CH_4, N_2O) from the ground one must look through an atmosphere that contains all of these gases and high dispersion spectroscopy has been proposed for detecting biogenic O_2 on Earth twins around M stars (Snellen *et al.*, 2013; Rodler and Lopez-Morales, 2014). But the best hope for characterizing Earth-like planets and for looking for signs of life comes from space-based telescopes.

NASA and ESA (the European Space Agency) have been developing space mission concepts to achieve this goal. One idea is to observe in the thermal-IR, where the *contrast ratio* between the planet and the star is more

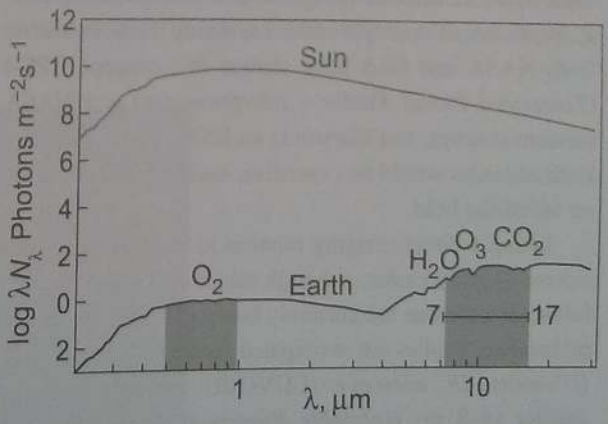

Figure 15.7 Diagram illustrating how the flux of photons from the Earth and Sun would appear if viewed from a distance of 10 pc. The shaded area is the wavelength region that would be observed by a telescope operating in the visible/near-IR (at left) or in the thermal-IR (at right). (From Kasting (2010). Reproduced with permission of Princeton University Press. Copyright 2012.)

favorable (Fig. 15.7). The Earth, for example, is 10^{10} times dimmer than the Sun in the mid-visible, but only 10^7 times dimmer in the thermal-IR. This arises from evaluating the Planck function (eq. (2.43)) in the visible and thermal-IR at emission temperatures of 255 K and 5780 K for the Earth and Sun, respectively, and also accounting for the two bodies' different size.

Apart from contrast ratio, another effect of wavelength must be considered, which is diffraction. The angular resolution, θ (in radians) of a telescope obeys

$$\theta \geq 1.22\frac{\lambda}{D} \tag{15.7}$$

Here, λ is the wavelength at which the observations are being made, and D is the diameter of the telescope mirror (or lens). Equality applies in eq. (15.7) at the *diffraction limit*, which is the smallest size (a so-called *Airy disk*) to which an optical system can focus. As mentioned before, a planet orbiting at 1 AU from star located at 10 pc distance forms an angle of ~0.1 arcsec or $\sim 5\times10^{-7}$ radians. For observations in the visible to near-IR, i.e., out to 1 μm = 1×10^{-6} m, we have $D = 1.22\lambda/\theta \approx 2$ m. Real telescopes involving coronagraphs operate at some multiple of this limit, typically 4 times, requiring a mirror diameter ≥ 8 m. Such large telescope missions have been advocated, e.g., a proposed *High Definition Space Telescope* (HDST) (Dalcanton *et al.*, 2015). Turned around, for a given mirror size, the smallest angle from a star at which an exoplanet can be detected is known as the *inner working angle*. In the thermal-IR, a wavelength of 10 μm is ~10 times longer, and so, according to eq. (15.7), you need an 80-m telescope. An 80 m mirror is too big for a single dish space telescope because it wouldn't fit in a launch vehicle, so thermal-IR designs for an Earth-finding telescope combine multiple detectors using *interferometry*. Both NASA and ESA have studied this concept. TPF-I (*Terrestrial Planet Finder – Interferometer*) is a NASA mission concept, and *Darwin* is an ESA mission concept. Both missions would be expensive, and both are currently on indefinite hold.

Doing a direct imaging mission in the visible is considered slightly easier, although this is still costly and no full-scale missions are currently being pursued. However, preliminary studies are anticipated for a (10–12 m) *Large UV-optical-IR telescope* (LUVOIR) and a somewhat smaller (4–8 m) *Habitable Planets Explorer* (HabEx). Some smaller probe-class (<$1B) missions have also been studied (see http://exep.jpl.nasa.gov/).

Visible direct imaging missions come in two varieties, depending on how one blocks the light from the star. One idea is to suppress the starlight internally within the telescope using a coronograph. This idea was studied under the name TPF-C (where the "C" stands for coronograph) (see http://exep.jpl.nasa.gov/TPF-C/tpf-C_index.cfm). The other idea is to use an external occulter, or *starshade*, to block out the light from the star. This technique is similar to what happens during a solar eclipse, when the Moon passes in front of the Sun. To provide adequate starlight suppression, an occulter would typically need to be 50–70 m in diameter, and it would need to fly at a distance of ~50 000 km from the telescope. So, this mission, like the interferometer, requires multiple spacecraft. The demands on the telescope mirror, however, are much less stringent than for the internal coronagraph, so this mission, sometimes called TPF-O (where the "O" stands for occulter), is under serious consideration. One such concept that has been studied in the USA is called *New Worlds Observer* (http://newworlds.colorado.edu/). NASA's smaller, probe-class analogs to these missions are concepts called Exo-C (Stapelfeldt *et al.*, 2015) and Exo-S (Seager *et al.*, 2015), where the "S" stands for "starshade."

15.3 Characterizing Exoplanet Atmospheres and Surfaces

We hope not only to find Earth-like planets, if they exist, but also to characterize their atmospheres and surfaces (see reviews by Crossfield, 2015; Seager, 2010, 2013). After all, if an Earth-mass planet is discovered within the habitable zone of some nearby star, the first question that will be asked is: "Is it habitable?" And the second will be: "Is there any evidence that it is indeed inhabited?"

A number of different characterization techniques can be imagined, ranging from relatively crude to highly detailed. The methods used will depend on what type of stellar system is being investigated and on when the measurements are being made. Today, for most systems, it is not possible to separate out the light from a planet from that of its parent star (except for a few super-Jupiters at large distances from their parent stars, which have already been directly imaged using ground-based telescopes); thus, our options are currently rather limited. Below, we look to the future instead.

15.3.1 The Near Term: Transit Spectra of Planets Around Low-Mass Stars

In the next 10–15 years, our best chance of characterizing habitable planets will be to look at nearby low-mass (late K and M) stars. The habitable zones around such stars are relatively close in, and so the probability that a potentially

habitable planet will transit is higher, according to eq. (15.6). The decreased orbital distances within the HZ are offset to some extent by the fact that a star's diameter tends to decrease with its mass, although not as rapidly as its habitable zone shrinks. A NASA mission called TESS (the *Transiting Exoplanet Survey Satellite*), scheduled for launch in 2017, will search for such transiting planets around nearby G and K stars. If such planets can be identified, then techniques that have already been used to study hot Jupiters can be used to characterize their atmospheres and surfaces. Later, an ESA mission, *PLATO* (Planetary Transits and Oscillations of stars) is planned for launch in 2024 to find further rocky exoplanets.

A transit spectrum can be obtained in two ways. When the planet is in front of the star, called *primary transit*, light passes through an annulus of the planet's atmosphere on its way to Earth (Fig. 15.8). By measuring the spectrum of the star when the planet is in front of it, and then subtracting the spectrum of the star when the planet has passed by, one can obtain a spectrum of the planet's atmosphere.

This primary transit allows *transmission spectroscopy*. During the transit, the whole spectrum gets dimmer, but at wavelengths where atmospheric gases absorb, the dimming is bigger, which can give insight into atmospheric composition and structure. The technique was first applied to Solar System planets (Smith and Hunten, 1990). As discussed in Sec.15.2.3, the flux decreases according to the planet : star area ratio as $(r_p/R_s)^2$; so if a planet's atmospheric opacity varies with wavelength, λ, the planet's apparent size changes. Thus, measuring $(r_p(\lambda)/R_s)^2$ gives a transmission spectrum. As the transit decrement is small, high precision is needed. Nonetheless, the technique has been used to study the atmospheres of hot Jupiters (Charbonneau *et al.*, 2002; Lecavalier Etangs *et al.*, 2008), Neptune-mass planets (Knutson *et al.*, 2014), and super-Earths (Bean *et al.*, 2010; Kreidberg *et al.*, 2014).

Clouds and hazes can confound transmission spectroscopy. In the Solar System, *all* thick planetary atmospheres have clouds, while the giant planets, Venus, and Titan have considerable haze. Haze, in particular, obscures deeper layers and, through absorption and multiple scattering, hazes can render transit spectra flat and featureless (Marley *et al.*, 2013; Robinson *et al.*, 2014b; Seager and Sasselov, 2000).

Another technique, which works well for some hot Jupiter-type planets, is called *secondary transit spectroscopy*. In this method, which works best in the thermal infrared where the planet/star contrast ratio is lowest, one first measures the spectrum of the planet + star when the planet is beside the star. Then, one takes another spectrum when the planet the planet is behind the star (secondary transit) and subtracts that one from the first (Fig. 15.9). Such spectroscopy has been performed successfully for hot Jupiter planets using the *Spitzer Space Telescope*, which operates at thermal-infrared wavelengths and has a roughly 1-m diameter mirror (e.g., Agol *et al.*, 2010).

The *James Webb Space Telescope* (JWST), scheduled to launch in 2018, has a 6.5-m mirror, so it just might be

Figure 15.8 Diagram illustrating the technique of primary transit spectroscopy. During transit, some of the light reaching the observer passes through the atmosphere of the planet. The planet blocks more light in a transit at wavelengths at which the atmosphere absorbs. (From Kasting (2010). Reproduced with permission of Princeton University Press. Copyright 2012.)

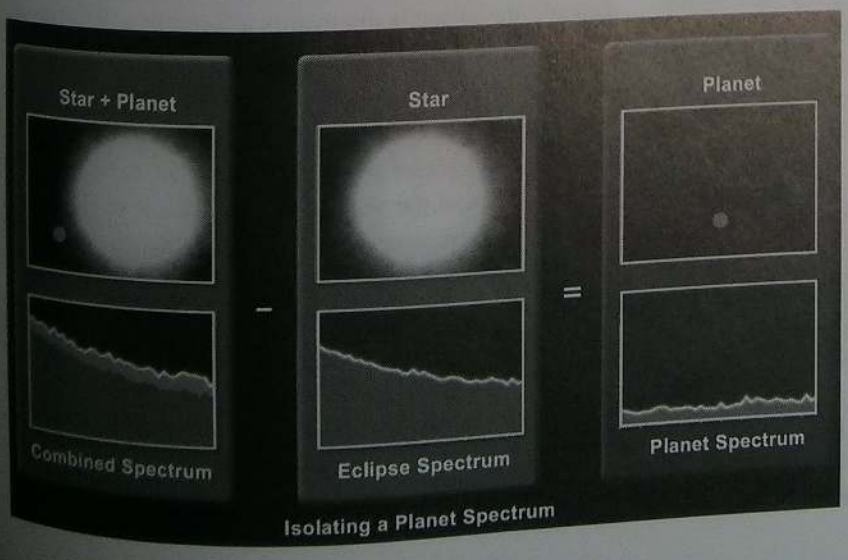

Figure 15.9 Diagram illustrating the technique of secondary transit spectroscopy. (From Kasting (2010). Reproduced with permission of Princeton University Press. Copyright 2012.)

able to obtain a spectrum of an Earth-like planet around a nearby M star (although not for any of the currently known transiting systems). That, of course, would be extremely exciting. We will not dwell on this topic here, though, because it remains to be seen if this can actually be done.

15.3.2 The Future: Direct Detection of Habitable Planets

Much more can be learned once we are able to image planets directly, that is, to separate out their light from that of the parent star. The TPF missions described earlier in this chapter would be able to do this. What can be discovered depends on how much of a planet's orbit can be observed and on how many photons are available. Let's consider the simplest observations and progress to the more complex.

15.3.2.1 *Searching for Liquid Water on a Planet's Surface*

If one can measure the planet's brightness over a significant fraction of its orbit, one can obtain information about its atmosphere and surface. Figure 15.10 illustrates the geometry of the observation for a planet whose orbital plane is aligned with the observer. (Such a planet would transit its star, but in this case we are interested in observing it out of transit.) In practice, it may be difficult to obtain a complete light curve because the planet gets harder to see as it moves closer to the star. Figure 15.11, though, illustrates the kind of information that might be obtained from even a partial orbit. If a planet behaves like a Lambertian sphere, i.e., if it scatters light equally in all directions (see Fig 2.7), then its brightness relative to its parent star is given by the following (Russell, 1916; Sobolev, 1975).

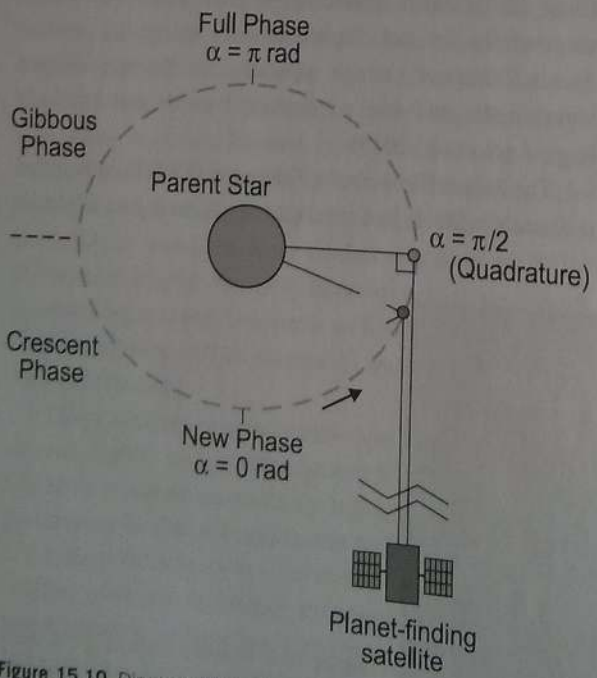

Figure 15.10 Diagram illustrating the geometry of an observation of an exoplanet out of transit. The orbital longitude, α, is defined as 0° when the planet is in front of the star. The phases of the planet are indicated, which are analogous to those we see of the Moon.

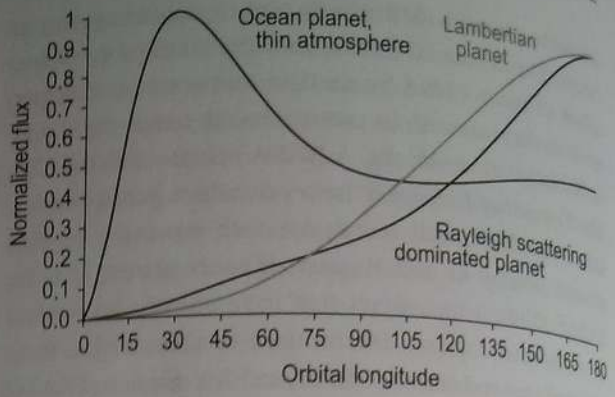

Figure 15.11 Light curves for different types of planets for an edge-on viewing geometry. (From Zugger *et al.* (2010). Reproduced with permission. Copyright 2010, American Astronomical Society.)

$$C(\alpha) = \frac{2}{3} A \left(\frac{r_p}{a}\right)^2 \left(\frac{\sin\alpha + (\pi - \alpha)\cos\alpha}{\pi}\right) \tag{15.8}$$

Here, A is the planet's Bond albedo, r_p is its radius, a is its orbital distance, and α is the orbital longitude. $C(\alpha)$ is wavelength-dependent, so technically A needs to be defined for the wavelength *range* of interest. At quadrature ($\alpha = \pi/2$), $C = (2/3)A(r/a)^2/\pi$. For an Earth-like planet orbiting a Sun-like star, $A \cong 0.3$ (Palle *et al.*, 2003), $r =$ 6371 km, and $a = 1.5\times10^8$ km, the contrast ratio C is $\sim 1.15\times10^{-10}$. This is why a visible TPF mission is so difficult: the planet is dimmer by a factor of $\sim 10^{10}$ than its parent star.

As one can see from Fig. 15.11, the light curve for a planet depends on its type of atmosphere and surface. A Lambertian planet would appear brightest right before it passed behind the star. (It would, of course, disappear from view somewhat before that.) Raleigh scattering mimics Lambertian scattering, so a planet with a thick atmosphere (and little absorption) would do the same thing. An ocean planet with a thin atmosphere would much different, however: its brightness would peak at an orbital longitude of ~30°. That's because the albedo of a liquid water surface depends strongly on the angle of the incident starlight. Light that is vertically incident is strongly absorbed, but light that strikes the surface at a

grazing angle is efficiently reflected. A *glint* spot is present where the angle of incidence equals the angle of reflection; it is spatially extended because of the finite size of the host star and the waviness of the liquid surface (Robinson *et al.*, 2010; Sagan *et al.*, 1993b). In principle, polarization could also be used to look for liquid water, but this appears not to be useful unless the planet's atmosphere is exceedingly thin (Zugger *et al.*, 2010).

15.3.2.2 *Using Color to Characterize Exoplanets*

To get a better estimate of an exoplanet's nature, the next step will be to measure its reflectance spectrum in several broad wavelength bands, at a resolution $R \equiv \lambda/\Delta\lambda \cong 5$. Photometric observations can be decomposed into three bands, and a planet can be placed on a color–color diagram, as shown in Fig. 15.12. In Fig. 15.12, the reflectance ratios of Solar System bodies were obtained from photometric observations of the EPOXI emission and other spacecraft with filter bands of 100 nm at the center wavelengths indicated (Crow *et al.*, 2011). (The EPOXI mission was what became of the *Deep Impact* spacecraft once it launched its probe into the comet 9P/Tempel.) In Fig. 15.12, the Earth occupies a special position compared to other Solar System bodies. Rocky planets with little or no atmosphere (Mercury, Moon, Mars) cluster in the red or near-IR corner of the diagram, while methane-rich gas giants (Uranus, Neptune) cluster on the opposite "blue" corner. Earth has a blue up-turn at short wavelengths and so is separated from the other planets. Earth is a *Pale Blue Dot* with a U-shaped spectrum in the visible to near-IR because of high reflectance in the blue (<450 nm) from Rayleigh scattering, weak absorption across the visible from ozone Chappuis bands (450–850 nm), white clouds reflecting across the visible, and an upward slope at 600–900 nm caused by continents and vegetation (Arnold *et al.*, 2002; Robinson *et al.*, 2011; Tinetti *et al.*, 2006).

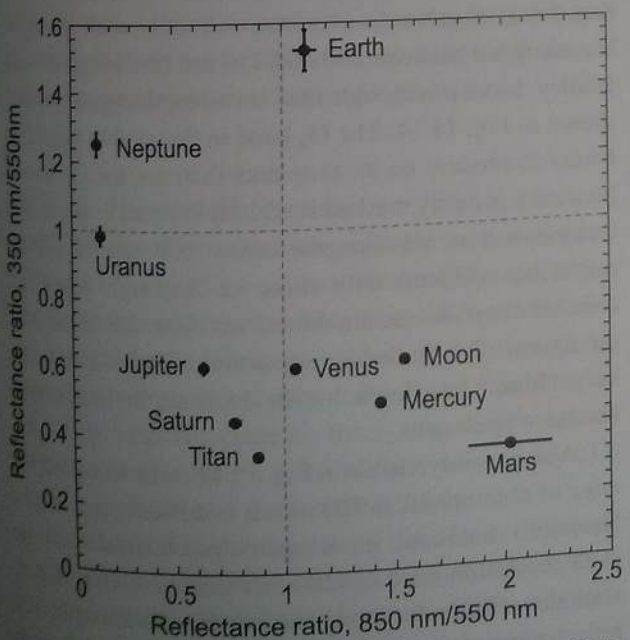

Figure 15.12 Color–color diagram for planets within our own Solar System using filters from the EPOXI mission. (From Crow *et al.* (2011). Reproduced with permission. Copyright 2011, American Astronomical Society.)

A study examining optimal photometric bins over 350–1000nm for identifying an exo-Earth shows that it can be difficult to distinguish an Earth from potential false positives. Icy worlds with thick atmospheres are similar in color because of Rayleigh scattering in the blue (Krissansen-Totton *et al.*, 2016b). Consequently, somewhat higher spectral resolution, R~10 is needed. The same study also shows that an early Earth-type planet shrouded in organic haze (an Archean analog) cannot be distinguished by color alone from a Titan-like planet. So, while color could provide some information, ultimately spectra are required to reveal the true nature of exoplanets and remove ambiguities.

15.3.2.3 *Measuring Planetary Rotation Rates, Land–Sea Distributions, and Planetary Obliquities*

Color can be used to derive other planetary characteristics if one can measure how a planet's color changes with time. Ocean surfaces are dark, whereas land surfaces are somewhat reddish, and clouds are gray. So, if one can measure the color variation as the planet rotates, it is possible to learn something about the planet's rotation rate (Oakley and Cash, 2009; Palle *et al.*, 2008) and about land–sea distributions (Cowan *et al.*, 2009).

NASA demonstrated the capability to examine color change with its EPOXI spacecraft. Figure 15.13 shows a map of planet Earth, along with its inferred geography from EPOXI data. All latitudinal information is lost because the planet was viewed as a single pixel, as would be the case for an exoplanet. But the longitudinal distribution of continents is readily apparent from the time series analysis of different color bands. Knowing whether a planet has continents as well as oceans could be valuable in evaluating its ability to harbor complex, or intelligent, life. This technique is limited to relatively bright planets (or very big direct imaging telescopes), however, because the integration time must be substantially shorter than the planet's rotation period.

Finally, if one is able to perform such a rotation rate analysis at a variety of orbital phase angles, one can in

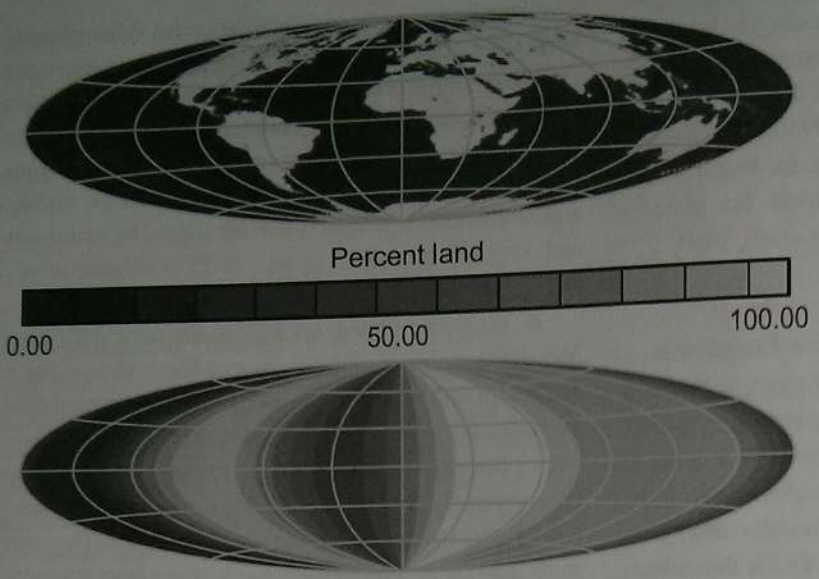

Figure 15.13 Inferred land–sea distribution for Earth, obtained from analysis of the *EPOXI* color time series. (From Cowan *et al.* (2009). Reproduced with permission. Copyright 2009, American Astronomical Society.) (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

principle determine an exoplanet's obliquity and obtain a 2-D map of continental geometry (Kawahara and Fujii, 2010). We say "in principle" because this probably requires being able to view the planet when it is relatively close to the star, and that, in turn, needs sufficient angular resolution. But even if resolution proved difficult with a first-generation direct imaging mission, it might be possible to do this at a later time by repeating the observation with a bigger space telescope.

15.3.2.4 *Visible/Near-IR Spectra*

Much more information is available if one can obtain a reasonably detailed ($R \cong 70$) spectrum of the planet's atmosphere. A spectrum of the Earth's atmosphere at UV/visible/near-IR wavelengths is shown in Fig. 15.14. This spectrum was taken using "Earthshine" data – light reflected from the dark side of the Moon (from Woolf *et al.*, 2002). These photons originated from the Sun, bounced off the Earth, and then back off the Moon. If one subtracts a spectrum of the bright side of the Moon, one obtains a reflection spectrum of the Earth. In the figure, the curve at the top represents the Earthshine data. The smooth curve running through the data, which matches the "clear sky" curve from lower down, is a model fit to the data. From this fit, one can detect three different gases: O_2, O_3, and H_2O. They are all seen as absorption bands. The brightness within these bands is less than it is elsewhere in the spectrum because the gases are absorbing some of the incident sunlight. O_2 itself has three different absorption bands that can be seen at this spectral resolution. The brightest of these is the O_2 "A" band at 760 nm. The "A" band is easy to observe – it would span two pixels in a spectrograph at $R = 70$ – and was singled out almost 30 years ago as a possible indicator of life on exoplanets (Owen, 1980). The O_2 "B" band at 690 nm is also easy to pick out, as are the three H_2O absorption bands at 720 nm, 820 nm, and 940 nm.

Ozone has broad absorption bands in the visible (the Chappuis bands) that extend from 450 nm to 850 nm and peak in the yellow-orange at ~600 nm (as mentioned in Sec. 2.4.2). O_3 absorbs even more strongly at shorter UV wavelengths, between 200 and 310 nm (the Huggins and Hartley bands), although that wavelength region is not shown in Fig. 15.14. The O_3 band in the visible might be harder to observe on an exoplanet than are the O_2 bands because it is easily masked by clouds (Segura *et al.*, 2003) and because at wavelengths below 600 nm its effects might be confused with those of Rayleigh scattering (labeled "ray" in the model curves near the bottom of the figure). The Earthshine spectrum in Fig. 15.14 is also very "blue," as shown by the increase in intensity at shorter wavelengths.

Also possibly visible in Fig. 15.14 is the so-called *red edge* of chlorophyll at 700 nm. It is difficult to pick out, however, because it is partially masked by an H_2O absorption band at 720 nm. Plants and algae absorb sunlight effectively shortward of this wavelength and reflect most of the sunlight longward of this wavelength (Kiang *et al.*, 2007a; Kiang *et al.*, 2007b; Sagan *et al.*, 1993b; Seager *et al.*, 2005). The red edge is a questionable biosignature for extraterrestrial life, partly because it

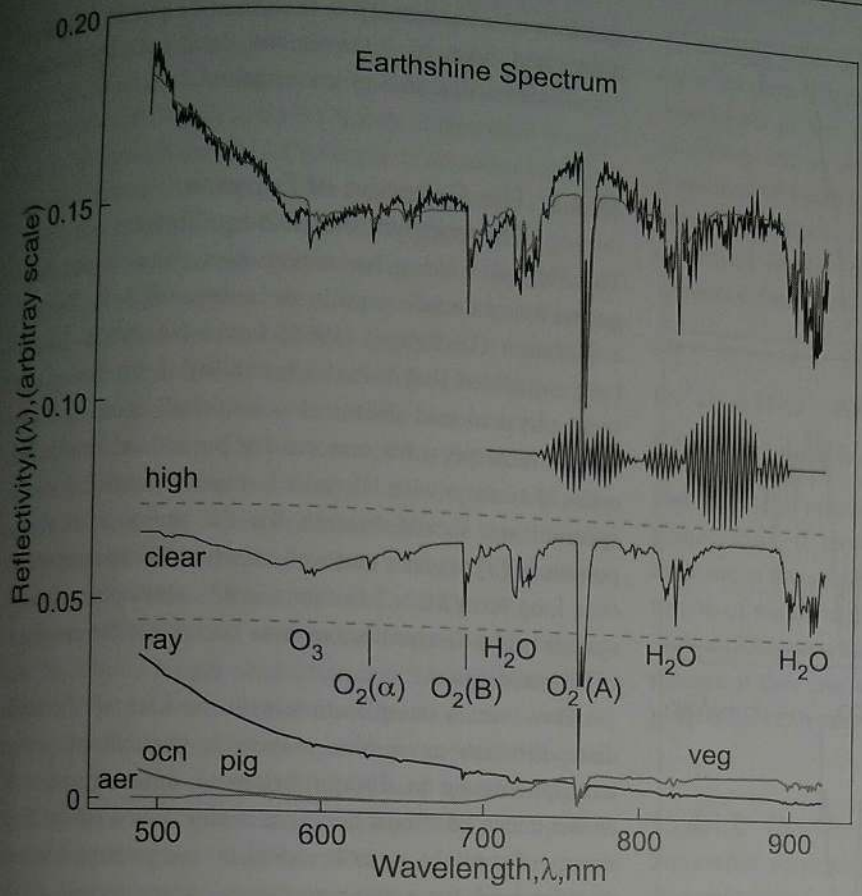

Figure 15.14 Visible/near-IR reflectivity spectrum of the Earth, taken from "Earthshine" data. The wiggly curve at the top shows the reflectivity data, which is the ratio of earthshine to moonshine corrected for phase, but with arbitrary scale. The smooth curve running through it shows a model fit (from seven components below). Five CCD interference fringes (inset and offset to the right) were subtracted from the data. The seven components were: "high" for reflectivity from a high cloud; "clear" for clear atmosphere transmission; "ray" for Rayleigh-scattered light; "veg" for reflected light from vegetated land; "pig" for reflected light from green-pigmented marine phytoplankton; "aer" for aerosol scattered light (negligible); and "ocn" for reflected light from the ocean (From Kasting, (2010). Reproduced with permission of Princeton University Press.) (Originally from Woolf *et al.* (2002). Reproduced with permission from The American Astronomical Society, Copyright 2002.)

is only marginally visible in disk-averaged spectra like Fig. 15.14 and partly because it is not certain that an alien biota would exhibit a spectral signature at this same wavelength. But we would certainly look for if we did obtain the spectrum of an extrasolar Earth.

15.3.2.5 *Thermal-IR Spectra*

Useful biosignatures are also available in the thermal-infrared. Figure 15.15 shows thermal-IR spectra of Venus, Earth, and Mars. Looking first at Venus and Mars, one can see that at this relatively low spectral resolution only a single feature is clearly visible: the 15-μm band of CO_2. This band is created by the ν_2 "bending" mode of the CO_2 molecule, and it is the primary reason why CO_2 is a strong greenhouse gas (see Sec. 2.5.4).

If we look at Earth's spectrum, the 15-μm CO_2 band is also clearly visible, even though Earth's CO_2 concentration is relatively low. Thus, the thermal-IR is an excellent place to look for CO_2. It is also a good way of distinguishing terrestrial (rocky) planets like Venus, Earth, and Mars from gas giant planets like those in the outer Solar System, because the gas giants lack CO_2. The atmospheres of some hot Jupiters contain measurable amounts of CO_2, as a consequence of rapid photochemistry, but these objects can be readily distinguished from terrestrial planets on the basis of their orbital distances. Earth's thermal-IR spectrum contains additional information. The absorption at short wavelengths (<8 μm) is caused by H_2O, as is the absorption at long wavelengths (>17 μm). The short-wavelength feature is the 6.3-μm rotation-vibration band, and the long-wavelength feature is the H_2O pure rotation band, which extends all the way out to the microwave region where it is used for heating food. So, as in the visible/near-IR, it should be possible to determine whether a planet has abundant water vapor in its atmosphere, but not necessarily on its surface.

Even more interesting is the strong band of O_3 centered at 9.6 μm. This band is clearly visible even though ozone is only a trace constituent of Earth's atmosphere, because the ozone resides mostly in the stratosphere. Ozone is formed photochemically from O_2 (see Sec. 3.3.1) so its presence indicates that the planet must have O_2 in its atmosphere. But, one also finds that O_3 can be detected even if only small amounts of O_2 are present (Leger *et al.*, 1993; Segura *et al.*, 2003). The reasons have

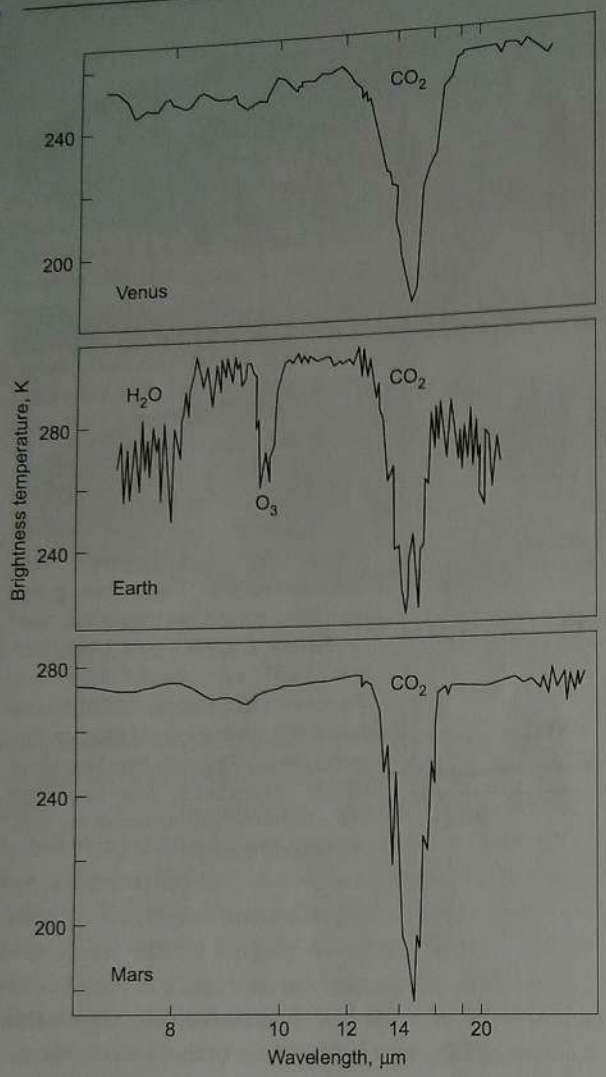

Figure 15.15 Thermal-IR spectra of Venus, Earth, and Mars. (From Kasting (2010). Original figure courtesy of Robert Hanel, NASA GSFC.)

to do partly with the nonlinear nature of the ozone photochemistry and partly with the fact that the stratospheric temperature drops as the amount of ozone decreases, making the absorption feature appear stronger. Thus, O_3 is, in some ways, an even more sensitive indicator of O_2 than is O_2 itself. By comparison, the O_2 "A" band becomes difficult to see below about 10% of Earth's O_2 abundance (Segura *et al.*, 2003).

15.4 Interpretation of Possible Biosignatures

Once a planet is found to be habitable, the next step will be to try to determine if it is actually inhabited. To do this, we will need to look for biosignature gases, some of which (e.g., O_2) have already been mentioned. But the question of what exactly to look for in a planet's atmosphere and how to interpret the data, once obtained, remains an active area of investigation.

15.4.1 The Criterion of Extreme Thermodynamic Disequilibrium

The idea of looking for remote biosignatures was suggested many decades ago in the context of Solar System exploration (Lederberg, 1965; Lovelock, 1965). Lederberg suggested that "kinetic instability in the context of local physical and chemical conditions" could indicate life. In retrospect, his concern for possible abiotic influences was very wise. Lovelock argued something more specific: we should "search for the presence of compounds in [a] planet's atmosphere which are incompatible on a long-term basis." In fact, Earth's atmosphere–ocean system is in a disequilibrium state because of the presence of biogenic gases.

However, a complication with the idea of chemical disequilibrium as a biosignature is that all planetary atmospheres are in disequilibrium to some degree as a consequence of atmospheric chemistry driven by the free energy of sunlight, particle radiation, and perhaps internal or tidal heat for some exoplanets. Some hypothetical, abiotic, CO-rich atmospheres could even be extremely out of equilibrium. For example, Kasting (1990) pointed out that such an atmosphere could have been produced by impacts on the early Earth. Zahnle *et al.* (2008) generated CO-rich atmospheres by way of photochemistry in a cold, dense early atmosphere on early Mars. As discussed in Sec. 12.2.3, the underlying reason why high-CO atmospheres are possible has to with a quirk of photochemistry: photolysis of CO_2 is relatively rapid, but recombination of (ground-state) O and CO to reform CO_2 is slow because the reaction is spin-forbidden; consequently, O atoms are more likely to recombine with each other to form O_2. That said, if CO ever was abundant in Earth's early atmosphere, it should have quickly fallen to low concentrations once life evolved. Thus, if anything, the presence of highly disequilibrium CO in a planet's atmosphere should be considered an "anti-biosignature" (Zahnle *et al.*, 2011).

15.4.1.1 *The Simultaneous Presence of O_2 and Reduced Gases*

Certain types of chemical disequilibrium, however, may be reliable biosignatures. Lovelock later reiterated and developed his original argument by suggesting that the simultaneous presence of O_2 and CH_4 would be a good indication of life (Hitchcock and Lovelock, 1967;

Lovelock, 1975). It is indeed difficult to see how these gases could be present simultaneously in appreciable concentrations without having a biological source for both, as the lifetime of CH_4 in today's O_2-rich atmosphere is only ~12 yr. Although the CH_4–O_2 couple is minor in terms of thermodynamic free energy (see below), kinetics requires a large flux of CH_4 into the atmosphere, and that flux is ~90% biogenic. Conversely, an atmosphere with a large abiogenic CH_4 flux would be unlikely to build up free O_2 because the global redox budget (eq. (8.40)) would be heavily weighted towards the reduced side. Kinetically speaking, not a single molecule of CH_4 should exist in the presence of so much O_2 were it not for the continuous input of CH_4 (as shown in Sec. 3.1.1). So, the CH_4–O_2 couple is important if we think in terms of kinetics and lifetimes rather than available free energy (Krissansen-Totton *et al.*, 2016a; Simoncini *et al.*, 2013).

O_2 and N_2O should also constitute a good biosignature. N_2O has a longer photochemical lifetime than CH_4, about 150 yr; however, it has even fewer natural sources and may therefore provide even stronger evidence for life. The simultaneous detection of O_2 and either CH_4 or N_2O is thus sometimes cited as the "Holy Grail" of remote life detection (Kasting, 2010).

Unfortunately, it would not be easy to observe these combinations of gases simultaneously on a planet like modern Earth because the concentrations of CH_4 and N_2O are only ~1.7 ppmv and 0.3 ppmv, respectively. Consequently, both CH_4 and N_2O are missing from the relatively low resolution spectra shown in Figs. 15.14 and 15.15.

These gases were seen, however, in a somewhat higher resolution, near-IR spectrum of the Earth obtained by the NIMS (Near Infrared Mapping Spectrometer) instrument on the *Galileo* spacecraft as it swung by Earth on its way out to Jupiter (Sagan *et al.*, 1993b). The bands observed by NIMS were between 2 μm and 5 μm, which are difficult to access for exoplanets because of the low photon flux at these wavelengths (Fig. 15.7). But if we saw an interesting Earth-analog planet on a first-generation direct imaging mission, one can imagine sending a bigger, follow-up mission that would be capable of taking spectra in this wavelength range. So, the Holy Grail may be accessible if one invests enough effort in looking for it.

There may also be Earth-like planets on which the O_2–CH_4 pair is naturally more observable. As discussed in Chapters 10 and 11, both gases may have been present in appreciable concentrations during the Proterozoic as a consequence of the very different nature of the marine biosphere. So, a planet resembling the Proterozoic Earth might be a good candidate for remote life detection.

O_2 and CH_4 might also be observed simultaneously on a modern Earth-like planet orbiting an M star. As mentioned in the previous section, atmospheric photochemistry on an M-star planet would be very different from that of Earth because of the low flux of visible and near-UV radiation. On Earth today, CH_4 is primarily destroyed in the troposphere by the following reaction sequence (see Sec. 3.1.2).

$$O_3 + h\nu \rightarrow O_2 + O(^1D) \qquad \lambda \leq 310\,\text{nm} \qquad (15.9)$$

$$O(^1D) + H_2O \rightarrow 2OH \qquad (15.10)$$

$$CH_4 + OH \rightarrow CH_3 + H_2O \qquad (15.11)$$

But the rate of reaction (15.9) would be slow on an M-star planet; thus, if methanogenic microbes were producing methane at the same rate as on modern Earth, the concentration of methane in the planet's atmosphere could be as high as 500 ppm (Segura *et al.*, 2005). So, perhaps M-star planets, if they can harbor life at all, will end up being the place where extraterrestrial life is first definitively detected.

15.4.1.2 *The O_2–N_2 Pair*

Somewhat surprisingly, the degree of thermodynamic disequilibrium in Earth's atmosphere (ignoring any interaction with the ocean) is not that large. Indeed, when disequilibrium is quantified as the Gibbs free energy available from reacting all the gases in a mole of air to equilibrium, it turns out that Earth's air has nearly a 100-fold smaller thermodynamic disequilibrium than does Mars (Table 15.2). (Mars' disequilibrium is the presence of CO and O_2 created by photochemistry.) The O_2–CH_4 pair is the dominant source of the disequilibrium in

Table 15.2 Comparison of the available Gibbs free energy in Solar System atmospheres defined here as the difference in Gibbs free energy between the observed composition and a theoretical equilibrium chemical composition calculated at the surface pressure–temperature conditions or 1 bar conditions for giant planets (Krissansen-Totton *et al.*, 2016a).

Planetary atmosphere	Available Gibbs free energy (Joules per mole of atmosphere)
Venus	0.06
Earth (atmosphere only)	1.5
Earth (atmosphere and ocean)	2326
Mars	136
Jupiter	0.001
Titan	1.2
Uranus	<0.1 (upper limit)

Earth's atmosphere, but the free energy is small because the concentration of CH_4 is small.

It is only when you allow Earth's atmosphere to react with the ocean water in such a calculation that Earth stands out as having a far larger disequilibrium compared to other bodies in the Solar System. In the air–ocean calculation, nearly all of Earth's disequilibrium is provided by O_2, N_2, and water. The equilibrium state of these substances is dilute nitric acid, i.e.,

$$2N_2(g) + 5O_2(g) + 2H_2O(l) \rightleftharpoons 4H^+(aq) + 4NO_3^-(aq) \quad (15.12)$$

In fact, the idea that this disequilibrium is the largest one at the Earth's surface was first noted by the chemist Gilbert Lewis (who discovered the covalent bond), although he didn't quantify it (Lewis and Randall, 1923, pp. 567–568).

Earth's disequilibrium is certainly biogenic, including the contribution from O_2–N_2–water. Oxygenic photosynthesis obviously produces the O_2, but the N_2 is also maintained by photosynthesis because the large quantities of photosynthetic organic carbon generate the anoxic conditions in sediments that allow denitrifying microbes to turn nitrate back into N_2.

We conclude that the concept of extreme thermodynamic chemical disequilibrium as a biosignature applies well to gases in a photosynthetic world, such as the O_2–CH_4 kinetic instability and the N_2–O_2–water disequilibrium, but the extent to which it can be generalized to other biospheres is yet to be fully investigated. The pairs identified so far are, in principle, detectable. Even N_2 has spectral features: N_2–N_2 dimers, best seen at 4.15 μm on Earth-like worlds (Schwieterman *et al.*, 2015).

15.4.2 Classification of Biosignature Gases

Seager *et al.* (2013a, b; 2012) and Seager and Bains (2015) have proposed a classification of potential biosignature gases into three categories. *Type I biosignatures* are by-products of existing thermodynamic redox gradients, such as CH_4 on the Archean Earth. Such biosignatures are equivocal because abiotic processes might synthesize them, if suitable kinetic pathways exist. Methane, for example, can be formed by serpentinization of ultramafic rocks by water containing dissolved CO_2 (Berndt *et al.*, 1996). Seager *et al.* suggest that NH_3 might be a Type I biosignature on a cold "Haber world" with a dense N_2–H_2 atmosphere, because it is difficult to produce NH_3 by gas-phase photochemistry. But NH_3 might also be produced by photolytically catalyzed reaction of N_2 with H_2O on surfaces containing TiO_2 (Schrauzer and Guth, 1977).

Type II biosignatures are biomass-building by-products,such as O_2 produced from photosynthesis. Anoxygenic photosynthesis also produces biomass-building by-products, as discussed in Sec. 9.6.5. H_2-, H_2S-, and Fe^{2+}-based photosynthesis yield H_2O, S, and $Fe(OH)_3$, respectively. Other forms of anoxygenic photosynthesis yield SO_4^{2-} and NO_3^{2-} (Seager *et al.*, 2012). Looking at this list, however, shows that most of these other by-products are either solids, liquids, or dissolved ions; thus, none of them would be likely to create an atmospheric biosignature. O_2 may be the only useful one.

Type III biosignatures are secondary metabolic products, such as dimethyl sulfide (DMS), OCS, CS_2, CH_3Cl, and higher hydrocarbons such as isoprene. Many of these species are volatile. But on Earth today, they are produced in such small amounts and/or are photolyzed sufficiently rapidly that they do not build up to high concentrations. If this is true on other Earth-like planets then they will likely not be observable by a first-generation direct imaging mission. But they might be useful targets for larger follow-up missions in the more distant future. Overall, although less obvious biosignature gases are interesting, the issue of whether they are ever likely to be detectable is surely paramount.

15.4.3 Is O_2 by Itself a Reliable Biosignature?

Let's return for a moment to Fig. 15.14, the visible/near-IR spectrum of Earth obtained from Earthshine data. As the figure shows, the only molecules that can be identified from these moderate-resolution data are O_2, O_3, and H_2O. O_3, of course, is formed photochemically from O_2, and so its information content is closely related. Should O_2 (and/or O_3) by itself be considered a reliable biosignature?

This question receives continued attention because it is not that easy to come up with a definitive answer. Nearly all of the O_2 in Earth's present atmosphere has been produced by photosynthesis followed by organic carbon burial. Furthermore, models of early pre-photosynthetic atmospheres show that it is difficult or impossible to build up significant abiotic O_2 levels on a planet with a large ocean like Earth orbiting a star like the Sun (Sec. 9.2). However, it is easy to conceive of abiotic methods for producing O_2 on planets that do not satisfy these conditions. The most obvious one is by way of a Venus-like runaway, or moist, greenhouse. Suppose that Venus started out with as much water as is present in Earth's oceans today, 1.4×10^{21} kg, and that Venus lost most of its water within the first few hundred million years of its history by photodissociation, followed by

escape of hydrogen to space (see Sec. 13.4.5). That process could conceivably have left an enormous amount of oxygen behind – enough to produce about 240 bars of surface pressure if it was all converted to O_2. But most of this oxygen was probably lost during accretion through contact between a steam atmosphere and an underlying magma ocean (Hamano *et al.*, 2013; Kurosawa, 2015). Alternatively, if the *moist greenhouse* model of Kasting (1988) is correct, then Venus' atmosphere might have remained O_2-rich for a billion years or more following the loss of its water. This "false positive" for life would probably not fool us, however, because the planet would be located suspiciously close to the inner edge of the habitable zone and the H_2O bands in the planet's atmosphere would be weak or nonexistent.

A second "false positive" that is easy to identify ahead of time would be a frozen planet like Mars near or outside the outer edge of the habitable zone (Kasting, 1997; Kasting, 2010). As noted in Ch. 12, Mars itself has about 0.17% O_2 in its atmosphere. This is too little to see spectroscopically from a distance because Mars' atmosphere is very thin; however, the process that produces it could operate on other planets as well. Mars' O_2 comes from photodissociation of H_2O and CO_2, followed by escape of hydrogen to space. The net source of O_2 is small. However, the sinks for O_2 are also very small. Volcanic outgassing is negligible, given undetectable SO_2 in the atmosphere (Krasnopolsky, 2012). Furthermore, Mars' surface is cold and dry, and so the rate of surface oxidation is slow. Mars' surface *is* highly oxidized, to be sure, but most of this oxidation may have happened a long time ago. In the absence of liquid water, surface erosion is relatively slow, and so fresh, non-oxidized rocks that serve as an O_2 sink are exposed only gradually.

In addition to surface oxidation, escape of oxygen atoms could limit the O_2 concentration in Mars' atmosphere. Mars is so small (~0.1 Earth mass) that oxygen escapes from Mars' atmosphere at a slow rate. As discussed in Sec. 12.2.4, O atom losses include dissociative recombination (eqs. (12.17) and (12.18)) and sputtering by the solar wind. However, if Mars were a little bigger, neither of these oxygen escape mechanisms would be likely to operate. The higher gravity would prevent O atom escape, and the warmer interior might be sufficient to generate a magnetic dynamo. If the planet remained too small to outgas hydrogen, then O_2 might conceivably accumulate to higher levels.

A more interesting question – because it might lead to confusion – is whether high abiotic concentrations of O_2 and/or O_3 could develop on a planet within the habitable zone. Once again, some false positives can be readily identified. Wordsworth and Pierrehumbert (2014) pointed out that planets that are deficient in N_2 might build up O_2 abiotically because there would be little background gas to prevent a moist greenhouse from developing. This type of pathological situation may or may not occur in reality, as it would be difficult to accrete water and carbon without bringing in nitrogen at the same time. However, skeptics could still point to such an explanation as an alternative to the existence of life. It may eventually be possible to rule out such this type of false positive by looking for the spectral signature of O_2–O_2 and N_2–N_2 dimers (Misra *et al.*, 2014; Schwieterman *et al.*, 2015).

Another false positive may exist for planets orbiting M stars. This one has already been introduced in Sec. 15.1.5.2, because it poses a fundamental problem for M-star planet habitability. Because of the long, bright, pre-main-sequence phase of the host star, planets around M stars may lose their water by a runaway or moist greenhouse. But this implies that they might also build up O_2-rich atmospheres, which could constitute false positives (Luger and Barnes, 2015). Like other post-runaway-greenhouse planets, however, these planets might be identified by their lack of H_2O absorption.

Finally, some recent calculations suggest that O_2 could accumulate on less extreme planets within the HZ, under certain circumstances. Hu *et al.* (2012) calculated that abiotic O_2 concentrations in a CO_2-rich atmosphere might be as high as 10^{-3} by volume, if surface outgassing of H_2 was low. This would likely be detectable spectroscopically, either using the O_2 0.76-μm band or from the associated O_3 signal (Segura *et al.*, 2003). An O_2 mixing ratio of 10^{-3} is above the Proterozoic O_2 level of 2×10^{-4} atm (10^{-3} PAL) estimated by Planavsky *et al.* (2014b), so such a detection would constitute a false positive. But Harman *et al.* (2015) do not get this result, for reasons that are not entirely clear.

Other calculations for abiotic O_2 on M stars have been debated. Tian *et al.* (2014) calculated that abiotic O_2 mixing ratios as high as 4×10^{-3} could be produced on planets with CO_2-rich atmospheres orbiting M stars, even when H_2 outgassing is included. In their model, the dearth of stellar visible/near-UV radiation on these planets lowers the efficiency of catalytic O_2 destruction mechanisms that operate on Solar System planets. Their prediction was not borne out in parallel calculations by Domagal-Goldman *et al.* (2014), but it was supported by Harman *et al.* (2015). All three of these models balanced both the atmospheric redox budget (eq. (8.30)) and the global redox budget (eq. (8.36)). As Harman *et al.* show, whether or not abiotic O_2 can accumulate in a planet's atmosphere depends critically on chemical

reactions that occur within the planet's ocean. The presence of dissolved ferrous iron, similar to Earth's oceans during the Archean (Ch. 12), will suppress atmospheric O_2. Direct reaction between dissolved O_2 and CO, perhaps catalyzed by metal ions, can also suppress O_2. Laboratory experiments are needed to study this reaction. So, while this question is not totally resolved, it remains possible that O_2 by itself is a good biosignature for planets within the habitable zones of FGK stars.

15.5 Parting Thoughts

This brings us to the end of our book. But it is certainly *not* the end of our study of atmospheric evolution. As a society, we are just partway through our exploration of the Solar System, and we are only beginning to study exoplanets. Thus, much of what we have written here, particularly in this last chapter, will be modified in the future. We hope, though, that we have provided a framework that will allow current and future generations of atmospheric, planetary, and Earth scientists, along with astronomers and astrobiologists, to place their discoveries in the context of the work that has preceded them. And we look forward to the construction of the future ground- and space-based telescopes that will eventually allow us to look for other planets like Earth and to test our theories of atmospheric, planetary, and biological evolution.