CHAPTER 7
THE NATURE OF EXOPLANETARY SYSTEMS

Now that the various methods of detecting exoplanets have been described, we turn to the results of the various searches. What types of planet have been detected? What types of planet await detection? Have we already discovered potential planetary habitats, or is this a discovery still to be made? In terms of the Drake equation (Section 6.1.1), what have we learned about the probability, p_p, of planets forming around a suitable star, and the average number, n_E, of suitable planets in a habitable zone per planetary system?

7.1 The discovery of exoplanetary systems

Ever since it has been known that the Sun is a star, people have longed to know whether there are planets around other stars. But it is only since the 1930s that astronomers have believed they might be able to find out. Since then, astrometry has been applied to nearby stars, notably by the Dutch astronomer Peter van de Kamp working at the Sproul Observatory, USA. From as early as 1937 he strove to detect the tiny stellar motions that would reveal the presence of companions of planetary mass. He even thought that he had detected some planets, but later work by others showed that he had not – the apparent stellar orbits were the result of changes made to his telescope in 1949 and 1957. Nevertheless, van de Kamp went to his grave believing that he had discovered at least one planet, around Barnard's star, just 5.9 light-years from the Sun – not much further than the closest star Proxima Centauri, at 4.2 light-years.

Then, in 1992, after decades of disappointment, the first exoplanets were confirmed. The US astronomers Alex Wolszczan and Dale Frail announced that they had detected two planets in orbit around a star of a rare type called a **pulsar**. Each planet had a mass just a few times that of the Earth. The claim that a pulsar had planets was greeted with considerable surprise by other astronomers, but this claim has withstood further investigation. The surprise stems from the way that a pulsar is formed, as the remnant of the explosion of a massive star at the end of its life – a **supernova** explosion. It had not been anticipated that planets could survive such a catastrophe. Perhaps they didn't. Perhaps the planets formed from debris left by the explosion, or perhaps they were captured from a companion star as the pulsar travelled near it. More important for us, life could not have survived the explosion, and even if the planets formed or arrived afterwards, they would be uninhabitable so close to the pulsar, with its deadly radiation. We will therefore discount these and other pulsar planets, and confine our attention to planets around stars that are more like the Sun.

Figure 7.1 Michel Mayor and Didier Queloz, discoverers of the first exoplanet around a Sun-like star. (Geneva Observatory)

In October 1995, two Swiss astronomers, Michel Mayor and Didier Queloz, of Geneva Observatory, announced the discovery of the first non-pulsar planet, in orbit around the Sun-like star 51 Pegasi (star number 51 in a catalogue of the constellation of Pegasus). The result was soon confirmed by others, and by early 1996 astronomers knew that the long drought in exoplanetary discoveries had ended. A steady stream of discoveries has

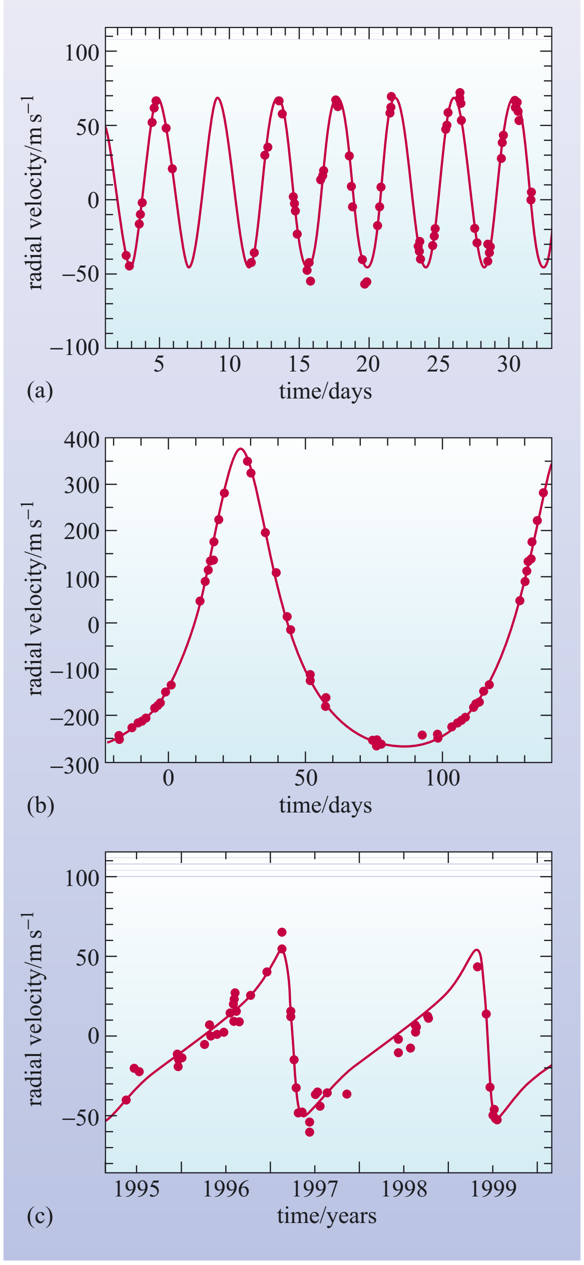

Figure 7.2 The variation of the radial velocity of some stars with planets, as obtained from the Doppler shifts in the stellar spectral lines. (a) 51 Pegasi (its planet is in a low-eccentricity orbit – data from late 1995); (b) 70 Virginis (its planet is in an eccentric orbit); (c) 16 Cygni B (its planet is in an eccentric orbit), where 'B' denotes star B in a multiple-star system, in this case a binary system, so the only other star is 16 Cygni A. ((a) Marcy & Butler)

followed, and still continues today. By mid-2011, 558 exoplanets (non pulsar) had been confirmed, at 469 stars, of which 56 had more than one known planet each. It is likely that all the others are also multiple-planet systems – the formation or survival of a single planet would be rare. The number of discovered planets is growing rapidly, and will be out of date when you read this.

Exoplanets are named after their stars. The letter 'b' is used to denote the first planet discovered, 'c' the next, and so on ('a' is not used). Thus the only known planet of 51 Pegasi is named 51 Pegasi b. If several planets in a system are discovered at the same time, then the lettering starts with the innermost and works outwards, as in Upsilon Andromedae b, c, and d.

The majority of exoplanets have been discovered using the Doppler spectroscopy method. This relies on the detection of the cyclic Doppler shifts in the stellar spectral lines induced by the radial velocity of the star (Section 6.6.2). Figure 7.2 shows data for three stars, along with best-fit curves. From the sinusoidal fit for 51 Pegasi in Figure 7.2a, it can be deduced that the orbit has a low eccentricity. The different-shaped curves in the other two cases indicate more eccentric orbits. The eccentricity can be deduced, though the details will not concern us.

The >400 stars known to have planets through Doppler spectroscopy by 2011 are all nearby – the most distant is at about 28000 light-years (SWEEPS-11), which is only about a quarter of the diameter of the disc of the Galaxy in which we live, and most are *much* closer. This proximity is because the closer a star the greater the flux of radiation we receive from it. The spectral lines are thus clearer, and the Doppler shifts are consequently easier to measure. (Remember that the Doppler shifts themselves are *independent* of the distance to the star – see Section 6.6.2.) One exoplanet discovered by the radial velocity method, around star HD 209458 (HD denotes a catalogue), has also been observed by the occultation (or transit) method (Figure 7.3). To observe a transit the planet's orbit has to be presented to us very nearly edge-on. Exoplanet orbits are orientated randomly on the sky, so on average, out of 100 exoplanets we can expect to observe transits only for about one or two. Another planet discovered by Doppler spectroscopy, Gliese 876 b (another catalogue), has since been detected astrometrically but there has as yet been no confirmed *discovery* of an exoplanet by that technique.

The first confirmed discovery of a planet by a method other than Doppler spectroscopy was of a planet orbiting the star OGLE-TR-56. This star was unnamed before the discovery, and its name carries that of the survey – the Optical Gravitational Lensing Experiment, which is carried out at Las Campanas Observatory in Chile. In spite of the survey name, the discovery was made by the transit method. This is because a survey looking for the changes in apparent stellar brightness due to gravitational micro-lensing (see Section 6.5) could also detect the changes during a transit. The

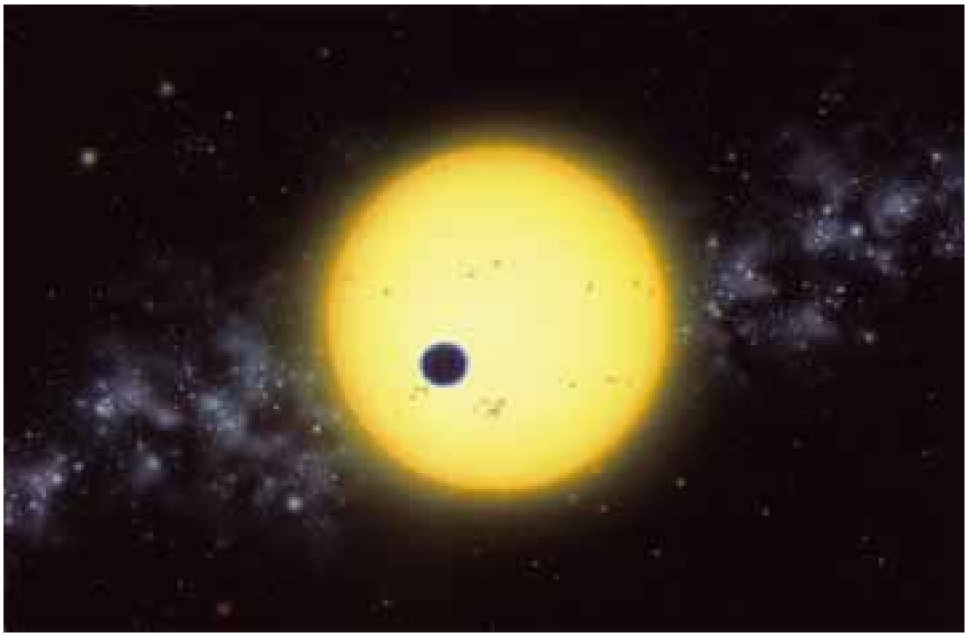

Figure 7.3 An artist's impression of the transit of HD 209458 b across its star, seen from Earth. See Figure 6.11 for the light curve. (Copyright © Lynette R. Cook)

transit and lensing methods have greater range than Doppler spectroscopy, and this is exemplified by the large distance of OGLE-TR-56, about 5000 light-years away. In spite of this distance, the mass of the planet has been obtained by Doppler spectroscopy, aided by the very short 1.2 day orbital period. The mass is $1.3M_J$, where M_J is the mass of Jupiter (318 times the mass of the Earth, M_E). Many other transit discoveries have since been made by other surveys.

We will now look at the known exoplanetary systems in more detail. In Section 7.2 the properties of the host stars are outlined, and then the properties of the exoplanets themselves. You will see that, as yet, largely because of observational selection effects, few Earth-mass planets have been discovered, but many giant planets of the sort unlikely to support life. Nevertheless, the discoveries that have been made have important consequences for the possible existence of Earth-mass planets, as discussed in Section 7.3. In Section 7.4 we summarize the properties of the planets so far discovered, and consider the prospects for future discoveries, including Earth-mass planets.

QUESTION 7.1

Using the graphs in Figure 7.2, measure the radial velocity amplitudes in Figure 7.2a–c. Express your results in (i) m s^{-1}, and (ii) as a multiple of the speed limit for cars in a residential district (30 miles per hour in the UK, which is 13.4 m s^{-1}).

7.2 Properties of exoplanetary systems

7.2.1 The stars in the known exoplanetary systems

Most of the stars presently known to have planets are similar to the Sun – they are Sun-like stars. Stars and stellar classification are outlined in Box 7.1.

BOX 7.1 STARS AND STELLAR CLASSIFICATION

Stars, at birth, are made mainly from the two lightest elements, hydrogen (H) and helium (He). These elements dominate the Universe, and the Sun when it was young would have consisted by mass of about 73% hydrogen, 25% helium and 2% everything else. In astronomy elements other than H and He are called heavy elements, which astronomers also call metals. The proportion of heavy elements is called the **metallicity** of the star. Note that over 99% of the Earth's mass is heavy elements, but that the giant planet Jupiter is much more like the young Sun, though almost certainly with a greater proportion of heavy elements, perhaps as much as 5% to 10%.

All stars we see today, including the Sun, had approximately the same composition when they were born, the most important variation being in metallicity, the range 0.05% to 3% covering most stars at birth. Their evolution depends mainly on their mass (unless there is interference by a close companion star). If the mass is greater than 0.08 solar masses ($0.08M_{\odot}$), then, at some point, the central core of the star becomes hot enough for nuclear fusion to be sustained, in which hydrogen is converted to helium. This releases energy that keeps the star in a fairly stable state. The star is then said to be a **main sequence star**, and the onset of this phase of a star's life marks its zero-age. During the main sequence phase the star roughly triples its luminosity (power output), the rate of increase getting steadily greater through this phase. The surface temperature (which is much lower than the core temperature, where nuclear fusion is occurring) changes much less – by the order of 10%. Table 7.1 shows how the luminosity and surface temperature depend on mass in typical cases, roughly half-way through the main sequence age.

■ Referring to Table 7.1, by what factors are the main sequence lifetime, luminosity and surface temperature of a star of mass $1.6M_{\odot}$ different from those of a star of mass $0.79M_{\odot}$?

❑ Main sequence lifetime is less by a factor of 5; luminosity is greater by a factor of 12; surface temperature is greater by a factor of 1.4.

The fourth column in Table 7.1 is the spectral class of the star. This is essentially a way of specifying the surface temperature. It is widely used in preference to surface temperature because it relates to observed properties of stellar spectra, from which the temperature is then deduced. These properties will not concern us, and neither will the reasons for the seemingly arbitrary association of letters with temperature – these are strictly historical. For our purposes we can take the spectral class as a surrogate for main-sequence surface temperature. The temperature decreases along the sequence O, B, A, F, G, K, M. Each class is subdivided 0, 1, … 9, with temperature decreasing as the numbers increase.

■ Place the following spectral classes in order of decreasing temperature: M5, A9, G2, B0, G5.

❑ B0, A9, G2, G5, M5.

A common mnemonic for the spectral class sequence is 'Oh Be A Fine Guy/Girl, Kiss Me'.

The proportion of stars born with a particular mass increases greatly as mass decreases. For example, there are about 25 times more stars of $0.5M_{\odot}$ than

Table 7.1 Properties of main sequence stars of different mass.

Mass/$M_{\odot}$	Luminosity/$L_{\odot}$	Surface temperature/K	Spectral class	Main sequence lifetime/Ma
23	87 000	35 000	O8	3.0
17.5	40 000	30 000	B0	15
2.9	50	9800	A0	500
1.6	5.7	7300	F0	3000
1.05	1.35	5940	G0	10 000
0.79	0.46	5150	K0	15 000
0.51	0.070	3840	M0	200 000
0.21	0.0066	3170	M5	very long

$2M_\odot$. The abundant main sequence M stars are called **M dwarfs**, because they are quite small.

The duration of the main sequence lifetime is very sensitive to the mass of the star, decreasing as mass increases, as shown in Table 7.1. The main sequence lifetime of a star ends when there is no longer enough hydrogen in its core to sustain sufficient energy output through fusion to helium. Subsequent events are dramatic and take place on timescales considerably shorter than the main sequence lifetime. For stars up to about $8M_\odot$ the star swells hugely and cools to become a **red giant**. It then flings off a large proportion of its mass, leaving a very hot, Earth-sized though massive remnant, which cools very slowly. This remnant is called a **white dwarf**. For a star of greater mass, the end is even more dramatic. Such a star swells to become a supergiant, it then undergoes a (Type II) supernova explosion in which nearly all of its mass is flung into space, leaving an exotic remnant, either a very dense object a few tens of kilometres in diameter, called a **neutron star**, or a **black hole**. Note that a pulsar is a particular kind of neutron star.

By 'Sun-like stars' we mean main sequence stars of spectral class F, G, or K, the Sun being G2, with properties not very different from those of the G0 star in Table 7.1. Their predominance in the known exoplanetary systems is because such stars attracted most of the initial search effort. The reason for this is that they suit the widely used Doppler spectroscopy method by having plenty of sharp spectral lines, good surface stability, and because there are plenty of bright examples. They also have sufficiently long main sequence lifetimes for any biosphere to become detectable from afar.

■ Why can't we include the lifetime after the main sequence phase?

❑ Any life that might have existed on a planet as its star ended its main sequence phase would not survive the post-main sequence events.

Recall from Chapter 2 that the presence of banded iron formations and carbon isotope evidence suggest that life may have emerged on Earth 600 to 700 Ma after the Earth was formed. But it would have been extremely difficult to detect this life from far away. It became much easier when life had greatly modified the atmosphere (Section 2.4.4). This had happened by the time the Earth was about 2000 Ma old. By then oxygenic photosynthesis had caused a considerable increase in atmospheric oxygen, and you will see in Chapter 8 how this can be detected from afar. Sun-like stars (F, G, and K) have sufficiently long main sequence lifetimes for life as we know it to be well established on any habitable planets.

You might wonder why M dwarfs were not initially favoured in the exoplanet searches – the proportion of stars increases greatly from O to M, and M dwarfs are by far the most common. M dwarfs also have huge main sequence lifetimes (Table 7.1). The reason for their neglect stems partly from their more complex spectral lines and their low luminosities. There was also a belief that the proximity of the habitable zone to the star was problematical. Recall (Section 2.3) that the habitable zone is that range of distances from a star in which the stellar radiation on an Earth-like planet would sustain water as a liquid on at least parts of the surface. Figure 7.4 shows the habitable zones for stars like the young Sun (G2) and for a young M0 star. A young M0 star has such a low luminosity that the habitable zone is close to the star, even closer for M1–M9 stars. For any planet in the habitable zone this must result in a tidal interaction between the star and the planet strong enough to establish synchronous rotation in the planet, i.e. it keeps one hemisphere facing the star, just as the Moon does to the Earth. Tidal interactions were discussed in Box 4.2, where their heating

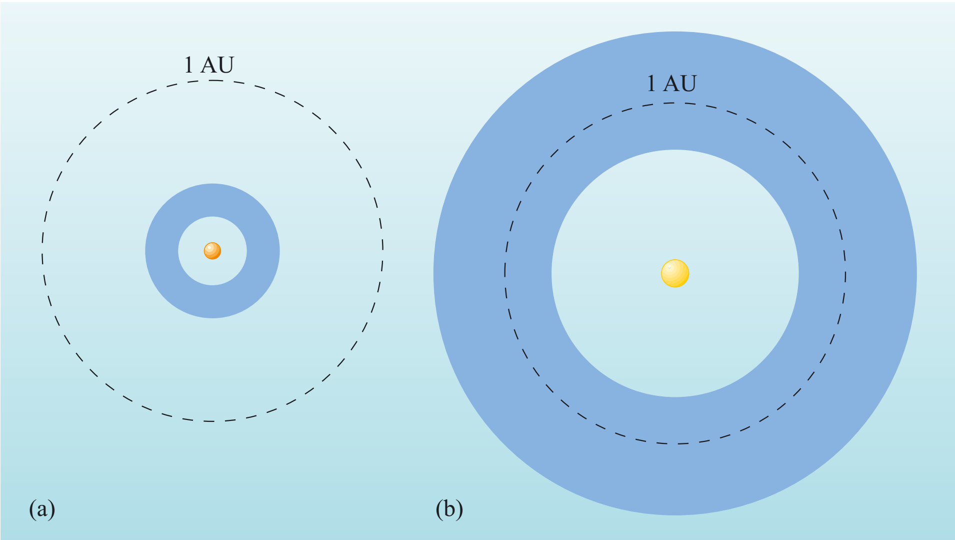

Figure 7.4 Habitable zones (blue-shading) around (a) an M0 main sequence star and (b) a G2 main sequence star (like the Sun), when the stars are young. Intermediate criteria have been adopted for the habitable zone boundaries. (1 AU is the radius of the Earth's orbit.)

effect was of primary concern, though synchronous rotation was also mentioned. When synchronous rotation is established the surface of the outward-facing hemisphere of the planet can get extremely cold, and it has been argued than this could cause the whole atmosphere to condense on this hemisphere. This would prevent life developing on the surface. However, if the atmosphere were sufficiently massive, then its circulation could carry heat from the star-facing hemisphere to the other hemisphere at a sufficient rate to prevent extensive condensation. This is a good reason to survey M dwarfs, and more of them are now being scrutinized.

Observational selection effects, for example the comparative ease with which hot jupiters are discovered (confirmed at 1–1.5% of Sun-like stars), make it difficult to derive reliable statistics, but it is now clear that a substantial fraction of nearby Sun-like stars have exoplanets. One analysis suggests that between 20 and 50% of Sun-like stars have at least one giant planet orbiting with a semimajor axis of less than 20 AU.

More than 10% of stars known to have planets have more than one. Table 7.2 shows the properties of three exoplanet systems known to have at least four planets each.

Table 7.2 Three exoplanet systems known to have at least four members.

Star	Spectral class	Exoplanet	$M_p \sin i_0/M_J$	Semimajor axis a/AU	P/days
55 Cancri	G8	e	0.024	0.038	2.82
		b	0.82	0.12	14.7
		c	0.17	0.24	44.3
		f	0.14	0.78	261
		d	3.8	5.77	5218
Gliese 581	M3	e	0.0061	0.03	3.15
		b	0.049	0.041	5.37
		c	0.0169	0.07	12.92
		g*	0.01	0.146	36.65
		d	0.022	0.22	66.8
		f*	0.023	0.738	433
HD 160691	G3	c	0.033	0.091	9.64
(µ Arae)		d	0.52	0.92	311
		b	1.7	1.5	643
		e	1.8	5.2	4206

* Unconfirmed.

- Which of the stars in Table 7.2 count as ‘Sun-like’?
- 55 Cancri and HD 160691, because their spectral type is in the range F, G and K (in fact, being G3, whereas the Sun is G2, HD 160691 is very much like the Sun). Gliese 581 is spectral type M3, and because of that is not regarded as ‘Sun-like’.

QUESTION 7.2

Redraw Figure 7.4 to show where the habitable zones would lie when the stars are about 5000 Ma old. Justify your estimates. Only an indicative shift of the zone is needed, not its actual position. *Hint*: think what happens to the luminosity of a main sequence star as it ages.

7.2.2 Exoplanet masses

Figure 7.5 shows the number of measured values of $M_p \sin i_0$ in the mass intervals 0 to $1M_J$, 1 to $2M_J$ etc. M_p is the actual mass of the planet and i_0 is the inclination of its orbit with respect to the plane of the sky – remember that the radial velocity method gives $M_p \sin i_0$ not M_p (Section 6.6.2). The least massive Doppler spectroscopy discovery so far is Gliese 581 e, which has $M_p \sin i_0$ of $0.0061M_J$. The smallest planets observed by transit as well as Doppler spectroscopy (so that $\sin i_0$ must be very close to 1) are Kepler-11 f (M_p of $0.0072M_J$), CoRoT-7 b (M_p of $0.015M_J$) and Kepler-10 b (M_p of $0.014M_J$). At the other extreme, we are not concerned with bodies exceeding $13M_J$, because they are better classified as stars than as planets. Such massive bodies will be of roughly stellar composition, made largely of hydrogen and helium. $13M_J$ is the approximate threshold above which the

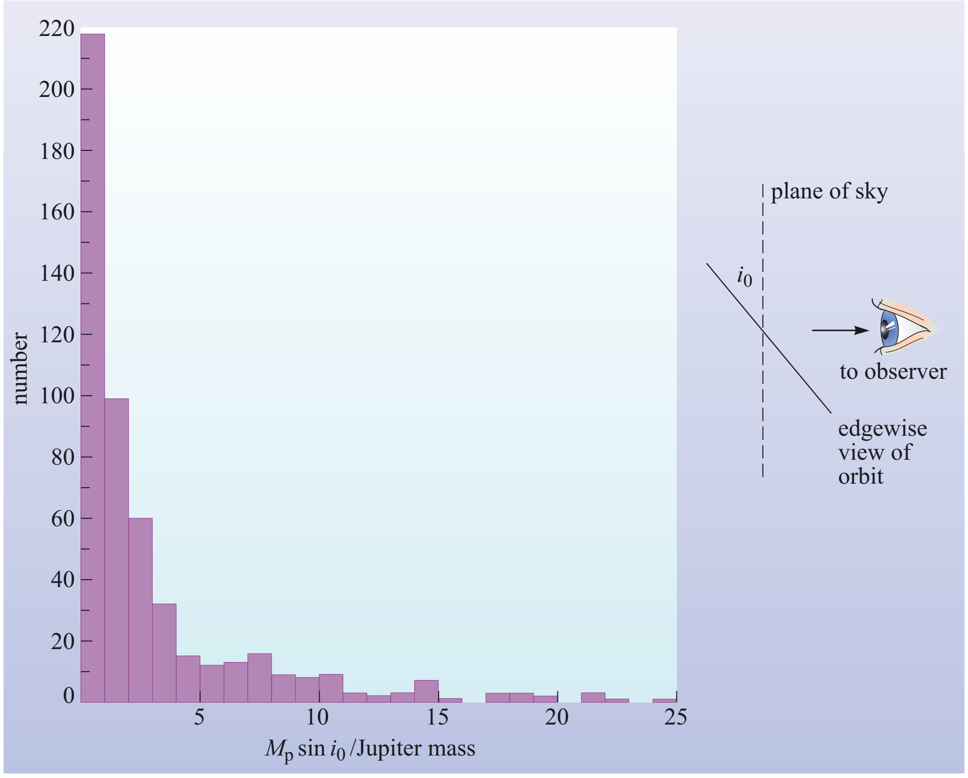

Figure 7.5 The distribution of $M_p \sin i_0$ for the known exoplanets in early 2011. Inset: illustration of i_0.

deep interior becomes hot enough for thermonuclear fusion to occur, involving the rare isotope of hydrogen, called deuterium (^{2}H). This fusion process can last the order of 1000 Ma, because even though deuterium is scarce, it is consumed at a low rate at the modest temperatures in these comparatively low-mass objects.

At masses greater than about $80M_J$ ($0.08M_\odot$) the interior temperatures are high enough to initiate the sustained fusion of the far more abundant isotope ^{1}H, a copious source of power, and so the star then enters on its protracted main sequence phase (Box 7.1). Hydrogen-rich objects in the approximate mass range $13M_J$ to $80M_J$ are called **brown dwarfs**. A large number have been discovered, particularly towards the upper end of this mass range.

Note that thermonuclear fusion is only one of several criteria that can be used to distinguish between a planet and a star, and though there is no consensus on which criterion is best, this one is useful because there seems to be a paucity of bodies with masses around $13M_J$ (Figure 7.5).

The question arises as to how many of the planets in Figure 7.5 are in fact brown dwarfs masquerading as planets, as they would if they have such low values of i_0 that the true masses are above $13M_J$.

■ If $M_p \sin i_0 = 4.5M_J$ and $i_0 = 5.7°$, what is the actual mass of this exoplanet?

❏ $M_p = 4.5M_J/\sin 5.7° = 4.5M_J/0.099 = 45M_J$.

Recall that a planet takes its name from its star, so HD 209458 b is the planet of the star HD 209458. A second planet carries the letter c, and so on.

HD 209458 b is certainly no brown dwarf. This is because it has been observed to transit its star (Section 7.1), and so we must be seeing the orbit nearly edge-on, in fact at $i_0 \geq 87°$, $\sin i_0 \approx 1$. We can also be fairly certain about Epsilon Eridani b (and the possible second planet Epsilon Eridani c), one of more than 20 exosystems known to have a dust ring. The ring is seen to be elliptical. Dust rings around other stars are also seen to be elliptical, but some are seen to be circular.
A circle seen at an angle looks elliptical. Therefore, assuming the ring of Epsilon Eridani to be actually circular, and that the planet orbits in the plane of the ring, a value of $i_0 \approx 46°$ is obtained. Thus given that for Epsilon Eridani b

$$M_p \sin i_0 = 0.86M_J$$

we get the true mass of Epsilon Eridani b

$$\begin{aligned} M_p &= \frac{0.86M_J}{\sin 46^\circ} \\ &= 1.39 \times 0.86M_J \\ &= 1.20M_J \end{aligned}$$

Useful estimates of i_0 can be obtained in some further cases, for example, from observations of the star that yield the inclination of its rotation axis with respect to our line of sight. If the planets orbit in planes that are approximately in the equatorial plane of the star then we get an estimate of i_0. In only a few of these cases is i_0 so small that the true mass exceeds $13M_J$. Overall, only a very small proportion of the bodies in Figure 7.5 could be brown dwarfs. This conclusion is supported on statistical grounds. If the orientations of the orbits of exoplanets are random, then it can be shown that the vast majority will have actual masses within a factor of two greater than their $M_p \sin i_0$ values. Most known exoplanets thus have masses of the order of the mass of Jupiter. But do they also resemble Jupiter in composition?

QUESTION 7.3

Consider Gliese 581 e the exoplanet with the smallest known $M_p \sin i_0$ of $0.0061M_J$. What is a likely upper limit on M_p for this planet? Express your answer in Earth-masses. Why do you think that no Earth-mass planets have yet been discovered by Doppler spectroscopy?

7.2.3 Exoplanet composition

Astronomers know from detailed observations and modelling that Jupiter has a composition not hugely different from the Sun and the other main sequence stars when they were very young, before nuclear fusion altered their compositions. By proportion of mass, the young Sun was 73% hydrogen, 25% helium, and 2% heavy elements (Box 7.1). Jupiter could have a composition close to this, or it could instead be enriched to the level of 5 to 10% of its mass as heavy elements, depending on its mode of formation (see Box 7.2). In any case it would still be dominated by hydrogen and helium. Most confirmed non-pulsar exoplanets have (minimum) masses similar to that of Jupiter, which is considerably greater than the mass of the Earth (Question 7.3). But could they nevertheless have Earth-like compositions, dominated by silicates and iron?

■ Suppose that an exoplanet has the same mass as Jupiter, $318M_E$. Assuming that its mean density is the same as that of the Earth, calculate the radius of this planet in terms of the Earth's radius R_E. (The volume V of a spherical body of radius R is $4\pi R^3/3$.)

❑ The density ρ is given by

$$\rho = \frac{M}{V} = \frac{M}{\left(\frac{4\pi R^3}{3}\right)}$$

where M is the mass.

This equation can be re-arranged to give $R = \sqrt[3]{\frac{M}{\frac{4\pi}{3}\rho}} = \left(\frac{M}{\frac{4\pi}{3}\rho}\right)^{1/3}$.

Therefore for the Jupiter-mass exoplanet

$$R = \sqrt[3]{\frac{318M_E}{\frac{4\pi}{3}\rho_E}}$$

but, for the Earth

$$R_E = \sqrt[3]{\frac{M_E}{\frac{4\pi}{3}\rho_E}}$$

therefore

$$\frac{R}{R_E} = \sqrt[3]{\frac{318M_E}{\frac{4\pi}{3}\rho_E}\frac{\frac{4\pi}{3}\rho_E}{M_E}}$$

$$= \sqrt[3]{318}$$

$$= 6.8$$

In fact, at this greater mass the internal pressures would be much greater than in the Earth, and so the interior would be compressed to a greater density. The actual radius for a Jupiter-mass planet with an Earth-like composition would be more like 5 to $6R_E$.

In comparison, Jupiter's radius $R_J = 11R_E$. This larger radius is because (at comparable pressures) hydrogen and helium are less dense than silicates and iron, so a planet of silicates and iron will have a smaller size than a hydrogen–helium planet of the same mass. Therefore, one way to infer composition is to measure the diameter of an exoplanet of known mass. In effect, this is a determination of its mean density.

After its discovery by Doppler spectroscopy, HD 209458 was found to decrease in apparent luminosity for a short period every 3.52 days, exactly the same as the orbital period of the star, which is the same as that of the planet. The periodic decreases must therefore be due to transits of the planet in front of the star (Figure 7.3). Thus, knowing that the orbit is presented to us approximately edgewise, its actual mass can then be fixed at $0.69M_J$. The amount of dimming of the star enables the radius of the planet to be fixed. This is about $1.35R_J$, which is about 15 times the Earth's radius. This huge size but modest mass certainly rules out a silicate–iron composition, but raises the question, why is it 1.35 times the radius of Jupiter but only 0.69 times Jupiter's mass? The answer is its short distance from its star, which is only 0.045 AU (1 AU is the semimajor axis of the Earth's orbit). The thermal expansion of its atmosphere even this close to its star is a minor effect. More important is the effect of the heat from the star on the planet's rate of cooling. If a hydrogen–helium rich planet of mass $0.69M_J$ had been at this sort of distance from the star from soon after its formation, then the stellar heat would have slowed the planet's rate of cooling, keeping it larger than it would have become if further out. From its size we can thus tell not only that it is a hydrogen–helium rich body, but also that it has been close to its star from soon after it formed. We shall return to this latter point in Section 7.3.1.

Another way to obtain composition, at least of the atmosphere, is to examine the imprint the planetary atmosphere places in the spectrum of its star. In this way, for HD 209458 it has been possible to detect the absorption due to sodium vapour in the planet's atmosphere. Sodium is comparatively easy to detect – detection of the presumed dominant hydrogen and helium is too difficult at present.

From analogous data for OGLE-TR-56 we have a radius about $1.3R_J$ and a mass about $0.9M_J$, giving a mean density about 40% that of Jupiter. The planet is even closer to its star than is HD 209458 b and presumably it is enlarged for much the same reasons. For most other exoplanets we have only *indirect* evidence for their composition. If they are silicate–iron in composition then the circumstellar discs from which they formed must have been highly enriched in heavy elements to make planets amounting to hundreds of Earth masses (Box 7.2, below). The degree of enrichment required seems too large given what we know about the composition of the sort of interstellar clouds from which stars and their planetary systems have been born. In any case, such enrichment would be seen as very high metallicity in the parent stars. Most metallicities are at the upper end of the range, but the enrichment is not large enough to support the hypothesis that the giant exoplanets are silicate–iron. Moreover, the main reason that a silicate–iron composition has been considered is the proximity of so many of the giant planets to their stars, closer than the Earth is to the Sun. If they formed at such distances they would probably have to be entirely silicate–iron because it is hard to see how anything else could condense at the high temperatures so close in (see Box 7.2 below). However, there is an alternative explanation for their proximity that does not involve them forming close in, as you will see in Section 7.3.1.

There is thus reasonable evidence that most exoplanets so far discovered, as well as having masses of the order of the mass of Jupiter, also have compositions that resemble Jupiter in that they are dominated by hydrogen and helium. Towards the lower end of the mass range, the masses are more like that of Saturn, about one-third the mass of Jupiter. But Saturn too is dominated by hydrogen and helium, so the conclusion also applies to smaller giants like these.

QUESTION 7.4

HD 209458 is a star rather similar to the Sun, which has a radius of $9.7R_J$. In Figure 6.10 you saw that HD 209458 b causes a 1.6% drop in apparent luminosity when it transits the star. Use Equation 6.9 to calculate the percentage drop in apparent luminosity of the star when a planet with a radius equal to that of the Earth ($0.090R_J$) makes a transit. If, due to effects of the Earth's atmosphere, a drop of no less than about 0.1% could be reliably detected, comment on whether this 'Earth' transit could be detected.

QUESTION 7.5

Suppose that a planet has $M_p \sin i_0 = 1.9M_J$. It is observed in transit, which gives a radius $1.2R_J$. Show (briefly) how it can be concluded that this is probably a hydrogen–helium giant.

7.2.4 Exoplanet orbits

Figure 7.6 shows $M_p \sin i_0$ (or $M_p \sin i_0$ where known) for exoplanets known in 2011. The diagonal line is for a constant radial velocity amplitude (assuming stars of one solar mass); the further above this line the 'easier' it is to detect a planet by Doppler spectroscopy, and the further below, the harder. You can see that different detection methods favour exoplanets falling in different regions of this diagram. Transit timing is a refinement of the transit method, in which variations in the timing of multiple transits are traced back to orbital perturbations caused by an unseen second planet.

This diagram does not convey the range of orbital eccentricities. About half the known exoplanets have orbital eccentricity $e > 0.1$, including about 15% for which $e > 0.4$, whereas in our Solar System the greatest planetary eccentricities are 0.2 for Mercury and 0.09 for Mars.

A striking feature of Figure 7.6 is how small are some of the semimajor axes – note that the scale is logarithmic. Jupiter and Saturn are shown for comparison – they are further out than most discovered exoplanets.

■ Several of the planets close to their stars in Figure 7.6 have semimajor axes around 0.04 AU. Express this as a fraction of the semimajor axis of the orbit of Mercury, the innermost planet in the Solar System, for which $a = 0.387$ AU, and as a multiple of the solar radius, 6.96×10^8 m.

❑ The fraction is 0.04 AU/0.387 AU = 0.10, about one-tenth the semimajor axis of Mercury. The solar radius = 6.96×10^8 m/1.50×10^{11} m AU^{-1} = 0.0046 AU. Thus the multiple is 0.04 AU/0.0046 AU = 8.7.

This proximity would be unremarkable if the exoplanets were low-mass iron–silicate bodies, as in the Solar System, but most are hydrogen–helium giants. To see why proximity is then remarkable you need to know something about the two current models for the formation of giant planets. A sufficient outline is provided in Box 7.2.

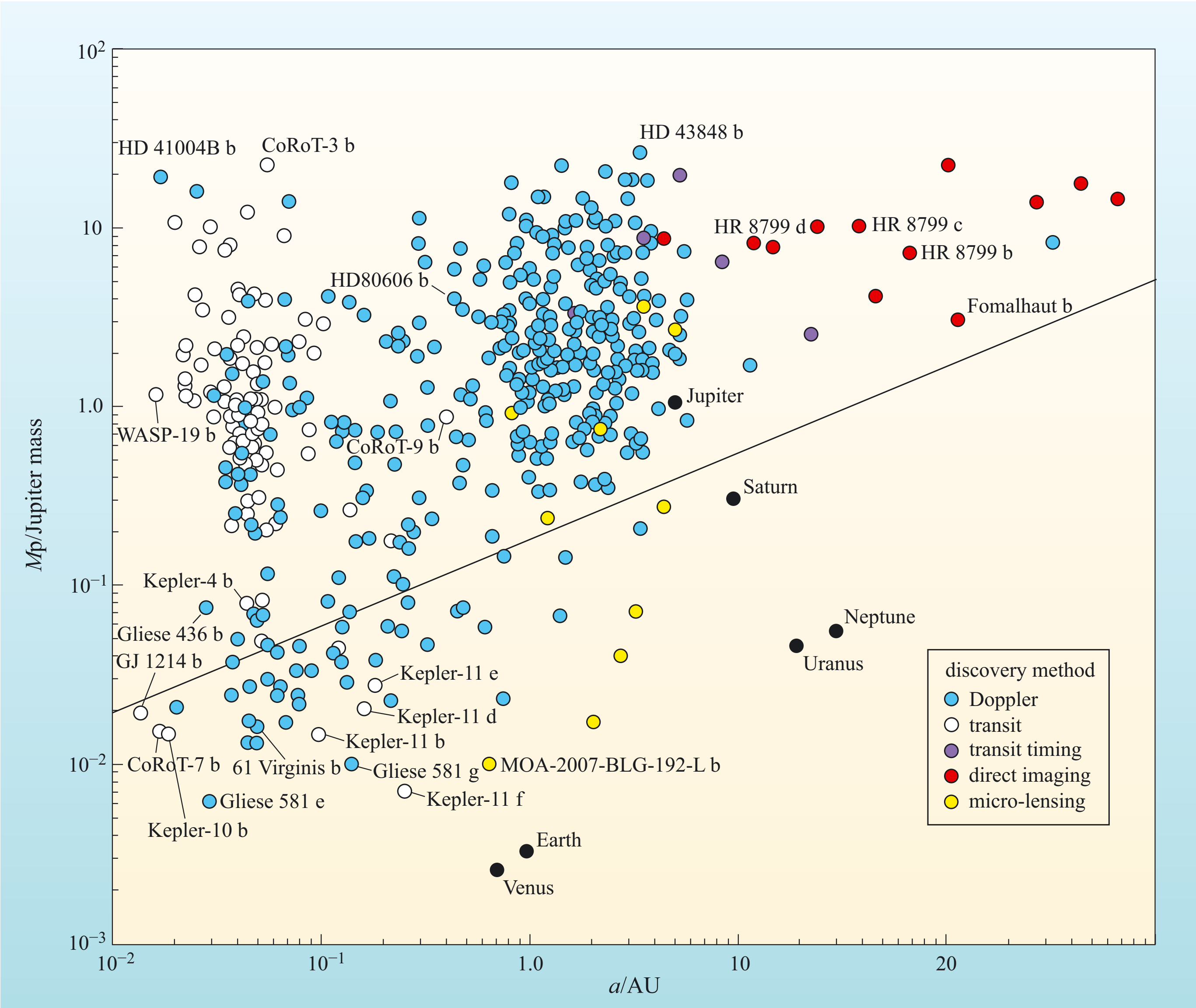

Figure 7.6 Mass of known non-pulsar exoplanets (in 2011) plotted against semimajor axis, distinguished by discovery method. In the case of planets known only by Doppler spectroscopy, the mass is $M_p \sin i_0$. Selected exoplanets are named, and some Solar System planets are shown for comparison. The diagonal line represents a radial velocity amplitude of 5 m s^{-1} (above the line is easier to detect by Doppler spectroscopy than below the line).

BOX 7.2 THE FORMATION OF PLANETS

Almost all astronomers adhere to the view that planets form from a circumstellar disc of material left over from the formation of the central star. Earlier, all this material was a slowly rotating fragment of an interstellar cloud, mainly of hydrogen and helium, that began to contract under its own gravity. As it contracted its rotation rate increased, for the same reason that the rotation rate of an ice-skater with extended arms increases as the arms are pulled in. The underlying reason is the conservation of a quantity called angular momentum. The net outcome of gravitational contraction, increasing rotation, and collisions of molecules and dust grains within the fragment is not at all obvious, but is as shown in Figure 7.7. Most of the mass is in a central condensation that becomes the star, and a few percent is in a flattened disc orbiting the star.

Though the star ends up with nearly all the mass, star and disc have the same elemental composition. In each of them about three-quarters of the mass is hydrogen, about one-quarter is helium, and the proportion of heavy elements – the metallicity – can be anything up to a few percent. In the disc a proportion of the heavy elements is in compounds, mainly with the abundant hydrogen (e.g. CH_4, NH_3 and H_2O, which are common icy materials) but also with each other (e.g. $FeMgSi_2O_6$, which is a silicate). Compounded or otherwise, nearly all of the heavy elements are initially in dust grains spread throughout the disc. The main exceptions are the chemically unreactive ones, such as neon and argon – these remain as gases. Low metallicities might not lead to planet formation, because little dust then forms.

When the star forms it heats the disc, and this evaporates some of the dust to give a radial variation in dust composition. Very close to the star there is no dust at all, just gas. Then there is a zone where dust consists of only the most condensable compounds (the most refractory), such as iron, and silicates. Further out a very important boundary is encountered. This is where water can condense, to form an icy component of the grains.

- Given that oxygen is the most abundant of the heavy elements, what is the relative abundance of water?
- With hydrogen even more abundant, water must be a particularly abundant compound.

Figure 7.7 An artist's impression (Julian Baum) of a young star and circumstellar disc from which planets will form. Note that this is an oblique view – the disc is circular. (Julian Baum/Take 27 Ltd)

The water is in fact about 20 times more abundant than anything else (except hydrogen and helium). The boundary where water condenses is called the **ice-line** (sometimes the snow-line). In the Solar System it was at about 4 AU, and at this boundary, because water and therefore water-ice is so abundant, there is a considerable increase in the mass in the dust grains.

In the inner disc the dust coagulates to form kilometre-sized planetesimals, and these eventually build up to form **planetary embryos** with masses of order $0.1M_E$, which then combine to form a few iron–silicate bodies with masses up to about that of the Earth. Beyond the ice-line, because of the abundance of water-ice, a few bodies build to the order of $10M_E$, mainly composed of water. At this mass the **kernels**, as they are called, gravitationally capture gas from the disc.

- What is this gas made of?
- It is predominantly hydrogen and helium.

Note the distinction between 'inside the habitable zone' which is between the inner and outer boundaries of the habitable zone, and 'interior to the habitable zone' which is the region closer to the star than the inner boundary of the habitable zone.

Thus, giant planets rich in hydrogen and helium are formed. Giants do not form interior to the ice-line because there is no water-ice to build the massive kernels. Gas capture stops when the circumstellar disc is dissipated, aided by convulsions in the young star that cause it to emit a violent wind – the so-called T-Tauri phase. Throughout gas capture, and perhaps for some time afterwards, icy–rocky planetesimals are captured, giving further heavy element enrichment, though the composition remains dominated by hydrogen and helium. In the Solar System, Jupiter has a mass of $318M_E$, of which, in this model, 5 to 10% would be heavy elements. Saturn, with a mass $95M_E$, if it has a similar total mass of heavy elements, would be proportionally more enriched. Saturn is less massive than Jupiter because it formed more slowly and therefore had captured less gas when the T-Tauri phase started. Uranus and Neptune, at about $15M_E$, must have formed even more slowly, and water is their main component.

That was the standard view until the late 1990s. Then, a theory that had been around in earlier decades and had never quite vanished, had a resurgence. This is the formation of giants in *one* stage, rather than the two-stage kernel-gas capture process. Models showed that at a sufficient distance from the star, roughly beyond the ice-line, the disc might become gravitationally unstable in such a way that giant planets contract directly from it. In this case the giants will initially have the same proportion of heavy elements as the star, up to about 3%, but the subsequent capture of icy–rocky planetesimals could raise this, even to 5 to 10%. Both theories are currently viable, and perhaps both processes operate, one predominating in some circumstellar discs, the other in the rest.

In either the two-stage or one-stage process of building giant planets, they form around or beyond the ice-line, about 4 AU in the case of Sun-like stars. Therefore, if giant planets do form in either of these two ways then those giant exoplanets that are now well within the ice-line of their star must have subsequently moved inwards. It was not long after the discovery of the first of these close-in giants, 51 Pegasi b in 1995, that theoreticians found migration mechanisms that could produce the 'hot jupiters', as they are called. That these mechanisms are fairly convincing adds weight to the view that the close-in giants are hydrogen–helium in composition. The mechanisms are outlined in the next section.

QUESTION 7.6

A few of the points in Figure 7.6 are labelled with the name of the exoplanetary system. A system more like the Solar System than nearly any other system is 47 Ursae Majoris (47 UMa).

(a) Make a table of the values of the minimum masses of its planets and the semimajor axes of their orbits.

(b) Draw the orbits of its giants with the orbits of Jupiter and the Earth to the same scale. You can approximate all the orbits as circles with the star at the centre. Label each giant orbit with the minimum mass of its giant in units of M_J. Include on your drawing the present-day habitable zone boundaries for 47 UMa, which extends from about 1.0 AU to 1.9 AU. (47 UMa is similar to the Sun in mass, though at 7000 Ma it is a bit older and thus its habitable zone is further out – see the answer to Question 7.2.).

7.3 Migration of exoplanets within exoplanetary systems

The discovery in the mid 1990s of 'hot jupiters', much closer to their stars than the models of their formation predicted, led within months to plausible mechanisms by which planets could migrate inwards. Of course, if the 'hot jupiters' could, somehow, have formed near to where we find them, then migration is unnecessary, but we don't understand how they could have formed so close in. By contrast, migration, extensive or limited, now seems almost inevitable. Theoreticians were not entirely being wise after the event, because migration had been predicted over a decade earlier, but largely overlooked. Migration mechanisms are outlined in Section 7.3.1. Then, in Section 7.3.2, we examine the implications of giant-planet migration for the formation and survival of Earth-mass planets.

7.3.1 Migration mechanisms, and consequences for giants

The key to migration is the gravitational effect of the giant planet on the circumstellar disc of gas and dust. The details are complex so only a qualitative outline is given here. At first the disc is symmetrical around an axis through its centre, as shown in Figure 7.7. But as the mass of the giant planet grows, its gravitational field produces spiral structures in the disc that destroy this symmetry, as in Figure 7.8. The disc mass considerably exceeds the mass of the planets that will form from it, so there's plenty of disc left. Consider the two-stage formation of a giant at the point where its icy–rocky kernel has a mass less than M_E. The spiral structure in the disc interior to the kernel has a gravitational influence on the kernel's orbit that tends to push it outwards, whereas the spiral structure exterior to the disc exerts a gravitational influence that tends to push it inwards. For any plausible disc model, the inwards push is the greater and so the

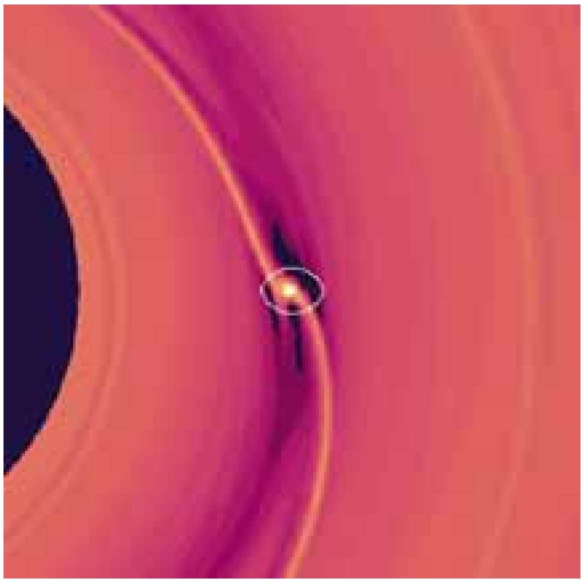

Figure 7.8 A computer model of a circumstellar disc with a spiral structure created by a planetary kernel. The disc is not modelled close to the star, hence the hole, which is an artefact. (Pawel Artymowicz)

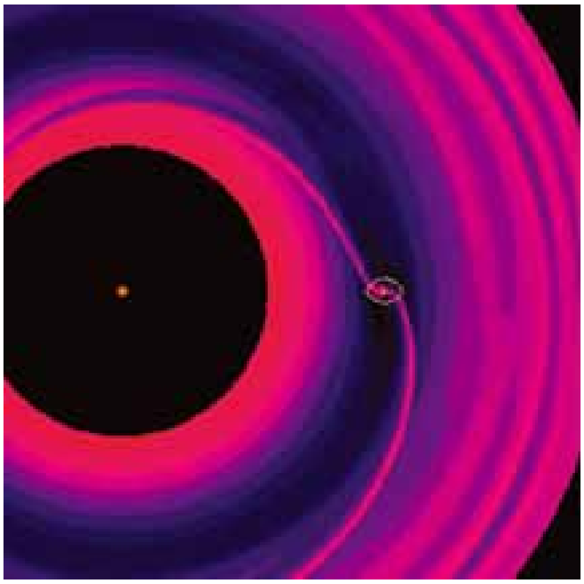

Figure 7.9 A massive kernel opens up a gap in its circumstellar disc. (Pawel Artymowicz)

net effect is inward migration. The rate of migration is proportional to the mass of the disc and also to the mass of the kernel, so as it grows it migrates inwards ever more rapidly. This is **Type I migration**. Note that the disc itself is also migrating inwards, but always more slowly than the kernel.

Type I migration continues until the kernel has grown to sufficient mass to open up a gap in the disc, as illustrated in Figure 7.9. The gap causes a major change in the migration. It now slows dramatically, by a factor of between 10 and 100, until the kernel and the disc are migrating inwards at the same rate. This is **Type II migration**. The kernel mass at which the transition takes place depends on various properties of the disc (its density, thickness, viscosity, temperature, and so on), and on the distance of the kernel from the star. An approximate range is $10M_E$ to $100M_E$, and so it is likely that there is a fully fledged giant kernel plus some captured gas at this transition. Gap formation reduces the rate at which the kernel acquires mass from the disc, but does not halt it, and so planets up to several M_J can form.

- If giant planets form in one stage, how is the story modified?
- There is no kernel build-up, and so it is likely that there will be no Type I migration – the giant enters the story at the point of a fully fledged Type II migration.

Migration needs to be halted, otherwise all the giant planets will end up in the star. Either the disc must be removed, or there must be counteracting effects. The disc will be removed partly through gradual infall to the star and partly through bursts of activity that young stars are observed to undergo, when outflowing stellar winds and intense UV radiation push the disc away (T-Tauri phase). Observations of young stars suggest that the disc lasts 1 to 10 Ma. This will be too long in some cases for the giant planet to survive, because the Type II migration time, depending on the disc properties and other parameters, can be less than this. There are however several ways in which opposing effects can appear, including tidal interactions between the planet and the star, magnetic interactions between the star and the disc, and evaporation by stellar radiation of a narrow zone in the disc a few AU from the star, creating a migration barrier. These are rather subtle effects and are peripheral to the story, so they are not detailed here, but note that they could save otherwise doomed planets.

So far we have considered a sole giant planet in a disc, yet the Solar System and at least a few of the known exosystems have more than one giant, and it is presumed that most or all of the others do too. Computer models of discs with two giants have shown that interactions between the giants can slow and even reverse Type II migration, and this is one way in which we could have ended up with a Solar System in which Jupiter and Saturn are still beyond the ice-line. It is also possible under special choices of circumstellar disc parameters to end up with limited migration without invoking giant–giant interactions.

The multiple-giant case can explain giants in highly eccentric orbits. Figure 7.10 shows the eccentricities *e* versus semimajor axis *a*, with Jupiter and Saturn for comparison. Some of these eccentricities are very high, particularly at larger *a*. It is difficult to produce such large eccentricities by formation from the circumstellar disc, without something extra. One way to get high *e* is when the orbital periods of the two giants are in a simple ratio, such as 1 : 2 or 1 : 3 – these are called orbital resonances (Box 4.2). In such a case the gravitational interaction between the giants can disturb the orbits in a cumulative way, resulting in large orbital changes. An everyday analogue is the pushing of a child on a swing – if you time your pushes correctly a large amplitude builds up.

Another way to achieve large eccentricities is through close encounters between giants. Simulations show that in each case a common outcome is that one giant is flung into the cold of interstellar space and the other is retained in a high *e* orbit. These scenarios are also the basis for an additional explanation of giants close to the star. If the surviving giant is in a high *e* orbit with a small **periastron distance** (the closest distance to its star), then tidal interactions with the star, perhaps aided by residual disc gas, will reduce *e* and we can end up with a giant in a small, low *e*, orbit.

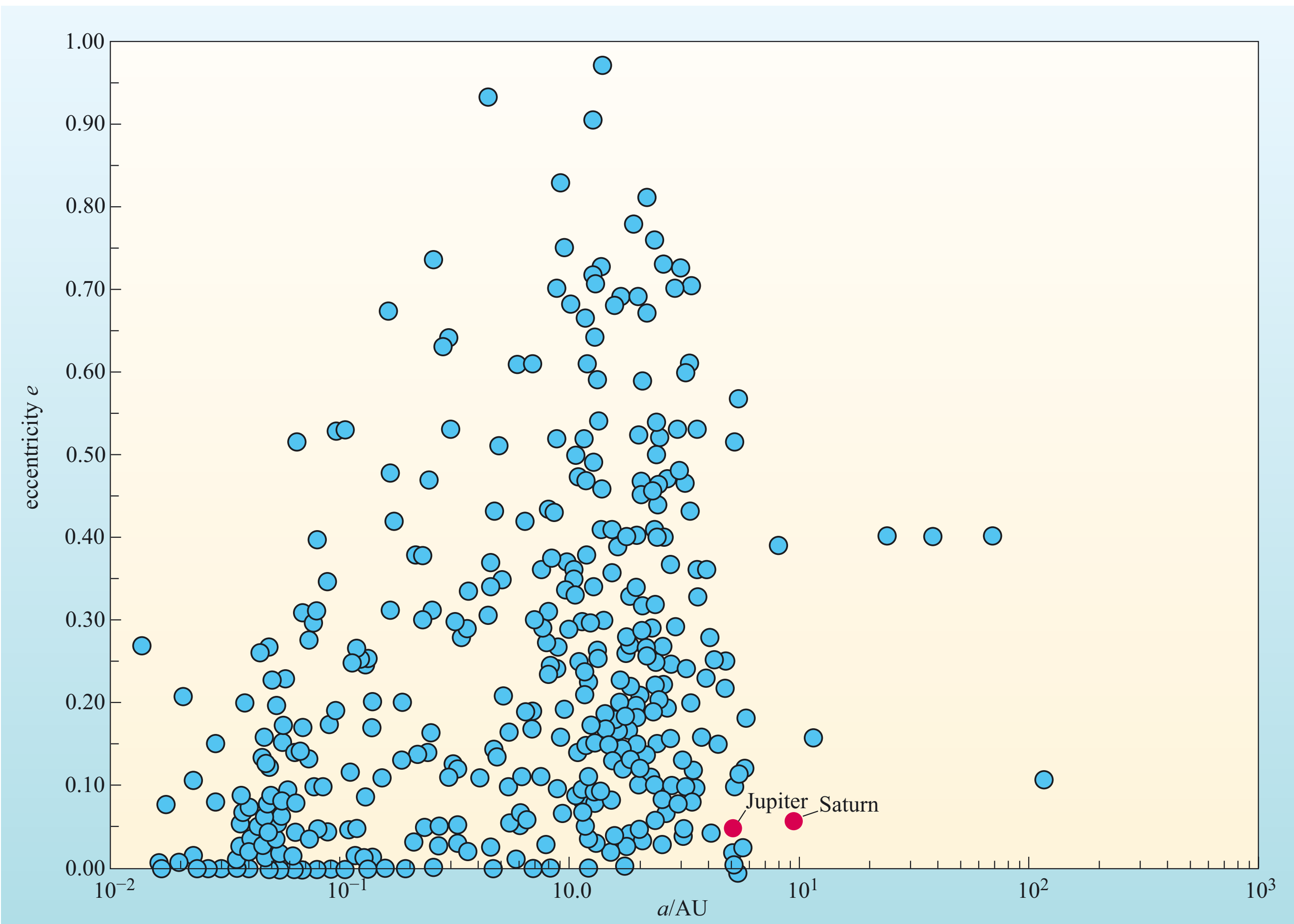

Figure 7.10 Eccentricities *e* versus semimajor axis *a* for known exoplanets in early 2011.

In conclusion, the system parameters specifying the star, disc, and the giant planets, are sufficient in number and sufficiently variable, that giant planets with a great range of masses, orbital semimajor axes, and eccentricities can result. We certainly have plausible explanations of the observed exoplanetary systems. The giant planets in our Solar System emerge as just one of many possible types of outcome. However, plausibility does not mean that it actually happened that way. There is no guarantee that the exosystems came into existence in the manner described above. The best current explanations have been given.

QUESTION 7.7

Take a look at Table 7.2 which shows some properties of two exoplanetary systems. For each one outline the sequence of events that could have led from each giant being an Earth-mass kernel beyond the ice-line to the systems we have today. (We have ignored two unconfirmed distant giant planets of Epsilon Eridani.)

Table 7.2 Properties of two exoplanetary systems.

System	$M_p \sin i_0/M_J$	Semimajor axis a/AU	Eccentricity
Epsilon Eridani	0.86	3.3	0.608
HD 168746	0.23	0.065	0.081

But what about Earth-mass planets? Could these form and survive in the habitable zones of the known exoplanetary systems? If not then the chances of finding potential habitats there are much reduced.

7.3.2 Implications of giant-planet migration for the formation and survival of Earth-mass planets

A major problem with the speedy Type I migration is that it carries Earth-mass planets rapidly into the star. Only by careful choice of disc parameters is it possible for planets with a mass of about $1M_E$ to migrate sufficiently slowly to outlast the disc and thus survive. This can seem rather contrived. Fortunately for our understanding, models indicate that the growth of $1M_E$ planets interior to the ice-line, which is where the habitable zone will lie, is likely to be rather slow, much slower than the growth of kernels beyond the ice-line. Type I migration during the 1 to 10 Ma lifetime of the circumstellar disc of gas is then slight, and when the disc disperses it is quite possible that this inner region will contain many embryos with masses up to $0.1M_E$, plus a swarm of lower-mass planetesimals from which further growth can occur.

Meanwhile, growth has been more rapid beyond the ice-line. In the two-stage process this is because the abundance of condensable water provides a swarm of planetesimals with low relative speeds. In the one-stage process it is because the disc instabilities occur early on. Subsequently, the region interior to the ice-line might or might not be traversed by a (growing) giant. If, as in the Solar System, the migration of a Jupiter equivalent has been, at most, slight, the terrestrial-planet region will not be disrupted by migration. In other systems, with more extensive migration, even if the giant(s) stops short of the habitable zone, embryos and planetesimals will be scattered as the giant moves inwards, sweeping orbital resonances across the habitable zone. If the habitable zone is traversed by a migrating giant then embryos and planetesimals will certainly be scattered. So, could Earth-mass planets form after migration has ceased, and could any such planets survive in the habitable zones today? Let's consider their formation first.

The few theoretical studies have shown that it is possible for Earth-mass planets to form in the habitable zone after giant migration is complete. In cases where the giant comes to reside beyond the habitable zone, sufficient embryos and planetesimals are left provided that the giant planet never comes close to the outer boundary of the habitable zone. In cases where the giant traverses the habitable zone and ends up parked near the star, well inwards of the interior boundary of the habitable zone, there might be so little material left that the formation of Earth-mass planets from planetesimals and embryos is very unlikely. If such 'hot jupiters' are ruled out, then in only about 10% of the exoplanetary systems known at the time of writing could Earth-mass planets have formed in the habitable zone. But should they be ruled out? There could be sufficient planetesimals left in some cases, and additionally there could be sufficient dust in the disc to form a new generation of planetesimals that lead to embryos and to an Earth-mass planet. If 'hot jupiters' are not ruled out then perhaps as many as one-half of the known exoplanetary systems could have Earth-mass planets in the habitable zone.

If, in any system, an Earth-mass planet somehow formed in the habitable zone, the next question is whether it could have *survived* there for the present age of the star. This is not the same question as that of formation, which takes less than 100 Ma, perhaps a lot less. By contrast, gravitational buffeting by the giant(s) militates against survival for the age of the star, which is typically billions of years. Computer studies have been made of representative systems. On the basis of these it can be concluded that in perhaps one-third of the known exoplanetary systems, Earth-mass planets, provided that they can form, will today still be in the habitable zone. If an Earth-mass planet does not remain confined to the habitable zone its usual fate is to be flung into the cold of interstellar space, rather than have a collision with the giant or the star.

Figure 7.11 shows the sort of regions in the habitable zone of typical systems where Earth-mass planets could survive. Figure 7.11a shows a giant planet in a low-eccentricity orbit near its star, very much interior to the habitable zone, and Figure 7.11b shows a giant in a more eccentric orbit not far outside the habitable zone. The systems awaiting discovery (Section 7.4) are likely to include ones that more resemble the Solar System, where the giants, Jupiter and Saturn, are well beyond the habitable zone (Figure 7.11c). In such systems the whole habitable zone, like ours, would be a safe harbour for Earth-mass planets.

QUESTION 7.8

Suppose that a giant planet was in an orbit that lay entirely inside the habitable zone (inside the zone, not interior to the zone).

(a) Where might an Earth-mass planet be found today in this system?

(b) Where might a planetary body be found that is actually habitable? (Remember how narrow is the definition of 'habitable zone'.)

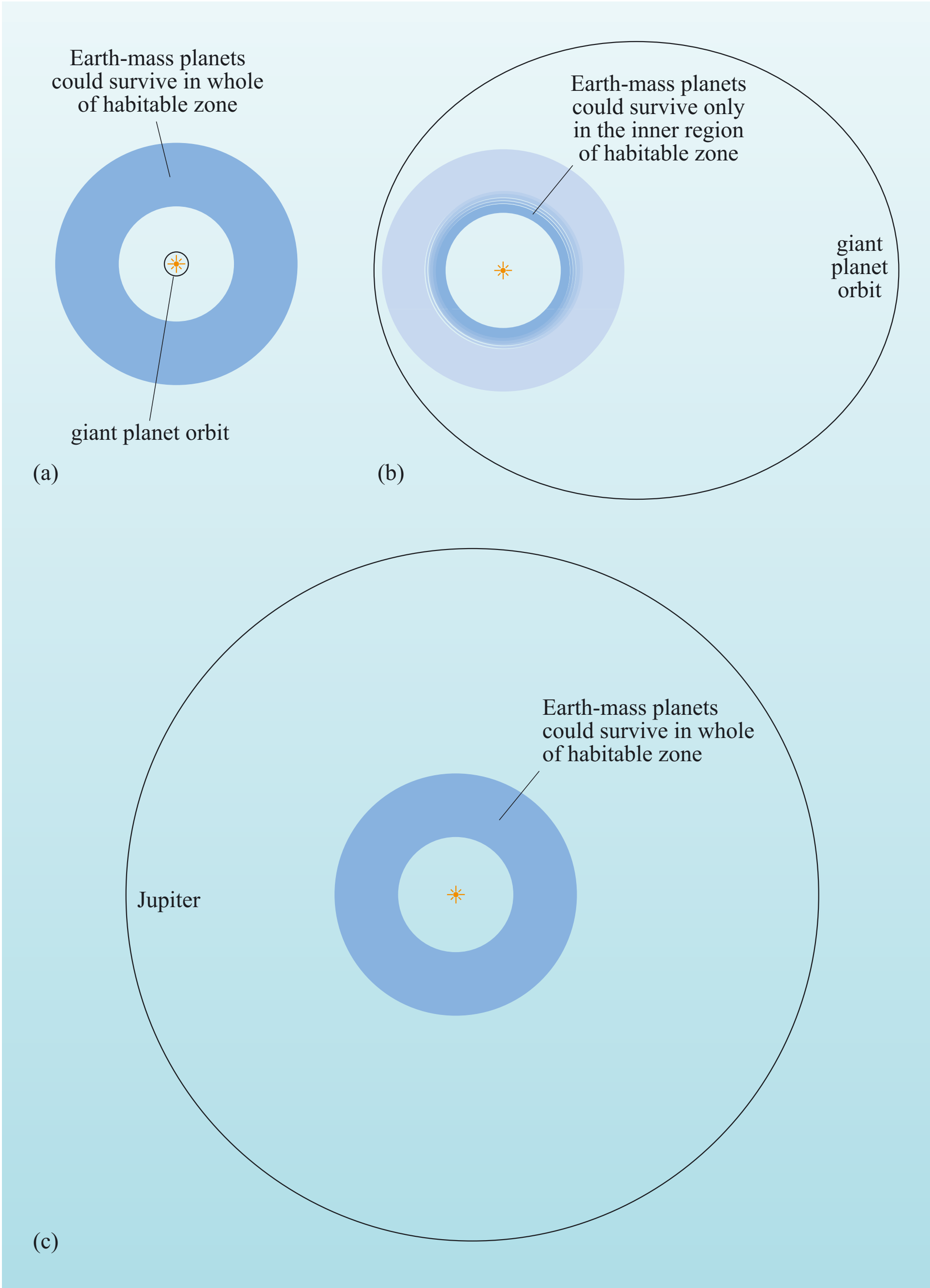

Figure 7.11 Survivable orbits for Earth-mass planets in the habitable zone where (a) the giant is very much interior to the habitable zone, (b) the giant is not far outside the habitable zone, (c) the giant is well outside the habitable zone as is the case with Jupiter in the Solar System.

7.4 The undiscovered exoplanets

7.4.1 The known exoplanetary systems – a summary

By early 2011 over 400 stars were known to have planets of which more than 50 had more than one planet each. The total of known exoplanets exceeded 500, of which about three-quarters had been discovered by the Doppler spectroscopy method. Most of the rest had been found by the transit method, and astrometry had not made any confirmed discoveries (Figure 7.6).

The properties of known exoplanetary systems can be summarized as follows:

- The most distant planets so far discovered are single planets around the stars SWEEPS-04 and SWEEPS-11 estimated to be at about 28 000 light-years distance. These were discovered by Doppler spectroscopy, and show that planets occur in the 'Galactic bulge' (towards the centre of our Galaxy, and not just in the disc where the Sun is).
- At least 10% of nearby Sun-like stars have giant planets within 4 AU of the star.
- Stars with high metallicity seem to be favoured for having planetary systems.
- As yet we have few spectroscopic observations of the atmospheres of transiting exoplanets. Most of those are hot jupiters (HD 209458 b, HD 189733 b), where (in addition to molecular hydrogen) sodium, water vapour, methane and magnesium silicate ($MgSiO_3$) condensates have been observed.
- No Earth-mass planets have yet been discovered, but we are beginning to find **super-Earths** (an informal term for an exoplanet above $1M_E$ but $< 10M_E$).
- The lowest mass exoplanets include Gliese 581 e, which has $M_p \sin i_0$ of $0.0061M_J$ ($1.9M_E$) whereas CoRoT-7 b and Kepler 10 b observed by transit as well as Doppler spectroscopy (so that $\sin i_0$ must be very close to 1) have M_p of 0.015 and $0.014M_J$ (4.8 and 4.6 M_E). Gliese 581 e is well inward of its star's habitable zone, but Gliese 581 c is a $5M_E$ ($M_p \sin i_0$) super-Earth that lies just about inside the habitable zone and may be water-covered. CoRoT-7 b is so close to its star that its day-side surface temperature must exceed 1800 °C, so that even silicates would be molten, and Kepler-10 b is not much more welcoming at >1300 °C.
- MOA-2007-BLG-192-L b is a $3.3M_E$ exoplanet orbiting a very low-mass star (a brown dwarf).
- The vast majority of known exoplanets are closer to their stars than Jupiter is to the Sun, and many of those are 'hot jupiters'.
- About half the known planets have orbital eccentricities > 0.1.
- A few exoplanets have been discovered in binary stellar systems, with the planets orbiting just one of the stars.

Figure 7.12 summarizes the sizes and radii of low-mass exoplanets known from transits by early 2011. Note that mass (and radius) are expressed relative to Earth rather than relative to Jupiter as in Figures 7.5 and 7.6. Some of these were discovered by the transit method (and then confirmed by Doppler spectroscopy), whereas others were discovered by Doppler spectroscopy and subsequently also detected by the transit method. This figure shows how close we had come to detecting Earth-like planets; note the lines indicating a theoretical relationship between radius and mass for different compositions (but beware that an apparent **waterworld** could instead have a large iron and/or silicate core surrounded by a high-mass atmosphere). Only a few planets qualify as super-Earths. Of these, only CoRoT-7 b (mass $4.8 \pm 0.8M_E$, radius $1.7 \pm 0.09R_E$), Kepler-10 b (mass

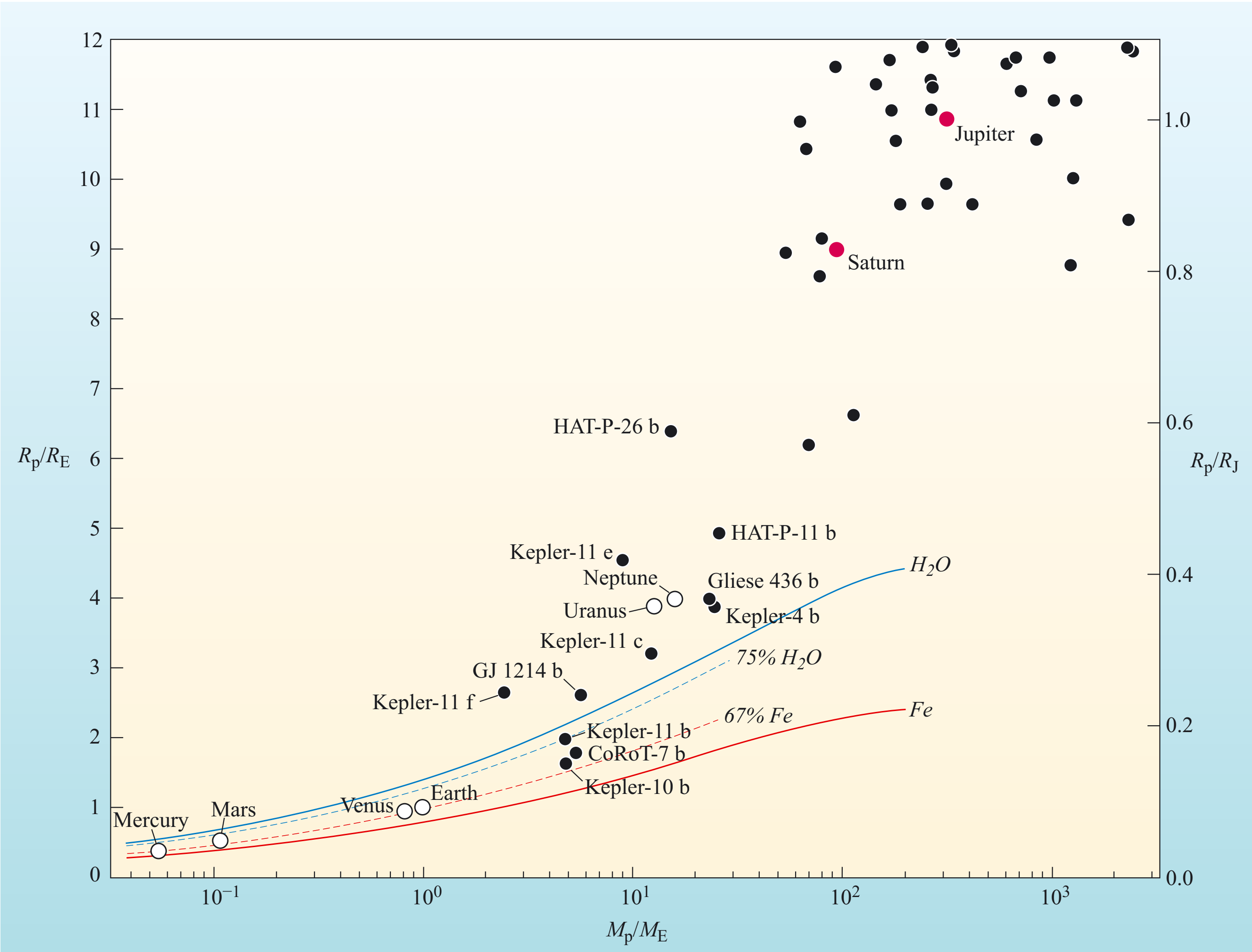

Figure 7.12 Radius of transiting exoplanets plotted against mass. The reference lines show the expected mass–radius relationships for various compositions. H_2O is pure water–ice; any planet above this line must have a substantial H/He envelope. 75% H_2O is water with 22% silicates and 3% Fe. 67% Fe is an Earth-like terrestrial planet (33% by mass silicates). Fe is pure iron; no exoplanet is expected below that line.

4.6 ± 1.3M_E , radius 1.4 ± 0.003R_E) and Kepler-11 b (mass ~ 4.3 M_E , radius ~ 2.0 R_E appear to be terrestrial planets, whereas GJ 1214 b (mass 4.8 ± 0.8M_E , radius 1.7 ± 0.09R_E) has too low a density to be a terrestrial planet and seems likely to be a 'waterworld' as hypothesized for Gliese 581 c. Note that none of the Gliese 581 planets are plotted on Figure 7.12, because transits do not occur (or, at least, have not been detected). As concluded in the answer to Question 7.3, the true mass of Gliese 581 e is unlikely to exceed 4M_E and by the same logic the true mass of Gliese 581 c is probably less than 10M_E , but we have no measures of their sizes.

7.4.2 What planets await discovery?

Exoplanets continue to be discovered, even within a few hundred light-years of the Sun. Moreover, we have not yet acquired a representative sample. In particular, observational selection effects are playing a major role. You met such effects in Chapter 6. Here we consider them further, first, those associated with the Doppler spectroscopy method that has been so fruitful.

The first selection effect arises from the need to collect enough light to display the spectral lines sufficiently clearly to measure their wavelengths accurately. This discriminates against large distances and explains why most of the exoplanetary systems discovered by Doppler spectroscopy are nearby (though it can reach further, as exemplified by SWEEPS-04 and SWEEPS-11). Another selection effect has been the preference given initially to solar-type stars – this is because they have an abundance of sharp spectral lines, and are bright (Section 7.2.1).

Other selection effects can be seen in Figure 7.13, which is derived from the equation for the observed radial velocity amplitude v_{rA} of the star, to which the Doppler shift is proportional. This equation was developed in Section 6.6.2. In Figure 7.13 v_{rA} is shown plotted against the orbital period P. Two lines are given, one for a $1M_\odot$ star, the other for a star with $0.5M_\odot$. The value used for $M_p \sin(i_0)$ is M_J in both cases – for other values note that v_{rA} is proportional to $M_p \sin(i_0)$.

■ Referring to Figure 7.13, what happens to v_{rA} if (i) $M_p \sin(i_0)$ is halved, (ii) P is decreased by a factor of 8?

❏ (i) v_{rA} is also halved (ii) v_{rA} is doubled.

So, the smaller the value of $M_p \sin i_0$ the smaller the Doppler-shift and the harder the planet is to detect, which is why so few exoplanets approaching Earth-mass have yet been detected by Doppler spectroscopy (Figure 7.6). Conversely, the smaller the value of P the greater the Doppler shift and the easier the planet is to detect. Overall, massive planets in short-period orbits are favoured by this technique. Short-period orbits have small semimajor axes, and so the Doppler spectroscopy technique favours planets close to their stars. The upper scale in Figure 7.13 gives values of the semimajor axis a for the planet for each of the two stellar masses.

Another advantage of small P is that the radial velocity goes through its cycle in a shorter time. About one orbital period of data is needed to identify a planet with reasonable certainty, and so this can be achieved more quickly the shorter the period. For a planet like Jupiter with an orbital period of 12 years, about 12 years of high quality data would be required.

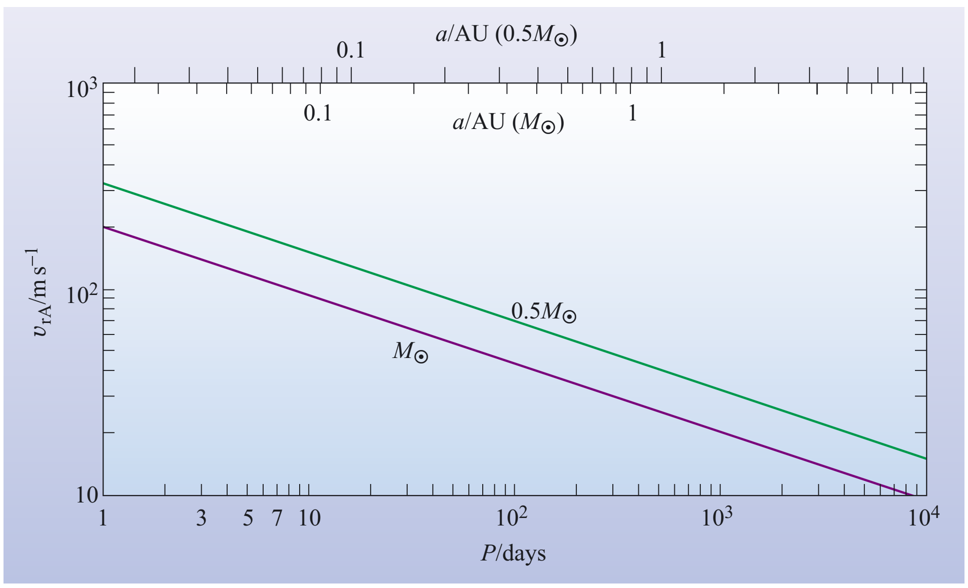

Figure 7.13 How the radial velocity amplitude v_{rA} varies with P (and a) for a solar-mass star and for a star of mass $0.5M_\odot$, using a value of $M_p \sin i_0$ of $1M_J$. Note that a is the semimajor axis of the planet's orbit.

For nearby solar-type stars, Doppler spectroscopy has discovered giant planets within 4 AU of the stars for about 10% of them. A further small proportion might also have such planets, undetected because the orbit is nearly face-on to us, thus reducing the Doppler signal. This technique has been in action long enough to begin to detect Jupiter-mass planets at the Jupiter distance of 5 AU. Radial velocities can currently be measured with a detection limit approaching 1 m s^{-1}, and this is sufficient for any Earth-mass planets around the low-mass M dwarfs, particularly in the habitable zones, which are close to such stars.

In the astrometric technique it is the periodic shift of the star's position that reveals the planet (Section 6.6.1). Figure 7.14 shows βd versus P where d is the distance to the star and β is the angular movement of the star in the sky as seen from a distance d. We give βd because then you can calculate β for any value of d. The planet has a mass M_J, and βd is proportional to this mass. One line in Figure 7.14 is for a solar-mass star and the other for a star of one-half this mass. The upper scale gives values of a for the planet. It is now the case that larger P is favoured, corresponding to larger a, though this means that a long series of observations is required.

■ For a given semimajor axis of the *star's* orbit, what happens to β if (i) d is doubled, (ii) the inclination i_0 is halved?

❑ (i) β is halved (so, distant stars are disfavoured), (ii) β is unchanged (this technique gives us the mass, not the minimum mass).

As yet astrometry has detected only one planet – Gliese 876 b, some years after its discovery. This detection was facilitated by the low stellar mass of $0.32M_\odot$, and the large mass of the giant, $3.3M_J$. Astrometry is ideal for detecting planets in large

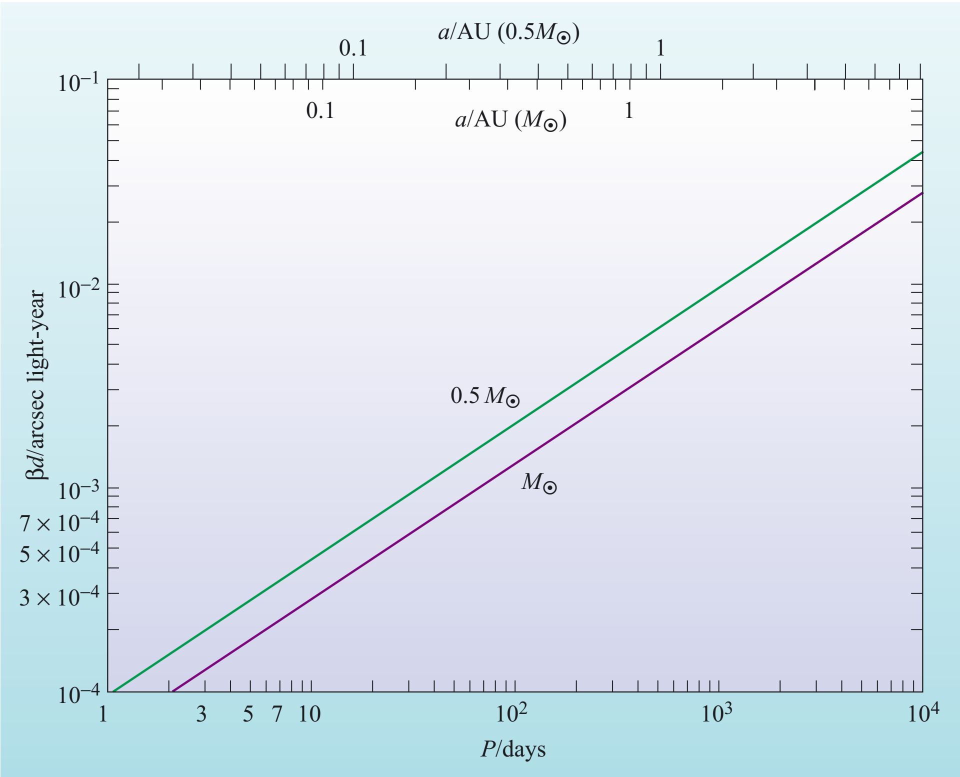

Figure 7.14 How, in astrometry, βd varies with P and a for a $1M_\odot$ star and for a star of mass $0.5M_\odot$, for a planet with a mass $1M_J$. Note that a is the semimajor axis of the planet's orbit.

orbits, and will surely do so when it acquires the required sensitivity, and as soon as observations have been made for long enough. The space telescope GAIA (Section 6.6.1) will have the giant planets in Solar System analogues within its capability out to the order of 1000 light-years. Earth-mass planets might be detectable around M dwarfs within a few tens of light-years.

Very close analogues to Jupiter (in terms of M_p and a) have already been found by Doppler spectroscopy, but no exoplanets matching Uranus and Neptune or the Earth have yet been discovered. Transit photometry, gravitational micro-lensing and direct imaging have begun to populate regions of Figure 7.6 where few planets had been found by Doppler spectroscopy. We can expect more Earth-mass planets to be found by the transit method as time goes by. However, transit photometry can only detect planets for which the orbital inclination i_0 is within a few degrees of 90°, so that the orbit is nearly edge-on to us. Moreover, the infrequency of transits and more stringent limits on i_0 for larger orbits means that the transit method favours detection of planets in smaller orbits (like Kepler-11 f) over those in larger orbits.

Though we have an increasing capability to detect Earth-mass planets, we will, of course, only detect them if they are there. Recall that theoretical studies have shown the following.

- Migration of giant planets makes it less likely that Earth-mass planets could form in the habitable zone of the star, though 10 to 50% of the known exoplanetary systems could have such planets.
- Perhaps in one-third of the known exoplanetary systems, Earth-mass planets, provided that they can form, might today still be in the habitable zones.

We can thus expect to find Earth-mass planets, even in habitable zones, though perhaps not commonly in those systems where the giant planets are within the ice-line.

To summarize, we expect the proportion of stars known to have planets to rise, perhaps considerably. Among these new discoveries there should be lower mass planets, and planets in larger orbits. At present the Solar System stands alone with its giants in large, low eccentricity orbits, and its whole habitable zone is consequently a particularly secure abode for the formation and survival of Earth-mass planets. We expect systems resembling the Solar System to be discovered, but we do not know whether they will be revealed as common. Simulations indicate that though the Solar System is probably unusual, systems like it should comprise an appreciable minority of exoplanetary systems. By about 2020 we should know if this is actually the case.

So, what of the two parameters in the Drake equation: the probability p_p of planets forming around a suitable star, and the average number n_E of suitable planets in a habitable zone per planetary system? It is not yet possible to put precise values on these parameters, but we do know that p_p is not vanishingly small, and could be as high as several tens of percent. More speculatively, it is also likely that n_E is not vanishingly small, and could be at least one-third, i.e. one per three systems, on average.

In the next chapter we anticipate the discovery of suitable planets and ask how we could determine if they are habitable, even inhabited.

QUESTION 7.9

Consider the case of an Earth-mass planet 1 AU from a solar-type star.

(a) In a few sentences, discuss whether the planet is in the habitable zone of its star.

(b) With the help of Figure 7.13 (which is for a $1M_J$ planet), calculate the radial-velocity amplitude v_{rA} of the star due to the Earth-mass planet ($M_J = 318M_E$). Hence decide whether this planet could be detected. Assume a detection limit on v_{rA} of 1 m s^{-1}.

(c) With the help of Figure 7.14 (which is also for a $1M_J$ planet), calculate βd due to the Earth-mass planet. Calculate the stellar distance d (in light-years) at which the star's motion could be detected, assuming an astrometric detection limit of 0.001 arcsec (which is the width of a finger from a distance of about 3000 km). Are there any stars closer to us than this?

7.4.3 A note on evidence from circumstellar discs

Information on the abundance and nature of exoplanetary systems can also be gathered from observations of circumstellar dust discs, which can be observed most readily by infrared telescopes either on the ground (e.g. Figure 6.6) or in orbit. About 15% of nearby Sun-like stars have been found to have dust discs, many of which may be analogous to our own Kuiper Belt.

The simple presence of a dust disc suggests the likelihood of planet formation, whereas structure, such as clumping, in the dust disc indicates the presence of planet-mass objects capable of perturbing the disc. Epsilon Eridani, a class K2 main sequence star that you met in Section 7.2.2, is one of the nearest and best-studied such systems. Its possible structure is illustrated in Figure 7.15.

QUESTION 7.10

Table 7.3 contains data on *undiscovered* planets of two main sequence stars in circular orbits. Consider the data carefully and then answer parts (i)–(v).

Table 7.3 For use with Question 7.10.

	Mass of planet M_p/M_J	Orbital inclination i_0/degrees	Orbital period of planet/days	Mass of star $M_*/M_\odot$	Distance of star/light-years
A	0.0030	89.9	200	1.0	25
B	2.0	30	2000	0.50	250

(i) Obtain the semimajor axis of the planets' orbits, expressing your answers in AU.

(ii) With aid of Figure 7.13 obtain the radial velocity amplitudes v_{rA} that would be observed in each case, expressing your answer in metres per second.

(iii) With the aid of Figure 7.14, obtain the angular movement in the sky β of the stars' positions that would be observed, expressing your answer in arc seconds.

(iv) State whether each planet is likely to be discovered by Doppler spectroscopy, astrometry, or transit photometry, or by more than one technique, giving reasons. You can assume that the detection limits for each technique are, respectively, 1 m s^{-1}, 0.001 arcsec, 0.1% apparent luminosity change.

(v) Discuss whether either planet is likely to be habitable, assuming that the star has an age of 5000 Ma in each case, and that residence in the habitable zone is a requirement.

QUESTION 7.11

The star CoRoT-7 has a second planet, CoRoT-7 c discovered by Doppler spectroscopy, and not yet detected by transit photometry. Its mass, $M_\mathrm{p} \sin i_0$, is estimated at $0.026M_\mathrm{J}$ and the semimajor axis of its orbit a is estimated as 0.046 AU.

(a) What is its $M_\mathrm{p} \sin i_0$, expressed in Earth-masses, and what can you say about this on the reasonable assumption that the orbits of CoRoT-7 c and CoRoT-7 b are almost co-planar (i.e. both close to CoRoT-7's ecliptic plane)?

(b) Does CoRoT-7 c orbit closer to, or further from, its star than CoRoT-7 b?

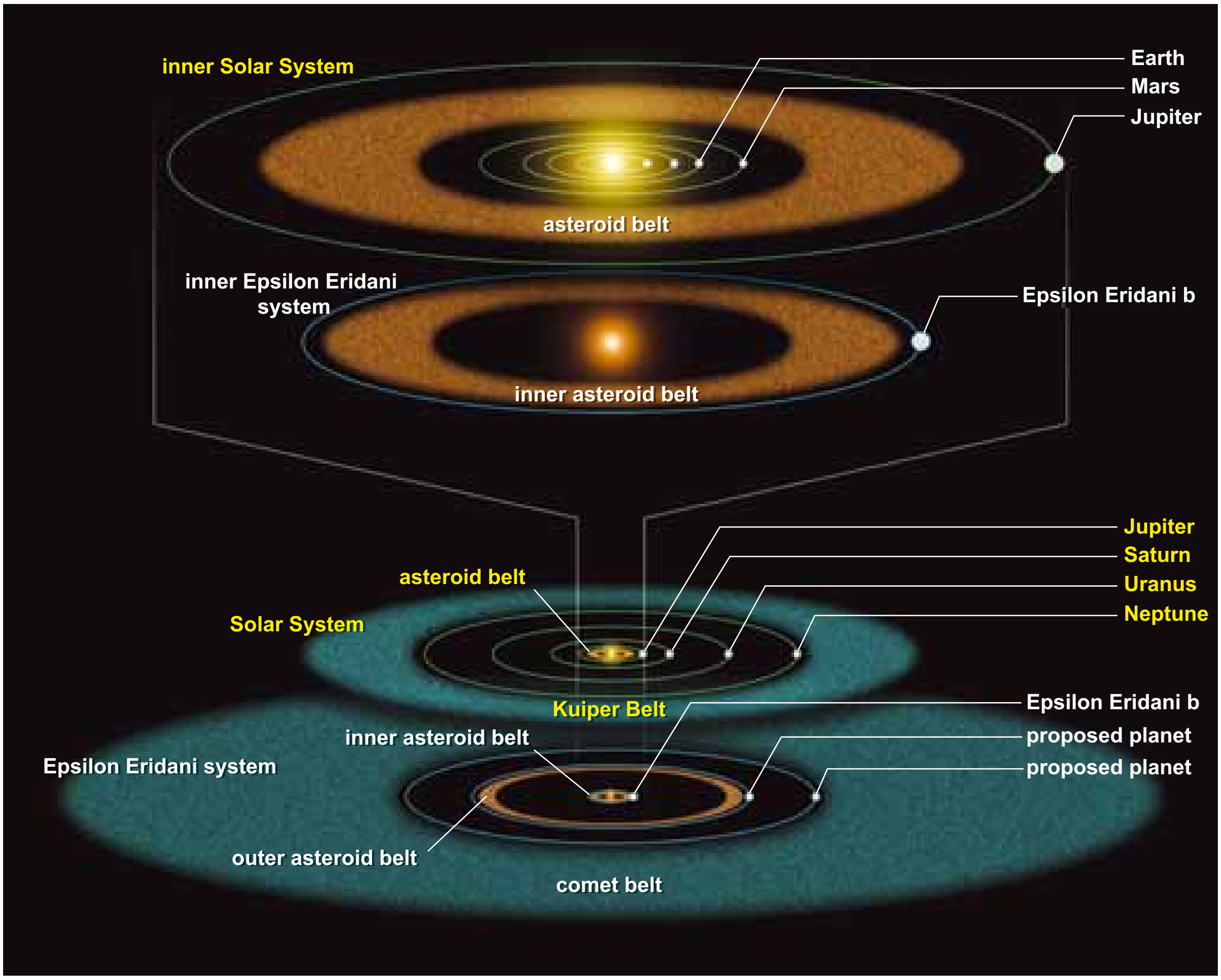

Figure 7.15 The Epsilon Eridani system, compared with the Solar System at the same scale. Epsilon Eridani b is known from Doppler spectroscopy. Features labelled inner and outer asteroid belt and comet belt have been observed by various infrared telescopes, whereas the two proposed planets are inferred from the structure of those belts. (NASA/JPL-Caltech)

7.5 Summary of Chapter 7

Much of the data here apply to mid-2011, and are subject to change, even in the short-term.

- Most of the Sun-like stars within a few hundred light-years of the Sun have been observed by Doppler spectroscopy long enough to reveal any planets of the order of Jupiter's mass in orbits within about 4 AU of the star. A proportion of stars less like the Sun have also been observed, particularly the abundant M dwarfs.
- Doppler spectroscopy has detected about 500 planets, and transit photometry has detected about 140 planets. The total number of confirmed exoplanets is about 560, including about 60 systems with more than one exoplanet.
- About 10% of the nearby Sun-like stars have giant planets within 4 AU of the star. A further small proportion might also have such planets, undetected, because the orbit is nearly face-on to us, thus reducing the Doppler signal.
- Stars with high metallicity seem to be favoured for having planets.
- The least massive exoplanet detected by Doppler spectroscopy has $M_p \sin i_0$ of $0.0061M_J$ ($1.9M_E$) whereas the least massive exoplanet yet seen by transit photometry has M_p of $0.0072M_J$ ($2.3M_E$).
- About a third of known exoplanets have orbital semimajor axis $a < 0.1$ AU, and only about half have $a > 1$AU. These statistics, notably the preponderance of 'hot jupiters', are heavily influenced by selection effects, because both Doppler spectroscopy and transit photometry favour detection of large planets in close orbits.
- About half the known exoplanets have orbital eccentricity $e > 0.1$, including about 15% for which $e > 0.4$.
- Migration of giant planets makes it less likely that Earth-mass planets could form in the habitable zone of the star, though at least 10% of the known exoplanetary systems could have such planets.
- Perhaps in one-third of the known exoplanetary systems, Earth-mass planets, provided that they can form, could still be in the habitable zones.
- In the exosystems yet to be discovered there is likely to be a small but nevertheless significant proportion that more resemble the Solar System, with the giants well beyond the habitable zone and Earth-mass planets safe within the habitable zone.
- In the Drake equation, the probability p_p of planets forming around a suitable star is not vanishingly small, and could be as high as a few tens of percent. The average number n_E of suitable planets in a habitable zone per planetary system is not vanishingly small, and (speculatively) could be at least one-third, i.e. one per three systems, on average. (The rate R at which suitable stars are born, was estimated in Chapter 6 as about 30 per year in our Galaxy.)