

---

*Fundamentals of*

# **EXOPLANETS and ASTROBIOLOGY**

**From the origin of the universe, stars and elements,  
and the creation of solar systems and exoplanets,  
to the arise and development of life**

*Uffe Gråe Jørgensen, March 2017*



## **Part 4: Exoplanets.**

for lectures 2017, March 14 and 16.

# 1

## Exoplanets

### — planets around other stars than the Sun

#### 1.1 Planets, exoplanets, and “exo-solar-systems”

Astronomy is a very old science, maybe the oldest. Many concepts are therefore introduced in the antique and would have been defined completely different if they were introduced today. This concerns also the concept *a planet*. Originally it meant a star-like wanderer, i.e. an object that moved on the sky. This definition is obviously no longer useful. Our knowledge and intuition tells us that planets are larger bodies that move around the Sun (or another star, in which case we also call them exoplanets). During recent years it has become clear that our solar system harbours hundreds of thousands of other objects than planets. Many of them are quite large (for example it is now believed that there are more than 35,000 planetary like objects in orbits beyond Neptune which have radii larger than 100 km), and a few are even larger than the smallest planet, Mercury. They got all kind of names, dependent on where in the solar system they were discovered; centaurs, asteroids, near-Earth objects, Kuiper-belt objects, trans-Neptunian objects, Oort-cloud comets, etc, and were for some times collectively referred to as solar system minor bodies. In 2006 members of the International Astronomical Union decided, in a somewhat dramatic vote, to change the definition, such that only the 8 historical planets are now called planets, defined physically as solar system bodies those gravity is large enough to shape them round, and which has cleared the region of their orbit for other solar system bodies. Other objects those gravity are large enough to shape them round, but not large enough to have (or for other reasons haven't) cleared the region of their orbit for other solar system bodies, are called dwarf planets. This group include Pluto, Ceres Eri, Makemake, and Haumea, and is likely to become very large in number as more and more large objects beyond Neptune are discovered. The remaining smaller solar system objects that are not just “rocks” or dust are called “small solar system bodies”.

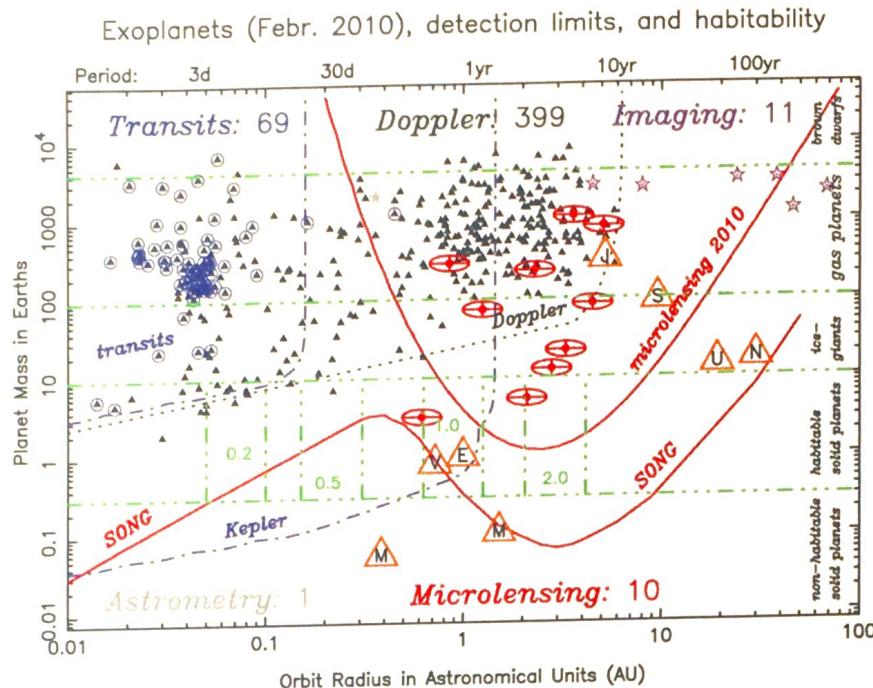
In the other end of the mass scale, it has been debated how large a planet can be before it can no longer be called a planet. Jupiter and Saturn radiates almost two times more energy than they receive from the Sun. Planets are therefore not just major bodies that reflects starlight, as opposed to stars that produce it intrinsically, as our intuition maybe would tell us to be a good definition. For a long time, most astronomers were tempted to distinguish stars and planets by the standard theory for the solar system formation that we discussed in the previous chapter. Then a star is the gaseous body that formed out of the contracting cloud by a gas-collapse (in the center in the case of single stars as the Sun, or at several places of the cloud in the case of binary or multiple stars), while planets are the (larger) bodies that form from dust condensation in the surrounding nebula, with a possible nebula collapse on top of the solid material (which then qualify

them to be gas-planets). However, part of the astronomical community is of the opinion that planets do not form from condensation, but from a sub-collapse in the nebula. If some, or all, of the planets are really formed in this way, then there is no principal difference in the formation process of stars and planets, and the formation process would then not be a good basis for the definition of what distinguish a planet from a star. Furthermore, we are of course not likely in any foreseeable future to be able to find a criterion we can use to see whether a given exoplanet candidate is formed from condensation or from gas collapse, so this alone makes the definition un-practical. Most astronomers therefore, for the moment, tend to lean to the practical definition of a planet being a larger object in space those mass is smaller than 13 times the mass of Jupiter. If it orbits a star (other than the Sun) it is called an exoplanet, and if it is not in orbit, it is called a free-floating planet. There is no need for calling it a free-floating exoplanet, since there are no known free-floating planets connected to our solar system. It will of course be impossible to know whether an exoplanet is round. It depends not only on its mass, but also on its composition. It will also be difficult to determine whether it cleared its orbit. Most of the exoplanets we know of seems to have changed orbit since they formed. With the present techniques, we are unlikely to detect exoplanets in the dwarf planet or minor body mass regime any time soon, and it seems most reasonable (in spite of the IAU 2006 definitions) to talk only about exoplanets, and not exo-dwarf-planets, even after we will discover exoplanets of quite small masses. Asteroids and comets were not given a definition in the IAU 2006 resolution, but we will be able to detect extrasolar analogues to the solar system asteroid belt within a foreseeable future, and we may already have seen comets orbiting other stars (for example the asymmetric ring around  $\epsilon$  Eridani), and it would make sense to term them extrasolar comets and extrasolar asteroid(belt)s.

If an object is between 13 and 80 Jupiter-masses ( $M_J$ ) we call it a brown dwarf, and if it is larger than 80 times the mass of Jupiter we call it a star. The  $80 M_J$  limit reflects the limit above which stable hydrogen burning will balance the gravitation, and thus stop the contraction for some length of time. The exact mass limits depend somewhat on metallicity. The  $13 M_J$  reflects the mass limit above which deuterium will burn in the center of the object. The  $13 M_J$  limit has the further advantage that the mass function seems to increase below  $13 M_J$  (i.e., there are more and more exoplanets when we go to lower masses), while there are very few objects with masses close to  $13 M_J$ , but again an increasing mass function toward higher masses above the  $13 M_J$  limit. As a result, there seems to be a difference between the population of objects below and above  $\sim 13 M_J$ , and it is likely that we will encounter only very few objects that are mis-classified as being planets or brown dwarfs in a physical sense, following this simple definition.

In the following we will assume that we know what a planet and an exoplanet is (i.e., that there is a workable definition). We will use the standard IAU nomenclature *star name* followed by the letter *b*, *c*, *d*, etc. to indicate the known planets orbiting the star. For example, 55 Cnc *b* is the first planet that was discovered around the star 55 Cnc, while 55 Cnc *c*, *d*, *e* were discovered during the subsequent years. If 55 Cnc had been a binary star, we would have called the two stellar components for 55 Cnc A and 55 Cnc B (capital A and B). For exoplanets discovered by use of the microlensing technique, there is a particular nomenclature challenge. As will be clear from the description in the following section about the microlensing technique, the name given to a microlensing event refers to the combined light from the background star (or stars, plus its possible planets) and the foreground star (or stars, plus its possible planets) that magnifies the light from the background star. When the first terrestrial-like exoplanet was discovered in connection with the gravitational microlensing candidate OGLE-BLG-05-390, it was therefore suggested to call the corresponding exoplanet (orbiting the invisible foreground star) OGLE-BLG-05-390Lb and not just OGLE-BLG-05-390b, in order to indicate that it orbited the foreground (i.e., lensing) star and not the background star. Had it been orbiting the background star it would probably have been called OGLE-BLG-05-390Sb, but we know of no exoplanets associated with the source stars in microlensing events yet.

During millenias humans have wondered whether we were alone in the universe, during the last almost hundred years astronomers have seriously discussed techniques that could in principle reveal whether there were planets like our Earth among the myriads of stars in the sky and the billions of stars in our Galaxy, and during the last two decades we have gone from the first actual discovery of a planet orbiting another star than the Sun to knowledge about the existence of thousands of exoplanets. But the field is still in its infancy in the sense that only patchy areas of the parameter space are reachable with existing instruments and techniques, so the coming years will see an explosive effort to understand how planets are formed and where they are,



**Figure 1.1.** Comparison of the region in the planetary mass versus orbit diagram, that the different exoplanetary search techniques and specific (existing or proposed) projects are most sensitive to. The position of the planets in our own solar system are marked by the letters M, V, E, M, J, S, U and N.

in the form of increasing investment in manpower and development of the techniques and instruments on the ground and in space that will be able to track down the existence of the kind of planets we still cannot easily reach with our present instruments – and this include most of the kind of planets we have in our own solar system, and which we used to assume was the norm rather than the exception.

We will start with discussing the different techniques that has been invented to search for exoplanets (= extra solar planets = planets orbiting other stars than the Sun), and then discuss the results and perspectives of these searches. Each of the techniques have their intrinsic strengths and weaknesses, and to some extend complement each other, in terms of which combination of mass and orbital radius they are most sensitive to, as is illustrated in Fig. 1.1. The radial velocity technique measures the gravitational "wobble" the planet gives the star it orbits, the transit technique measures the dimming of the stellar light that appears when a planet moves in-between the star and us, and the microlensing technique measures the anomaly in the light curve of background stars when they pass behind stars with planets relative to when they pass behind stars without planets. Of the 1944 exoplanets that by August 14, 2015, were listed in the authoritative web catalog exoplanet.eu, 609 were discovered by the radial velocity technique. Transit detected planets counted for 1214 of the listed discoveries. Of the ground based observed potential exoplanet transits (i.e., that observations have shown some kind of periodic dip in the light curve that could look like a transit by an exoplanet), experience has shown that only 1 out of 10 are actually due to a planetary transit. The exoplanet.eu data base therefore requires exoplanet transit candidates to have been confirmed by another technique (usually the radial velocity method) before they are listed as new exoplanets. If one were including the many exoplanet transit candidates discovered from the Kepler satellite (of which probably more than 90% are real exoplanets) but not yet confirmed by other techniques, the total number of transit exoplanets exceeds 3000, and therefore count for the majority of known planets. This illustrates that it is no longer meaningful to give a precise number of known exoplanets, since it depends very much on the definition of candidates and confirmed, so all we can really say is that we know a few thousand exoplanets now. Exoplanets listed in exoplanet.eu, as discovered by the direct imaging method counted by August 2015 for 60 of the listed 1946 exoplanets,

while microlensing observations had revealed 38, timing 19 (including planet sized clumps of mass orbiting pulsars, whose origin may not be planetary), and transit timing variation 4. In the future all the techniques will no doubt reveal many more new exoplanets, and most likely the majority of the known exoplanets will in a few years have come from the Gaia astrometric survey that is predicted to be able to reveal 10,000 or more planets of a similar type to those seen by the radial velocity method.

Even planets in no orbits, so-called free-floating planets, are postulated. The direct image of a free floating planet obviously looks just like a dot, and can only be understood as a planet (rather than a star) if it is seen in a cluster as an object of lower luminosity than the smallest stars (and brown dwarfs), but it has always been difficult from such observations to be sure that they really are free floating planets in the cluster rather than more distant background stars not physically belonging to the cluster. Free floating planets have, however, also been inferred from the gravitational effect they in a few cases seems to have had on the light from passing background stars, and if this interpretation of the light changes are correct, it may indicate that free floating planets are more common than planets orbiting stars – a conclusion not unexpected from theoretical considerations about the formation of planetary systems. They may have formed as stars, from a collapsing gas cloud, or they may be planets that once were in orbit around a star, but were expelled in a gravitational interaction with another star. The fact that a number of exoplanets can be shown to orbit their host star the “wrong way around” (i.e., retrograde) could be understood as some stars having captured free floating planets expelled from nearby other stars’ exoplanetary system, thereby indirectly supporting the idea that it is normal for stars to eject a major fraction of their original planetary system. Some theories predict that our own solar system once had several more large gas planets that were expelled from our solar system shortly after its formation, and now exist as free floating planets somewhere in the interstellar neighbourhood.

Fig. 1.1 compares the orbit-mass sensitivity regions of the methods based on the Doppler, transit, microlensing, and astrometric techniques. In the following sections we will describe in some more detail the radial velocity technique, the transit method, the astrometric method, and the microlensing technique. Each of them have their strengths and weaknesses, and Fig. 1.1 illustrates how they complement each other in terms of the parameter space of planetary masses and orbits that they are most sensitive to.

## 1.2 The radial velocity method

From a baryonic point of view, the solar system is basically a two body system, with 99.9% of the mass ( $M_{\odot} = 1.989 \cdot 10^{33}$  g) in the Sun, and 0.1% of the mass ( $M_J = 1.899 \cdot 10^{30}$  g) in Jupiter, with Jupiter orbiting the Sun. Seen as a binary system, the distances  $r_{\odot}$  and  $r_J$  of the mass center of the Sun and Jupiter from their common center of gravity is therefore given by

$$M_{\odot} \cdot r_{\odot} = M_J \cdot r_J \Rightarrow r_{\odot} = 1.07 R_{\odot} \quad (1.1)$$

where  $R_{\odot} = 6.96 \cdot 10^5$  km is the solar (mean) radius; i.e. that the center of gravity of the Sun and Jupiter is only 50,000 km above the solar surface.

The Sun (and Jupiter) moves with a period  $P_{\odot} = P_J = 11.9$  years around their common center of gravity, which gives the Sun an orbital velocity of

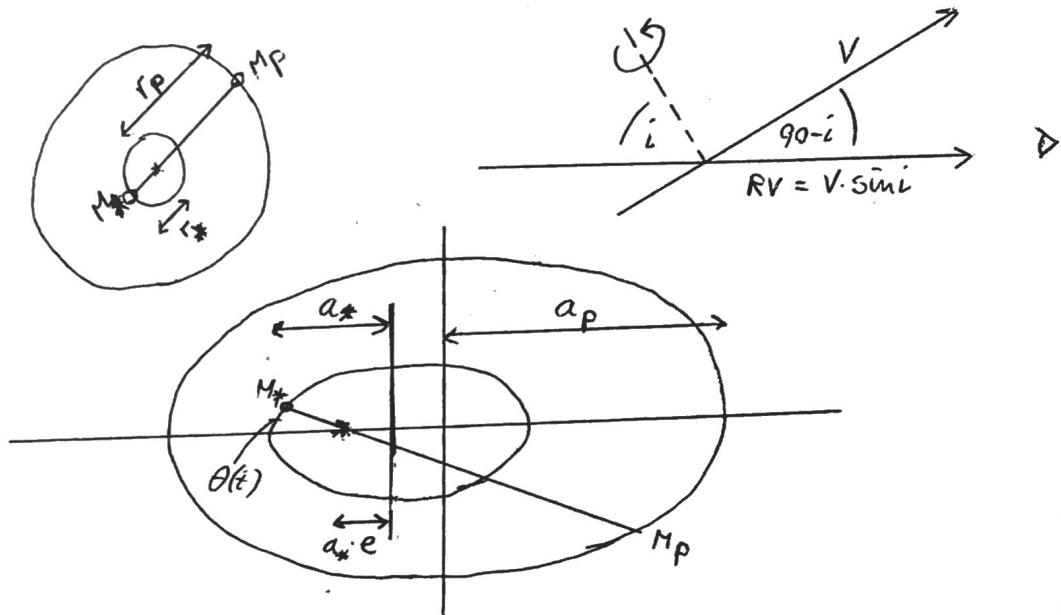
$$V_{\odot} = 2\pi r_{\odot}/P_J = 12.5 \text{ m/s} \quad (1.2)$$

If we stood on a planet orbiting another star in our Galaxy, we would be able to measure a Doppler motion of the Sun due to its motion around the common center of gravity of the Sun and Jupiter, those amplitude (i.e. the measured radial velocity semi amplitude,  $\Delta RV = (RV_{max} - RV_{min})/2$ ) would be

$$\Delta RV_{\odot} = V_{\odot} \cdot \sin i \quad (1.3)$$

where  $i$  is the inclination angle of the Sun-Jupiter plane relative the observer. If we saw the system “edge on” (i.e.,  $\sin i = 1$ )  $\Delta RV_{\odot}$  and  $V_{\odot}$  would be identical, whereas a system seen “from the top” would have  $\Delta RV_{\odot}=0$ .

## 1. EXOPLANETS



**Figure 1.2.** The geometry of a star and a planet orbiting their common center of gravity in circular orbits and in elliptical orbits, and the relation between inclination  $i$ , radial velocity  $RV$  and orbital velocity  $V$ .

Similar estimates for the second largest planet, Saturn, with mass  $M_S = 0.3M_J = 0.0003M_\odot$ , and orbital period  $P_S = 29.4$  years, gives the Sun a  $V_\odot = 2.7$  m/s, while the Earth imposes  $V_\odot = 9$  cm/s.

Generally, we will of course not know the exoplanetary orbital radii (corresponding to  $r_J$  in Eq. 1.1), and we would need to apply Kepler's 3<sup>rd</sup> law,

$$\frac{P^2}{r^3} = \frac{4\pi^2}{G(M_* + M_p)} \quad (1.4)$$

where  $r = r_* + r_p$  is the sum of the orbital radii (for circular orbits, or the sum of the two semi major axes for elliptical orbits) of respectively the star and the planet, around their common center of gravity.  $M_*$  and  $M_p$  are the masses of the star and the planet, and  $G$  the gravitational constant.  $P$  is the (common) orbital period.

If we for the moment assume circular orbits, then the geometry in Fig. 1.2 combined with Eq. 1.1 give us

$$(r_* + r_p)^3 = r_*^3(1 + r_p/r_*)^3 = r_*^3((M_p + M_*)/M_p)^3 \quad (1.5)$$

and the orbital period

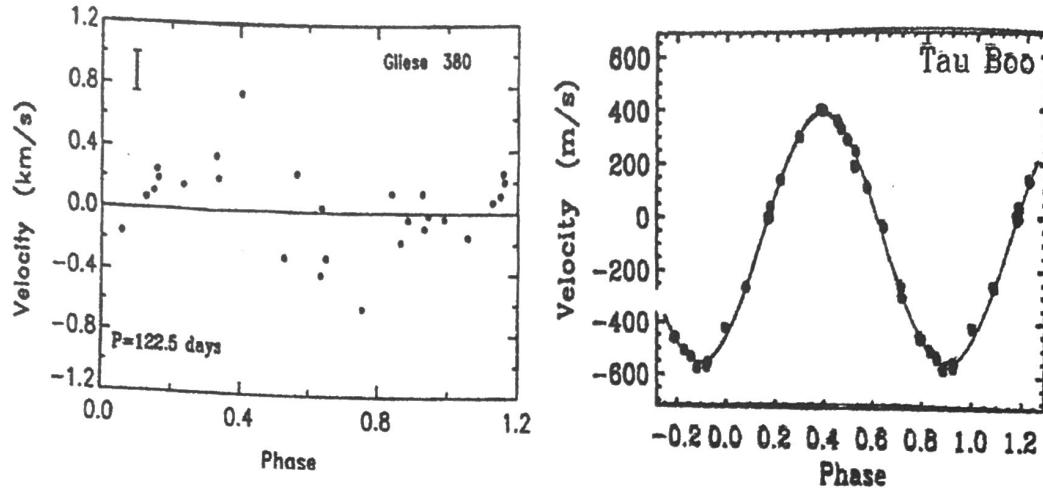
$$P_* = \frac{2\pi r_*}{v_*} = \frac{2\pi r_* \sin i}{\Delta RV} \quad (1.6)$$

Inserting Eq. 1.5 and 1.6 into Kepler's law (Eq. 1.4 using the form  $P^2/r^3 = P^3/(P \cdot r^3)$ ), give us

$$M_p \sin i = \left( \frac{P_* M_*^2 (\Delta RV)^3}{2\pi G} \right)^{1/3} \quad (1.7)$$

where we have used  $M_p + M_* \approx M_*$ .

If now  $M_*$  can be determined from spectral classification of the star, and  $P_*$  has been measured (the sinusoidal period) together with  $RV$  (the amplitude of the sinus curve), then  $M_p \sin i$  can be calculated from



**Figure 1.3.** Comparison of attempts to detect exoplanets by radial velocity measurements in the early 1960's (left panel) with present day planetary detections (right panel). In the left panel the uncertainty in the individual measurements is given by the vertical bar, while the uncertainty in the right panel is smaller than the size of the dots. The scales on the axes of the two plots are almost identical (actually even a factor two larger in the left panel). The revolutionary change from the early 1960's to today in the ability to detect exoplanets, from measurements of the stellar radial velocity, is the enormous technological improvement in the accuracy of the measurements.

Eq. 1.7. Remark, however, that we can only determine  $M_p \sin i$ , not the planetary mass  $M_p$  itself. If we cannot determine the inclination  $i$  from other methods, we will therefore only determine a lower limit of the planetary mass. However, the expectation value of  $i$  for randomly oriented systems (see Eq. 1.18 and the associated Fig. 1.13 connected to the description of transiting exoplanets) is such that 87% of the planets have a real mass between 1 and 2 times the measured  $M \sin i$ . On average the real mass is  $1.27 \times M \sin i$ , and statistically the real mass will therefore most often be close to the measured value of  $M_p \sin i$ , but we have to have in mind that for individual systems the real value of  $M_p$  can be considerably larger than the measured value of  $M \sin i$ , and we cannot exclude that a specific system harbours a brown dwarf rather than a predicted planet, unless other constraints exclude this.

Historically, the search for the sinusoidal Doppler motion caused by an orbiting planet ("the reflex motion" or "stellar wobble" as it is often termed), has taken place throughout most of the second half of the 20<sup>th</sup> century, but during most of this time the accuracy and stability of the spectrographs were too poor for reliable detections, and several suggested exoplanet discoveries were never generally accepted in the community as planetary candidates. Fig. 1.3 compare a (erroneously) claimed planetary candidate from a scientific paper of early 1960's, with a series of recent measurements of a real exoplanet.

A typical spectral accuracy in the 1960's was a few hundred m/s, but steadily improving to 10 m/s by the end of the 1980's, and reaching values below 1 m/s today. With the coming *ESPRESSO* instrument at the VLT telescope it is the hope achieve accuracies of 10 cm/s in 20 minutes exposures of V=7.5 mag stars, which would allow to survey the nearest  $\sim 100$  stars to detect planets down to a few earth-masses out to orbital radii of  $\sim 1$  AU, and possibly reach  $1 M_\oplus$  planets in the habitable zone around a few M or K dwarfs. In the mid 2020's it may be possible to obtain  $\sim 1$  cm/s accuracy with the planned *CODEX* instrument on the E-ELT telescope. As a rule of thumb an accuracy of approximately a factor 4 better than the radial velocity amplitude is needed in order to identify the wobble from observations spanning one orbital period. I.e., around 1990 one could, for the first time, realistically expect to be able to identify the effect on a solar type star of a planet of say 3 times Jupiter's mass orbiting in a Jupiter-like orbit, but only after having assembled radial velocity data of the same quality for additionally 10 to 20 years. By that time,

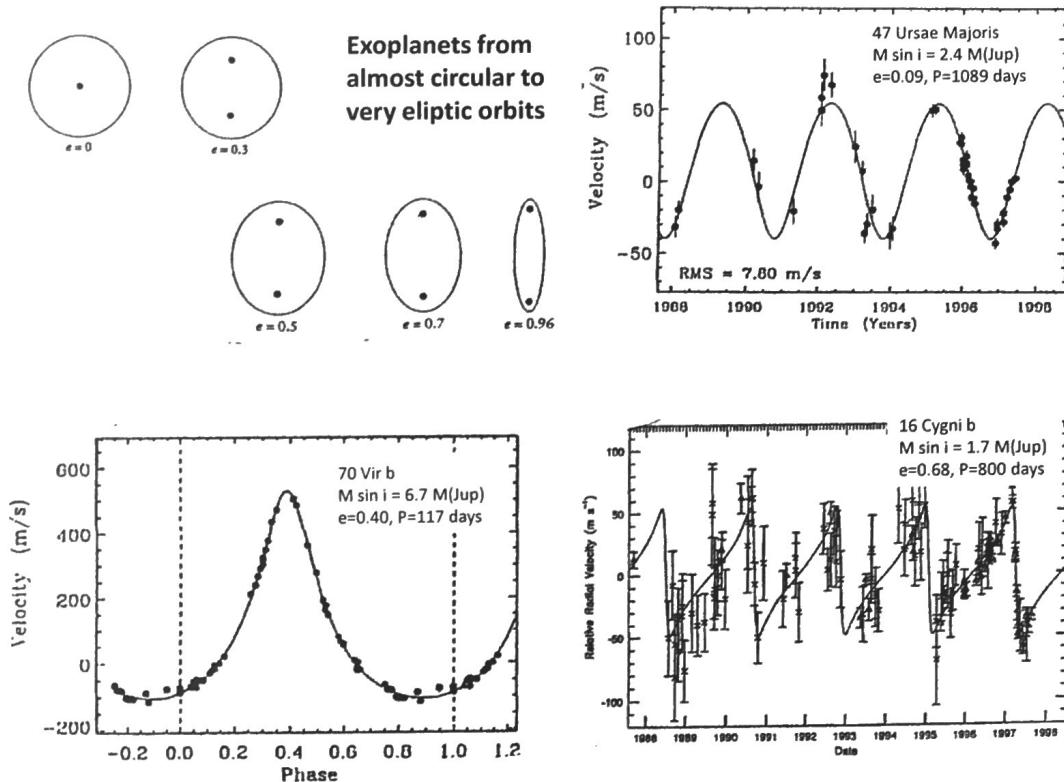


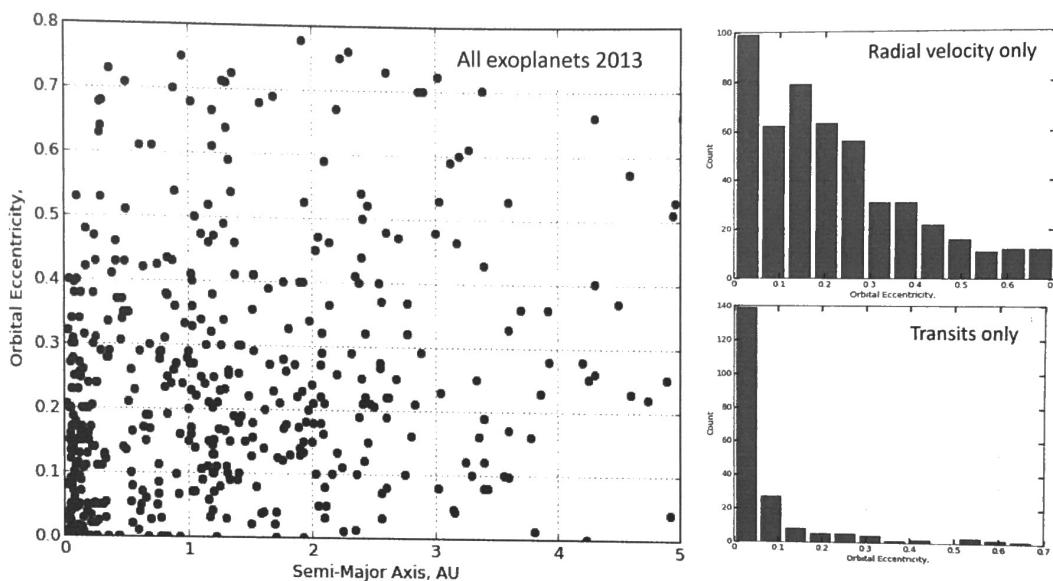
Figure 1.4. Some examples of measured light curves of stars orbited by exoplanets of different eccentricity.

several groups were systematically monitoring a total of about 300 individual stars. Very surprisingly the first confirmed sinusoidal Doppler motions were identified already few years later, but for solar type stars with Jupiter-mass planets in smaller than Mercury-like orbits – soon to be named hot-Jupiter exoplanets. Because of the extremely small orbits and correspondingly small orbital periods (down to a few days), it was possible to measure many orbital periods during few weeks or months of observations, and hence reach a sufficient accuracy to confirm the existence of planets.

During the first almost two decades, the radial velocity technique was by far the most successful method in terms of number of exoplanet detections. By September 2011, identifications of 633 of a total of 671 exoplanets orbiting main sequence stars involved radial velocity observations. Most of them were giant planets inside the frost line of formation. The radial velocity technique is most sensitive to such planets, and it was therefore an open question whether the dominance of giant exoplanets in small orbits was a pure observational effect or was reflecting that such planets are really the most abundant planets in the universe. We saw in a previous chapter that as far as we understand the formation of planetary systems, giant planets can only form outside the frost line. Only beyond the frost line was the proto-planetary nebula cold enough that water could condense out of the nebula as ice. Only when water – the most abundant condensable molecule in the nebula – could contribute to the build-up of the solid planet, could the planet grow large enough that the surrounding hydrogen-helium gas could collapse onto the solid planet and form a giant gas-planet. Should we now believe that this scenario is wrong, because almost all the observed exoplanets were in contradiction with the theory taken face on?

A major problem in deriving statistics from radial velocity observations, is the very inhomogeneous accuracy and sample selection between the science teams and over time. A particular valuable sample already

at an early time was therefore the 1300 solar-type stars monitored in the homogeneous Lick radial velocity survey. 75 of the 1300 monitored stars showed giant exoplanets in small orbits, indicating that 5% of solar-type stars in the solar neighbourhood have planets not predicted by the standard theory for planetary system formation. Other surveys show similar results (see Fig. 1.32). This still tell us nothing about how common our own solar system is, with its 4 giant planets at large orbits and its 4 small terrestrial planets inside the ice-line, but it does tell us that at least 5% of stars form planetary systems with planets very different from those in our own solar system. As we will return to in a little while, the most likely explanation for the giant planets in small orbits, is that they shortly after their formation migrated from their formation place outside the snow-line and in to the small orbits we find them in today. This fact has been a great inspiration to improve the formation theory of our own solar system too, with the inclusion of minor, but important, planetary migration in the outer parts of the solar system.



**Figure 1.5.** The eccentricity distribution in the left panel (eccentricity as function of semi major axis in AU) is puzzling, because it more resembles the distribution of eccentricity in double star systems than the much more circular orbits expected for planet formation. The right upper panel show the eccentricity distribution (number of planets as function of eccentricity for  $0.0 < e < 0.7$ ) for exoplanets discovered with the radial velocity method, while the right lower panel show the same thing for the transit planets only.

In deriving Eq. 1.7, we assumed the orbits to be circular. In general the orbits are of course not circular. In fact it is one of the surprising things that the orbits are very elliptical; more resembling double star orbits than planetary orbits as we know them from our own solar system (see Fig. 1.5). This makes it necessary to include more parameters in the fit to the observed light curve. Fig. 1.2 shows in the lower panel the geometry of the elliptic orbits, and from a little more complicated geometrical considerations than those for the circular movement above, we conclude that the radial velocity of one of the components (here the star) can conveniently be described as

$$RV_* = K_*(e \cdot \cos\omega + \cos[\Theta_*(t) + \omega]) \quad (1.8)$$

where  $e$  is the eccentricity of the orbit.  $\omega$  is seen from the centre of mass the angle between the line of sight toward the observer and toward the periastron (periastron is the point in the orbit where the star is closest to the centre of mass, as opposed to apoastron which is the point most far away from the centre of mass).  $\Theta_*(t)$  is the periastron angular distance, such that  $\Theta_*(t) + \omega$  is the angle between the direction from the centre of mass to the Earth and toward the position of the star at time  $t$  (see Fig. 1.2).  $K_*$  is the observed radial

velocity semi amplitude of the star,

$$\begin{aligned} K_* &= \Delta RV = (RV_{max} - RV_{min})/2 \\ &= \frac{2\pi a_* \sin i}{P\sqrt{1-e^2}} \end{aligned} \quad (1.9)$$

where  $RV_*(t)$  the observed radial velocity curve.

By introducing Kepler's 3rd law (Eq. 1.4 with the semi-major axes  $a_*$  and  $a_p$  instead of the radii  $r_*$  and  $r_p$ ) into Eq. 1.8 with  $K_*$  from Eq. 1.9, we get

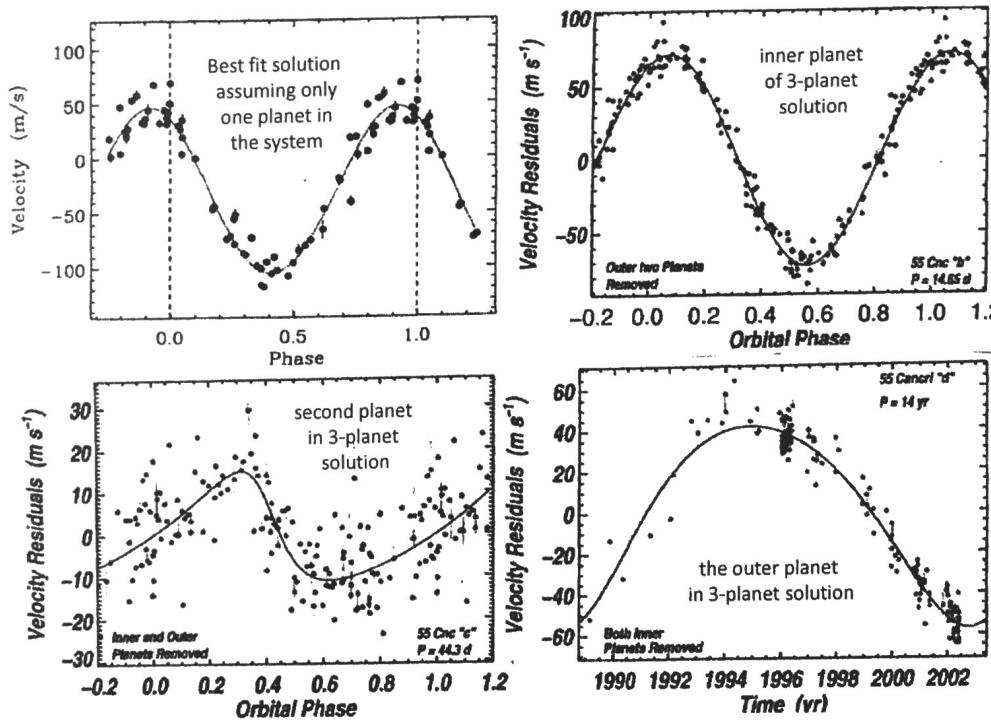
$$M_p \sin i = \tilde{c} K_* \sqrt{1-e^2} P^{1/3} (M_* + M_p)^{2/3} \quad (1.10)$$

If  $K_*$  is measured in km/s,  $P$  in days, then  $\tilde{c} = 4.7 10^{-4}$  will correspond to  $M_p$  in units of  $M_\odot$ .

We see from Eq. 1.8 that if  $e=0$  (circular orbit), then the measured radial velocity curve,  $RV_*(t)$ , will be a sinus curve, and Eq. 1.10 becomes Eq. 1.7.

### 1.2.1 Multi-planet systems

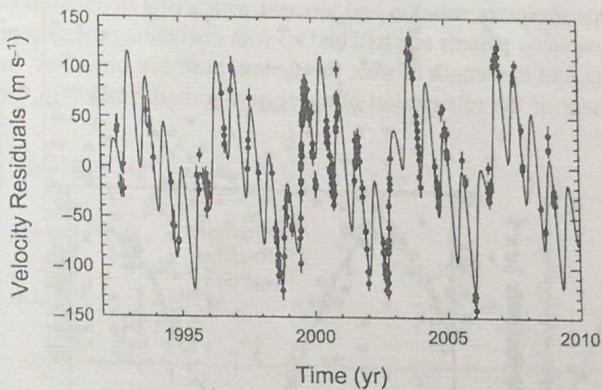
In deriving Eq. 1.1 to 1.10 we have assumed that the star is orbited by one planet only. If more planets orbit the star, one could proceed iteratively by finding a best fit solution to the most obvious planetary period from visual inspection of the data, then subtracting this planetary solution and proceed with a best fit solution to the most obvious period of the residuals, etc. How many planets one will end up with concluding the system consist of, will depend on the quality of the data and the length of time the system has been observed. In figure 1.6 is given an example of the development of the recognition of successively more planets in the system.



**Figure 1.6.** The fit to the radial velocity observations of the star 55 Cnc, based on the assumption of the existence of one exoplanet only (first panel), and based on the existence of 3 planets (panel 2 – 4 for each of the planets).

The exoplanet 55 Cnc b was one of the first exoplanets to be discovered. It has a period of  $P=15$  days and orbits a G8V (i.e., almost solar) type star. However, it was soon realized that there was a systematic

deviation from the almost sinusoidal velocity curve of 55 Cnc, and it was suggested that the deviation was due to a planet with a quite long orbital period which systematically pushed the 55 Cnc b data in and out of the best-fit sinus curve. Later when more and better data became available, it was believed that even a two-planet solution was not an optimal fit to the data, but that at least three planets are necessary in order to model the observations. Fig. 1.6 shows the early data and a one-planet fit to these, as well as a more recent interpretation based on an analysis that assumes the existence of three planets orbiting 55 Cnc; a ( $M_p \sin i = 0.84 M_{Jup}$ ) inner planet with an orbital period of 14.7 days, a mid-planet of  $0.20 M_{Jup}$  and  $P = 44.3$  days, and finally an outer planet of  $4.2 M_{Jup}$  and  $P = 5484$  days = 15 years. Even more recent interpretations of the velocity curve of 55 Cnc, find a slightly better fit to the data by assuming that there are 4 planets instead of 3, with the new one (55 Cnc e) being a Neptune mass planet with an orbital period of only 2.8 days and an orbital radius of 0.04 AU; ten times smaller than the orbit of Mercury. This changes the best fit of the other planets too, such that Cnc d was then estimated to be at an orbital radius of 5.3 AU. This is just outside the snow-line in our solar system, and makes Cnc d one of the few radial-velocity exoplanets which are in agreement with the standard theory of solar system formation without planetary migration. The most recent interpretation with more data is now a 5 planet solution, with the best fit model of the innermost planet now moving in to 0.016 AU (with half the mass of Neptune) and the outermost out to 5.76 AU (with  $3.8 M_{Jup}$ ).



**Upsilon Andromedae (UpsAnd)**  
is an F8V (left) + M4.5V (right)  
double star system with four  
Jupiter-like planets orbiting the  
larger of the stars.

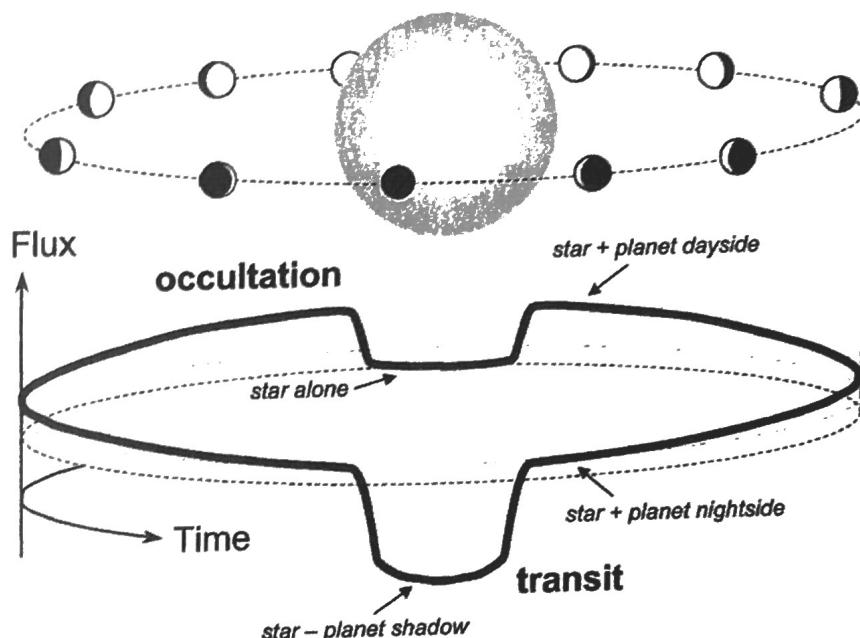


**Figure 1.7.** The double star Upsilon Andromeda with its four jupiterlike exoplanets. Lower panel is an artist impression of the system (not to scale), while the upper panel is the residual radial velocity light curve (dots) and the best multi-Kelplerian orbit fit (solid line), after the signal from the inner most planet has been subtracted.

An even more complex system is Upsilon Andromedae. This is a double star system consisting of a  $1.3 M_\odot$ , 2.5 Gyr old F8 main sequence star, star Ups And A, and a smaller M4.5 V star, Ups And B, orbiting one another in a large orbit of 750 AU. Ups And A is orbited by four Jupiter-sized exoplanets, Ups And b,c,d and e, of mass 0.6, 1.8, 3.8, and  $1.0 M_{Jup}$  at a distance of 0.06, 0.8, 2.5, and 5.2 AU, respectively. For an F8 V star, the orbit of Ups And b must be very close to, or even within, the habitable zone. Ups And e is some of the closest we know of a Jupiter-twin, orbiting almost at Jupiter's orbit in our solar system (at 5.2456 AU compared to Jupiter's 5.2043 AU), and with a mass  $M = 0.96 \pm 0.14 M_{Jup}$  in an almost perfectly circular orbit ( $e=0.00536$ ). Fig. 1.7 shows an artist impression of the system (with the distances not at the same scale as the planetary and stellar sizes), and the best-fit residual radial velocity curve to the observations after the subtraction of the effect of the innermost planet. Today a couple of hundreds multiple exoplanetary systems are known, and although the majority of the known exoplanets are single planet systems, this is probably more an observational effect of only being able to see the largest of the exoplanets in the system than a real physical fact.

### 1.3 The transit method

If the orbital *inclination* (see Fig. 1.2) of an exoplanet is close to  $90^\circ$ , it will pass in front of the star once per orbit, hereby blocking a small part of the star light; we call this phenomenon a planetary transit, and say that the planet transits the star. If the photometry is very accurate, we may also see the tiny dimming of the combined light from the star and the planet when the planet slides in behind the star; we call this phenomenon an occultation (or sometimes the *secondary eclipse*). Both phenomena could in principle equally well have been called eclipses, but we usually reserve this word for eclipsing binaries (two stars passing in front of one another), solar and lunar eclipses, and other phenomena where the mutually occulting objects are of comparable angular size to one another. The period from when the planetary disk first time touches the stellar disk to when it is fully in front of the disk is called the ingress, and the two points are called first contact point (or exterior ingress) and second contact point (or interior egress). Similarly the exit of the planet from the stellar disk is called the 3rd and 4th contact points (or interior and exterior egress) and the corresponding period the transit egress. Fig. 1.8 show the geometry of the exoplanetary transit and occultation and the corresponding schematics of the transit light curve.



**Figure 1.8.** The exoplanetary transit and occultation. The upper panel illustrates the geometry, while the lower panel shows the total flux of the planet and star (fat line) at the various phases, and the stellar flux alone (dotted line).

Since the planet will turn its dark side toward us during the transit, we can to a first approximation think of it as a black, circular disk that passes in front of an evenly luminous stellar disk, and which has a radius given by

$$\Delta L/L_* = (R_p/R_*)^2 \quad (1.11)$$

where  $R_*$  is the stellar radius.

It will be useful for the more detailed discussion below of the wealth of information that lies in the transit lightcurve, to be a bit more precise than in Eq. 1.11 about the dimming of the stellar light that a transiting planet will cause. As almost always, the words used have a slightly different meaning in science and in daily

speaking, which at first can be somewhat confusing, and even within different areas of natural science the words are used differently. In optics one often use the words candela, lux and lumen (now commonly listed on electrical bulbs), and radiant and luminous flux, because the eye's response function historically (and in modern attempts to list the bulbs as how much energy they affect our eyes with) has been mixed into the definition. We therefore recall for the description below that in astronomy the terms luminosity, flux, and intensity are often used a bit interchangeable, but that they do have each their strict meaning. Intensity is the energy (e.g. Joule or erg) that is radiated in a given direction from a square unit (e.g.  $\text{cm}^2$ ) per time unit (e.g. per second), through a cone of opening angle of one steradian. Flux is the total energy radiated per square unit per second (i.e. the intensity integrated over all directions), while luminosity is the total energy output of the star (i.e. the flux times the surface area). What we are really interested in is the energy that reach our telescope per time unit. This is also a flux, but in order to distinguish it from the stellar flux  $F_*$  (i.e. the flux leaving the star) it is often termed flux density  $f_*$  (i.e. energy arriving per square unit of detector per second). We will use the terms  $F_p$  and  $f_p$  correspondingly for the planet. For the purpose of deriving the change in the flux density during a transit (i.e., the transit light curve), we will start with defining  $I$  to be the intensity emitted from the star (or planet) in our direction. This is the most useful term to begin with, because it is the main quantity that varies over the stellar surface and because the planetary intensity differs with the planetary position relative the star (the phases of the planet).

We now have that the received flux density before (or after) the transit is

$$f_{\text{tot}} = f_* + f_p = I_* \frac{\pi R_*^2}{d^2} + I_p \frac{\pi R_p^2}{d^2} \quad (1.12)$$

where  $R_*$  and  $R_p$  are the stellar and planetary radii,  $I_*$  and  $I_p$  are the stellar and planetary intensities emitted in our direction, and  $d$  is the distance between us and the transiting system. In the same way the total flux density received during the transit is

$$f_{\text{trans}} = I_* \frac{\pi R_*^2 - \pi R_p^2}{d^2} + I_p \frac{\pi R_p^2}{d^2} \quad (1.13)$$

The relative change in received flux density (i.e., the light curve) is therefore

$$\begin{aligned} \frac{f_{\text{trans}}}{f_{\text{tot}}} &= \frac{I_*(R_*^2 - R_p^2) + I_p R_p^2}{I_* R_*^2 + I_p R_p^2} \\ &\approx 1 - \frac{R_p^2}{R_*^2} + \frac{I_p R_p^2}{I_* R_*^2} \end{aligned} \quad (1.14)$$

where we have used that the planetary flux density is small compared to the stellar flux density ( $I_* R_*^2 + I_p R_p^2 \approx I_* R_*^2$ ). We see that if  $I_p$  is zero, Eq. 1.14 becomes Eq. 1.11.

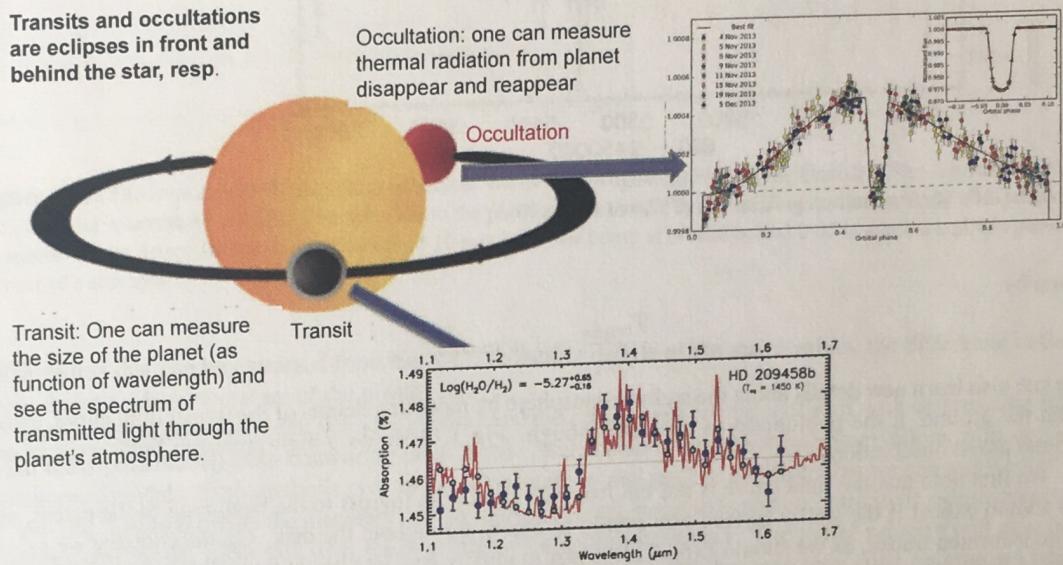
In reality we should of course use the angular radii,  $\theta = R/d$ , instead of the physical radii in Eq. 1.14, but for an extrasolar system the distance to the system will always be so large that it is a good approximation to assume the same distance  $d$  to the star and the planet (so that  $d$  cancels in deriving Eq. 1.14). This is in contrast to the transits inside the solar system itself, as e.g. the well studied historical Mercury and Venus transits of the Sun.

A typical hot-Jupiter extrasolar planet could have an effective (atmospheric) temperature of  $T_{\text{eff}} = 1500 \text{ K}$ , and its intrinsic surface flux is therefore substantial; in particular in the red and infrared. If the planetary atmosphere is dense and the planet rotates fast, the day- and night-sides may have the same temperature, while a thinner and slower rotating atmosphere (or distinct surface features if the planet is solid) could give rise to different day- and night temperatures. If both the transit and the occultation are observable (which is usually only possible from space, but for some cases can be obtained from accurate ground based observations as well), one can determine the ratio between the intensity of the light from the night-side of the planet,  $I_p(t_{\text{tr}})$ , and its day-side,  $I_p(t_{\text{occ}})$ , and hence gain some knowledge about the heat transport efficiency of the planet (see Fig. 1.9). Introducing into Eq. 1.14 that we really see the planetary night side during the transit and the

## 1. EXOPLANETS

planetary day side right before and after the occultation, gives us that

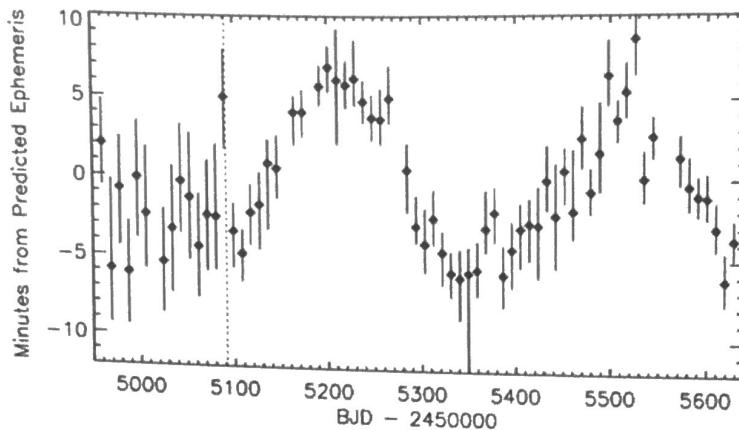
$$\begin{aligned}\frac{f_{tot} - f_{trans}}{f_{tot}} &= \frac{R_p^2}{R_*^2} - \frac{I_p(t_{tr}) R_p^2}{I_* R_*^2} \text{ and} \\ \frac{f_{tot} - f_{occ}}{f_{tot}} &= \frac{R_p^2}{R_*^2} \frac{I_p(t_{occ})}{I_*}\end{aligned}\quad (1.15)$$



**Figure 1.9.** From transits and occultations one can obtain information about the atmosphere, such as composition and the difference in day- and night-time temperature of the atmosphere.

With high precision observations of successive transits, the orbital period can be determined with an accuracy of approximately a second. With such high timing precision, one would be able to discuss tiny delays and variations in the transit caused by for example orbiting moons pulling the planet a bit back and forth, and gravitational effects from other (possible undetected) planets in the system, especially ones in resonances. The relative low cadence of the Kepler observations give the Kepler-timing a lower accuracy than can be obtained from the ground, but the completeness and huge number of successive transits observed by Kepler, makes the (now publicly available) data set very valuable for identifying TTVs. Figure 1.10 shows the transit timing variations of Kepler 9b as discovered in 2011. Kepler 9 is almost a solar twin, but only 1 Gyr old. A clear transit signal showed a  $45 M_\oplus$  in a 19 days orbit. However, the transit timing soon revealed periodic variations with a very large amplitude of  $\approx \pm 6$  minutes as shown in Fig. 1.10 from the discovery paper. At first it was believed that an additional  $6 M_{Jup}$  planet, Kepler 9c, further out in the system (with an orbital period of 160 days) was causing the perturbations, but more detailed observations and modeling now points at a much smaller ( $31 M_\oplus$ ) planet moving in mean motion resonance with Kepler 9b (i.e., orbiting the star once each time Kepler 9b orbits twice). Additional modeling have speculated about the existence of a third planet in the system, Kepler 9d with a period and mass of 1.59 days and  $5.25 M_\oplus$ , respectively. Kepler 9c was the first out of now a handful of exoplanets having been discovered by the TTV technique. Dynamic modeling of TTV planets allow determination of the planetary masses from the transit data alone, without additional observations from for example radial velocity measurements.

If also the occultations can be determined, then details of the orbital form, such as eccentricity and the detailed orientation of the orbital ellipse in space, can be determined. For example, the duration of the transit,  $T_{trans}$ , relative the length of the occultation,  $T_{occ}$ , is a function of the orbital eccentricity,  $e$ , and the periastron argument,  $\omega$  (i.e. the direction in space of the periastron, the point of the planetary orbit closest to the star),



**Figure 1.10.** The transit timing variations (TTV) of Kepler9b.

given by

$$\frac{T_{\text{trans}}}{T_{\text{occ}}} = 1 + es \sin \omega \quad (1.16)$$

We can also learn new details about the stellar atmosphere by modeling details of the transit light curve, even from the ground, if the photometry is accurate enough. Fig. 1.11 shows a high-precision light curve from ground based observations.

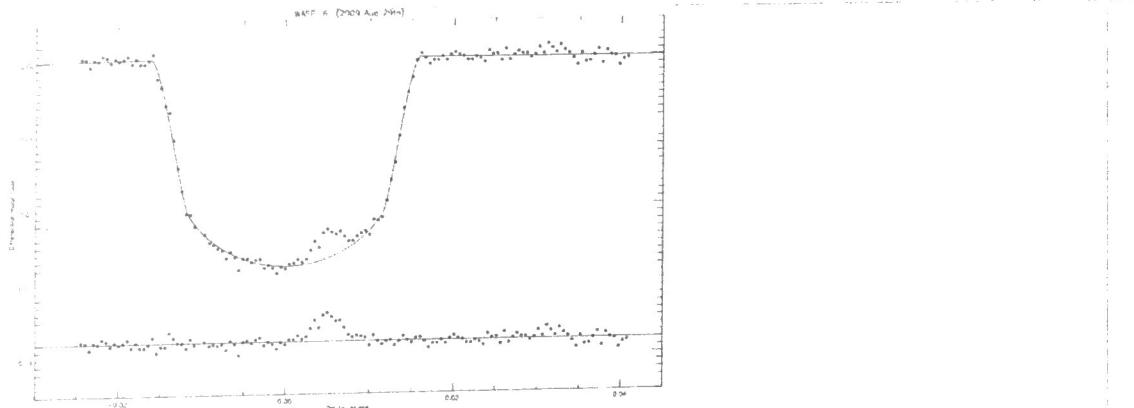
We first note that the light curve is not flat from the end of the ingress to the beginning of the egress, as we should expect if the stellar surface was evenly luminous throughout the disk. On the contrary we see a more u-formed transit, as we should expect if the star gradually gets less bright toward the edge of the disk. This is the stellar limb darkening effect, due to the fact that we see deeper into the photosphere to hotter layers at the middle of the star than at the edge of the disk. One easily sees the limb darkening on the Sun with the naked eye at these occasions when the cloud cover is exactly such that one can look directly into the Sun on the sky. Then the solar limb look less bright than the centre, and a bit illusionary makes the Sun look like a 3-dimensional sphere in the sky. Limb darkening can be computed accurately from stellar atmosphere models, but are more commonly fitted to the light curve by assuming that the intensity of the light,  $I(\mu)$ , is a quadratic function of the relative distance from the centre ( $\mu = 0$ ) of the stellar disk to the rim ( $\mu = 1$ ) of the disk,

$$I(\mu) = 1 - u_1(1 - \mu) - u_2(1 - \mu)^2 \quad (1.17)$$

where  $u_1$  and  $u_2$  are the quadratic limb darkening coefficients to be determined from the fit to the light curve. Substituting Eq. 1.17 into Eq. 1.14 or 1.15 and performing corresponding numerical fits to observed data will allow us to determine the limb darkening coefficients  $u_1$  and  $u_2$ , which can be checked against a corresponding stellar atmosphere, and eventually correct the calculated temperature structure to fit the data, and thus improve our understanding of the stellar atmosphere. We see from Eq. 1.15 that when the planet transits the meridian of the star,  $I(\mu)$  is at maximum, and therefore the negative term in the light curve at minimum, i.e., the light curve is at its deepest, and to both sides of that  $I_\star$  is decreasing quadratic as function of  $\mu$ , giving rise to the u-shaped form seen in Fig. 1.11.

We also notice a clear increase of the light shortly after the middle of the transit. This is caused by the planet moving in front of a spot on the surface of the star. Since star spots are less bright than the surrounding stellar disk (just like sun spots), the planet is blocking less bright light while it moves over the star spot, and hence the total luminosity increases and give rise to the bump seen in the light curve in Fig. 1.11. In principle one can model the form, size and duration of the star spots from such observations, and one can also determine the stellar rotation period if the same spot can be identified on successive transits.

Determination of  $R_p$  therefore requires a more careful analysis of the light curve than Eq. 1.11 indicates, but it also allow us to extract more information from the transit light curve than just the relative planetary to



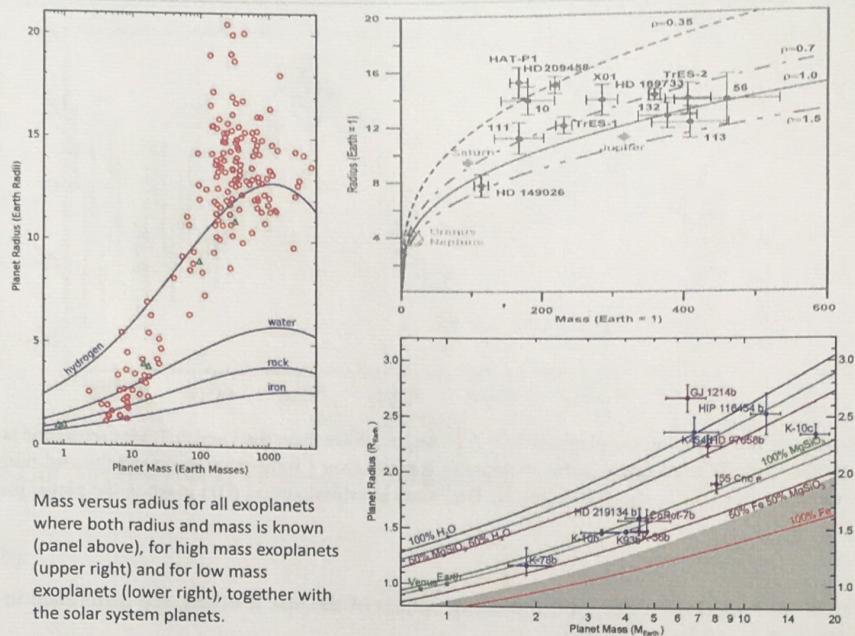
**Figure 1.11.** Observations of the transiting exoplanet WASP-6 in August 2009 from the Danish 1.54m telescope at the ESO-La-Silla observatory in Chile. The accuracy in the photometry is better than 1 milli-magnitude and the mid-point of the transit can be determined with an accuracy of 10 seconds. The bump at orbital period 0.05 is when the planet passes in front of a star spot.

stellar radius that can be extracted from Eq. 1.11, including details of the star it orbits, the difference in the day and night side temperature of the planet, etc.

We see that in both the radial velocity method and the transit method, we will start with determining the stellar mass (and radius) from comparing photometric (and possible spectroscopic) observations with stellar evolutionary models. Both methods give us the (planetary) orbital period, and from this we can use Kepler's 3<sup>rd</sup> law (Eq. 1.4) to estimate the planetary orbital radius (or semi major axis and eccentricity if radial velocity measurements exist). After this step, the two methods are complementary. The radial velocity will give us  $M_p \sin i$ , but the transit will give us  $R_p$  and automatically also  $i \approx 90^\circ$  because only edge on systems will transit. If an exoplanet can be identified with both the radial velocity method and the transit method, we will therefore know both  $M_p$  and  $R_p$  and hence also the mean density  $\rho_{\text{av}} = 3M_p/4\pi R_p^3$  of the planet.

The first case where it was possibly to determine both the radial velocity variations and transits, was for the star HD 209458. In this very important case, radial velocity measurements showed  $M_p \sin i = 0.69 M_{\text{Jupiter}}$ , and the transit measurements showed (beside  $\sin i \approx 1$ )  $R_p = 1.32 R_{\text{Jupiter}}$ , which means that  $\rho_{\text{av}} = 0.4 \text{ g/cm}^3$  (compared to 1.33 for Jupiter and 0.69 for Saturn). Fig. 1.12 shows the derived mean density for the exoplanets where both radial velocity and transit measurements exist.

The fact that the same exoplanets have been seen both in radial velocity and transit measurements, first of all eliminated completely the doubt that had been among part of the scientific community in the beginning about whether the periodic stellar radial velocity variations should be understood as due to orbiting exoplanets or due to an intrinsic stellar phenomena. For some time after the first exoplanet announcements in 1995, it was discussed whether the variations could be due to stellar pulsations, rotations, etc, instead of being a planetary effect. The double confirmations also solved a longstanding question about what is really unexpected in the observed hot-Jupiter exoplanets. Our standard theory for the solar system formation (as outlined in a previous chapter) predicts that gas planets will form only beyond the frost line ( $\approx 3 \text{ AU}$  for solar type stars), while planets inside this limit only can be composed of rare elements (elements with high condensation temperature, such as metals and "rocks") and therefore have to be small (unless the proto-stellar nebula was very massive, which would contradict several observations and theoretical arguments). Although most explanations of the hot-Jupiter exoplanets assume that they are formed beyond the frost line and later migrated into their present small orbits, we could *a priori* not exclude that they were gigantic Earth-like planets of rock and metals, formed where they are now (i.e. without migration), by a mechanism different from the one that took place in our own solar system, although it would be a tough challenge for planetary system formation models to explain this. The fact that the densities of the close in giant exoplanets where both transit and radial velocity measurements exist, all have densities close to those of Jupiter ( $\rho = 1.33 \text{ g/cm}^3$ ) and Saturn ( $\rho = 0.69 \text{ g/cm}^3$ ) clearly show that these exoplanets are true gas planets. In other words, it is the fact that the



**Figure 1.12.** The planetary radius versus mass for a number of exoplanets where both radius and mass is known (panel above), for high mass exoplanets (upper right) and for low mass exoplanets (lower right), together with the position of the planets in our own solar system.

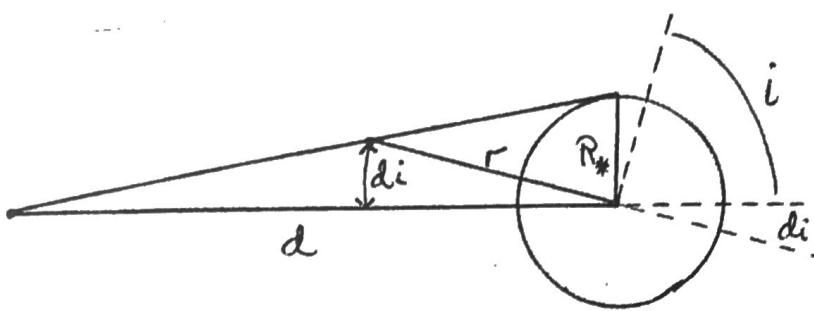
“hot-Jupiter” exoplanets are in small orbits (i.e., are “hot”) that is puzzling, not that they are big (i.e., are “Jupiters”). Today many exoplanets have been studied with both transit and radial velocity observations, and one can begin drawing conclusions about the distribution of planets of various densities, and start thinking about their composition. Fig. 1.12 show the mass versus radius for exoplanets that have been measured with both methods, together with the relation between radius and mass for homogeneous planets of various mean densities, and the position of the giant planets in our own solar system in the diagram.

In order for a planet to transit a star, the Earth needs to be very close to the orbital (“ecliptic”) plane of the exoplanetary system. The further away the planet is from the star, the closer we need to be to the orbital plane in order for the exoplanet to transit the star. There are all reasons to believe that the orbital planes of extrasolar systems are distributed randomly relative to the direction toward the Earth (the inclination of the solar system itself is  $60^\circ$  toward the Galactic plane, which can be seen with the naked eye on a clear evening or morning when one can see the zodiacal light stretching up from the horizon in a  $60^\circ$  angle with the Milky Way plane). Therefore the probability of the plane to be within an interval  $di$  of any inclination angles is independent of inclination, and in particular the probability  $P_{trans}$  of the system to be within an inclination  $di$  from edge on ( $i=90^\circ$ ) is

$$P_{trans} = \frac{2di}{\pi} = \frac{2}{\pi} \left( \frac{R_*}{r} - \frac{R_*}{d} \right) \approx \frac{2}{\pi} \frac{R_*}{r} \quad (1.18)$$

where we have used that (1) there are 4 orientations relative edge on, which will make the plane be seen as within the angle  $di$  radians from edge on, (2) that  $di$  is small, and (3) that  $d \gg r$ . See also Fig. 1.13.

In order to estimate  $P_{trans}$  from Eq. 1.18, we must first determine  $r$ , which can be done from transits or radial velocity measurements and using Kepler’s laws and photometric and/or spectroscopic estimates of the stellar mass. Since Kepler’s law relates  $r$  to orbital period  $P$  and stellar mass  $M_*$ , we can also express



**Figure 1.13.** The geometry of an exoplanet transiting the star it orbits.

$P_{\text{trans}}$  in these variables, and get from Eq. 1.18

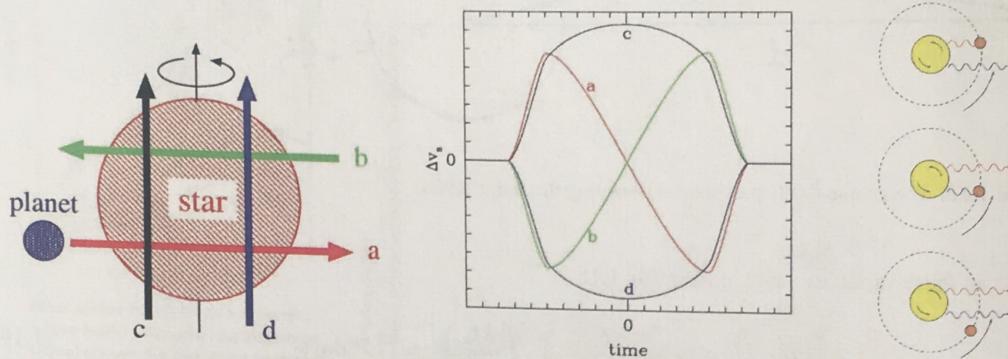
$$P_{\text{trans}} = 0.128 \cdot \left( \frac{R_*}{R_\odot} \right) \cdot \left( \frac{M_*}{M_\odot} \right)^{-1/3} \cdot P[\text{days}]^{-2/3} \quad (1.19)$$

Exoplanets are today known with periods as short as less than a day to longer than 10 years. We see from Eq. 1.19 that such planets will have probabilities from more than 10% to well below a per mille to transiting their host stars. Since several hundreds of exoplanets have now been discovered with the radial velocity method, we see that we will expect several of them to also show transit signals, and that the chances are biggest for the stars with small periods. HD 209458b is among the very short period exoplanets with a period of 3.5 days, and it is therefore not unexpected that exactly this planet was the first radial velocity discovered planet to be also identified in transit observations. Today, however, a few thousand exoplanets and exoplanet candidates are known from discovery through transit surveys, foremost the Kepler satellite experiment. Many of these have now been searched for radial velocity signals, and the number of exoplanets where both planetary radii (from transits) and planetary mass (from radial velocity) are known (and therefore average density can be estimated and guesses of compositions made) is therefore increasing rapidly (as seen in Fig. 1.12). However, far from all Kepler-transit planets can be checked for radial velocity signals, because the chosen Kepler candidates generally are too weak for high resolution spectroscopy to be performed on them. This choice was made in order to have sufficiently many stars in Kepler's fixed position field on the sky. Following the great success of Kepler, it was decided to follow up with a new satellite, called TESS, which will be observing only brighter stars, and all over the sky. These stars can all be followed up with high precision radial velocity measurements from the ground, and since they are bright, many will also be nearby, and include the discoveries of the planets where our highest priority should be focused on sending signals for initiating the first communication with extraterrestrial civilizations — if this is what we should do.

### 1.3.1 The Rossiter-McLaughlin effect

For the majority of stars, one can only obtain an integrated spectrum of the stellar disk, i.e. a combined spectrum from all the parts of the visible surface. Due to the stellar rotation some parts of the surface will move toward us and other parts away from us (relative to the centre of the disk). The observed integrated spectrum will therefore be the sum of many red-shifted and blue-shifted spectra, which will cause the spectral lines to be (rotationally) broadened compared to what they would be if the star was non-rotating. This fact can be used to determine the inclination angle and rotation direction of a transiting planet relative to the stellar rotational axis. If the planet orbits the star in the same direction as the star rotates (as all the planets in our solar system), then the planet will first transit parts of the star which are blueshifted (rotating toward us), and hence diminishing the contribution of the blueshifted part of the stellar light, which will make the integrated spectral lines move red-ward. Similar when the transit approach egress, the planet will diminish the contribution of the redshifted stellar light and make the integrated spectral lines shift toward blue. This

movement of the centre of the spectral lines is called the Rossiter-McLaughlin effect (although actually first described by J.R. Holt already in 1893 in connection with spectra of eclipsing variable stars). By measuring the details of how the spectral lines develop during the transit, one can model the path the planet has taken over the stellar disk. The schematics of the effect is illustrated in Fig.1.14.



**Figure 1.14.** Schematic illustration of the shift  $\Delta\nu_s$  in central wavelength (middle pannel) of the stellar spectral lines due to a planet transiting the star in different crossing directions (a, b, c, d in the left panel) relative to the stellar rotational axis. The right-most panel illustrates the situation a, where the planet rotates the same direction as the star, such that first the blue part of the spectral lines are reduced and later the red part.

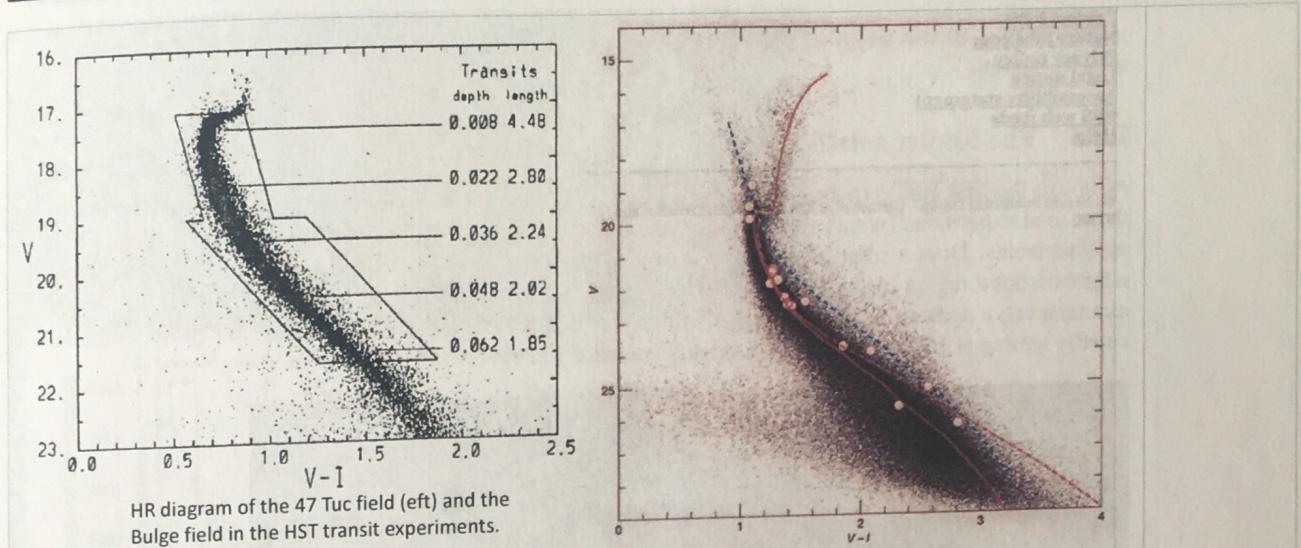
A very surprising result from systematic measurements of the Rossiter-McLaughlin effect of transiting planets, is that seemingly up to 25% of all hot-jupiter planets have retrograde orbital motion (orbits the star in opposite direction of the stellar rotation). This may mean that the effect that migrated the hot-jupiter planets into their small orbits can have been much more violent and dynamical than a simple friction-driven migration, or that some stars flip over due to interaction between the proto-star and the disk, or that many hot-jupiters are captured free floating planets born in orbits around other stars than where they ended up.

### 1.3.2 The HST campaigns

Both the radial velocity method and the transit method have the disadvantage that they are biased toward detecting what we (still) tend to consider the "unusual planetary systems" – giant planets in small orbits. However, the transit method has the advantage over the radial velocity method that it can be applied to more distant stars. The radial velocity method requires high resolution (typically  $R \approx 60,000$ ) spectra of the star, and is therefore limited to relatively bright (and therefore nearby) stars, typically brighter than  $m_v \approx 8$ , or within a couple of hundred light years. The transit method only requires stars bright enough that accurate photometry can be obtained, which has been achieved for stars as far away as more than 10,000 light years.

The highest photometric accuracy is obtainable from space where the observational challenges with subtracting the sky background sufficiently accurate are effectively eliminated. Two campaigns on the Hubble Space Telescope (HST), followed by the dedicated Kepler satellite, have taken advantage of this. The first HST project monitored a field of the globular cluster 47 Tuc continuously during 8.3 days (see Fig. 1.15, left panel), and another project later monitored a dense field toward the Galactic bulge for a similar amount of time (see Fig. 1.15, right panel). The fact that so distant fields can be monitored, makes it possible to expose many stars onto a single frame.

The observed field of 47 Tuc contain about 34,000 stars between  $m_v=17.1$  and 21.6 that could be monitored in two filters every  $\approx 6$  minutes. For stars of  $m_v=18.4$  (corresponding to  $M_*=0.81M_\odot$  and  $R_*=0.92R_\odot$ ) a photometric accuracy of 0.0053 mag was obtained, which is seen from Eq. 1.11 to be sufficient to detect transits by planets slightly smaller than Jupiter. We further see from Eq. 1.19 that for such stars, the probabilities of orbital orientations giving rise to transits of planets with orbital period 2.5 days and 5.0 days are 6.9% and 4.3%, respectively. From the radial velocity studies of the stars in the solar neighbourhood, we know



**Figure 1.15.** Left panel: The HST colour-magnitude diagram of the globular cluster 47 Tuc. The box shows the position of the 34,091 stars searched for exoplanetary transits. For 5 selected magnitudes are listed the expected  $\Delta mag$  (“depth”) for a  $R_p = 1.3R_{\text{Jupiter}}$  exoplanet transit, and the calculated corresponding duration (“length”) of the transit. Right panel: The HST colour-magnitude diagram of a field toward the Galactic Bulge. In total the values for 250,000 stars are plotted in the figure. The 16 stars that showed short-periodic dips in their intensity (and therefore potentially are orbited by giant gas planets in very small orbits) are marked by open circles.

that approximately 1% of the solar type stars have Jupiter like planets with  $P < 5$  days. Applying this to the 34,000 stars in 47 Tuc, makes us expect (if all stars were  $m_v=18.4$  and the same rate as in the solar neighbourhood applied)  $N \approx 34,000 \cdot 0.01 \cdot 0.043 = 15$  of the stars to show a visible transit during the campaign. However, the result of the observations was that none of the 34,000 stars were transited by a short-periodic Jupiter-like planet during the 8.3 days of observation, and we must therefore conclude that the dense (more than 1000 stars per pc<sup>3</sup> compared to less than 1 in the solar neighbourhood), low metallicity ([Fe/H]=−0.7) environment of 47 Tuc has resulted in less than 1/10 as many hot-Jupiter planetary systems as in the solar neighbourhood.

Seen in retrospect with today’s knowledge, the result was of little surprise, but at that time the number of known exoplanets was far too small to see (and understand) their dependence on the metallicity of their host stars (see the discussion in connection with Fig. 1.30 later on in this chapter). A new campaign was therefore scheduled (with equally much observing time allocated on top of the 47 Tuc observations, which was already the most time consuming project which had ever been allocated on the Space Telescope), which should monitor 250,000 stars in the direction of the Galactic center. These would represent a statistical distribution of stars in fields of several different densities and metallicities, and one should therefore expect results more similar to the results from the solar neighbourhood than those from 47 Tuc. Again 8 days of continuing monitoring was allocated, and this time 16 stars with short-periodic luminosity dips were identified in the data. Most of these were too weak for radial velocity investigations to be performed, but 4 of the stars were within reach of high-resolution spectroscopy, and they were all confirmed to be true exoplanets. However, the same calculation as we did above show us that there should have been more than 100 transiting Jupiter-sized exoplanets visible in this experiment. Even if we assume all 16 candidates to be real exoplanets, there are still at least a factor 6 too few if compared to the rate of such exoplanets around the stars in the solar neighbourhood. On the other hand, the result is statistically consistent with the result from 47 Tuc, because  $36/250 \times 16 = 2$ , would make us expect to have found only 2 (or less) transiting exoplanets in the 47 Tuc sample. Figure 1.15 (right panel) shows the distribution of all the monitored stars in the second HST experiment in a colour-magnitude diagram with the 16 candidates marked. Note that the stars with exoplanets are distributed relatively even throughout the stellar main sequence, and not markedly concentrated toward

solar-type stars, as was speculated earlier on to be normal, judged from the early solar neighbourhood radial velocity surveys.

### 1.3.3 The Kepler satellite

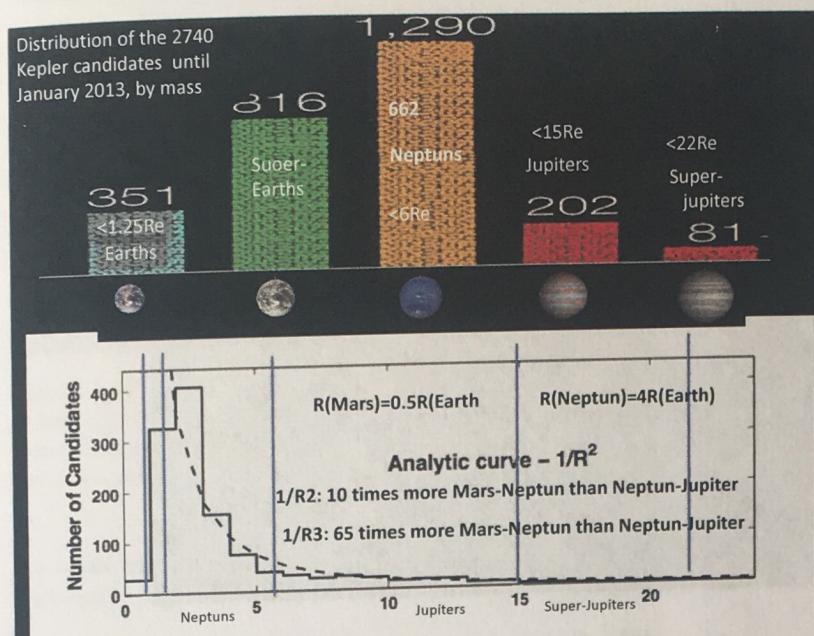
Although the first HST experiment can be well explained as a metallicity effect, it is still mysterious why the second experiment (which should not be metallicity biased) gave the same low abundance as the 47 Tuc measurements. Does it mean that the abundance of exoplanets is particularly high around stars in the solar neighbourhood region of the Galaxy? Following the successful and somewhat strange results of the HST experiments, a dedicated satellite, Kepler, which was basically to do the same as the HST project, but dedicatedly aiming at 160,000 stars was built and launched in 2009.



**Figure 1.16.** The 2740 exoplanet candidates announced by the Kepler team until January 2013. Blue, yellow, and red dots refer to year of announcement, with the newest having a tendency of being of lower planetary mass, based on longer and more accurate analysis of the light curves. The radius of Earth, Neptune, and Jupiter are marked as horizontal lines for comparison. The number of large to small planets are roughly independent of distance out to  $P \approx 1$  month, after which the detections are more biased due to the relatively large orbital periods compared to the observational period. Inset in upper left corner shows all the planets discovered by ground based transit surveys prior to the Kepler mission.

The Kepler satellite is a 1 m space telescope dedicated to monitoring of a specific  $100 \text{ deg}^2$  field in the direction of the constellation Cygnus, where it follows 160,000 stars to look for transiting exoplanets, much the same way as was done in the two HST experiments described above, but continuously for the 4 years satellite lifetime (originally scheduled to 3.5 years; then prolonged to 6 years after the first exciting results; but then technically ran into problems in May 2013 after 4 years of operation). Being in space, it is possible to obtain very high photometric accuracy, and the aim was to obtain 20 ppm accuracy on 6.5 hour integrations on 12 mag solar-type stars, which would make it possible to identify truly Earth-sized planets around sun-like stars. The Earth itself would give rise to a 80 ppm dip, during 13 hours, in the light from the Sun during transits seen from deep space, and Earth analogues would therefore be detectable from the Kepler data.

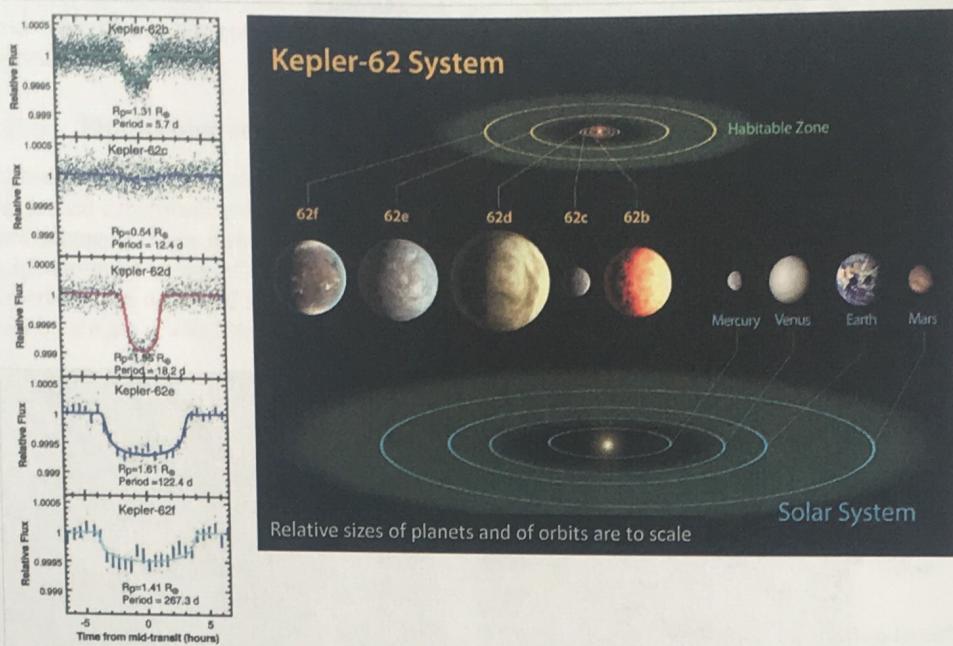
In February 2011 the first set of data was released, representing the first 4 months of observations, from the period May to September 2009 (blue dots in Fig. 1.16). These data included more than 1200 candidate transiting exoplanets, of which 68 (or 6%) have radii less than  $1.25 R_{\oplus}$  (corresponding to  $M \approx 2M_{\oplus}$ ). The



**Figure 1.17.** The ratio of small to large exoplanets in the Kepler data indicates a size distribution of somewhere between  $N(R) \propto 1/R^2$  and  $N(R) \propto 1/R^3$ .

full release in 2013 included 2740 candidates, with many more small planets relative to the large, due to the larger sampling time and hence better statistical possibility to extract the weak signal from small planets out of the noise level. It is not obvious how to correct for the decreased sensitivity to small planets relative to the sensitivity to larger planets, in order to get the true size distribution function, but the data in Fig 1.17 obviously show a rapidly increasing number of planets with decreasing radius; indicating a true functional behavior of maybe  $N(R) \propto R^{-2}$  or  $N(R) \propto R^{-3}$ . We will discuss the best present estimates from the literature below and compare them with the results from other techniques, but just here stress that the former case ( $N(R) \propto R^{-2}$ ) implies that there are 10 times more Mars-to-Neptune-sized planets in the Kepler data than Neptune-to-Jupiter-sized planets, and if  $N(R) \propto R^{-3}$  it will mean that there are 65 times more Mars-to-Neptune-sized planets than Neptune-to-Jupiter-sized planets. Some of the most cited theoretical models of planetary formation predicts a ratio of approximately 1:50, in good agreement with the Kepler data. One of the most exciting conclusion from this statistics, is that the mass-function in the Kepler estimates is independent of orbital radius. Although the statistically significant data only represent orbital periods up to a bit more than a month (see Fig. 1.16), it gives confidence in the theoretical results which predicts that ratio of  $\sim 1:50$  is independent of orbital radius all the way from very small radii to orbits as large as almost Jupiter's orbit in our solar system. If this result holds further observational confrontations, it means that the real number of exoplanets is an overwhelmingly large factor more than what the radial velocity giant planets have revealed, and that small Earth-sized exoplanets are likely to be the dominant type of planets in the Universe.

Due to the very high precision in the Kepler photometry, Kepler has revealed some of the smallest exoplanets discovered to date, including all the confirmed sub-Earth mass exoplanets, a handful of which is within or close to the habitable zone, as will be discussed below. Due to the high sensitivity to small planets in conjunction with the high abundance of such planets, Kepler has also revealed a number of exciting multi-planet systems with several small planets, including the Kepler 62 system, where 5 small planets of radius between 0.5 and 2 times the Earth's radius orbit a 7 Gyr old K2V star ( $T_{\text{eff}} = 4870 \text{ K}$ ,  $M = 0.69 M_{\odot}$ ). The outermost planets, Kepler-62e and -62f, are super-Earth-size ( $1.25 R_{\oplus} < \text{planet radius} < 2.0 R_{\oplus}$ ) planets in the habitable zone of their host star, respectively receiving  $1.2 \pm 0.2$  times and  $0.41 \pm 0.05$  times the solar flux at Earth's orbit. The existence of Kepler 62c is on the margin of what is possible to establish based on Kepler's data.



**Figure 1.18.** The five planets Kepler-62b, c, d, e, and f of size 1.31, 0.54, 1.95, 1.61 and 1.41 Earth radii ( $R_{\oplus}$ ), orbiting a K2V star at periods of 5.7, 12.4, 18.2, 122.4, and 267.3 days, respectively. Left panel shows the actual Kepler data and light curve fit, while the right panel is an artist interpretation with the relative size of the Kepler 62 planets and the solar system planets drawn to scale and the relative size of the orbits drawn to scale too.

## 1.4 The microlensing technique.

Both Newton's and Einstein's theories of gravity predicted that the stars close to the sun apparently would have changed place if observed during a solar eclipse, due to the bending of the light by the gravitational field of the sun. The amount of predicted bending is a factor two different in the two theories, and it was one of the first triumphs of Einstein's theory of relativity, that the observed amount of deflection (i.e. bending) of the star light seen during the total solar eclipse in 1919 was the value predicted by Einstein's theory, not Newton's. In 1936 Einstein wrote a paper where he calculated what the bending of the light will look like if the foreground object is not the Sun, but a point sized and far away star that deflects the light from background stars even further away. It became the first, and now highly cited, paper on microlensing. In his paper he describes the effect of a stellar gravitational field (the lens) magnifying the light of a background star by focusing its light toward us, in analogue with how a magnifying glass does. His reasoning follows the basic geometry sketched in Fig. 1.20.

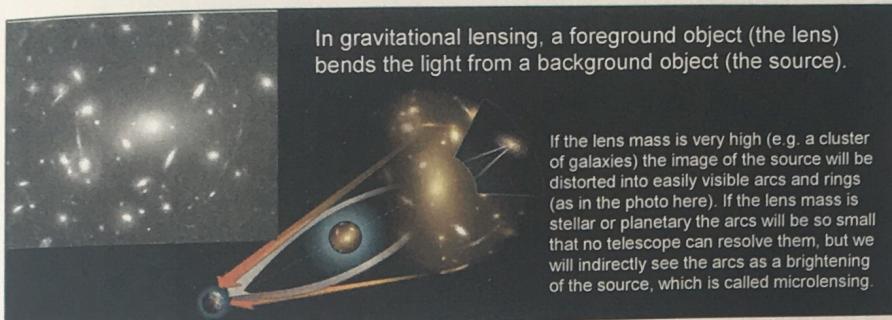
### 1.4.1 The stellar microlensing.

The central equation in Einstein's work is the so-called lens equation. In its most trivial form, from the geometry of Fig. 1.20, it just simply says that the angle  $\beta$  between the line of sight to the lens and the true position of the source is the angle  $\Theta$  between the lens and the apparent place of the source minus the angle  $\alpha$  between the apparent and true places of the source.

$$\beta = \Theta - \alpha \quad (1.20)$$

The theory of general relativity predicts that the bending angle  $\alpha_b$  of a photon passing a distance  $r$  from an object of mass  $M_L$  (the lens) is given by

$$\alpha_b = 4GM_L/c^2r \quad (1.21)$$



**Figure 1.19.** An illustration of the gravitational lensing from a whole cluster of galaxies. The mass of the cluster is so large that the light from the background objects are bent and stretched into easily visible arcs and rings. For stellar masses the same bending and stretching takes place, but the arcs are too small to be visible individually, and are only seen indirectly as a brightening of the background object.

For small angles (using  $\tan(\alpha) \approx \alpha$  and  $\tan(\alpha_b) \approx \alpha_b$ ), it is seen from Fig. 1.20 that  $\alpha$  (in Eq. 1.20) is related to the bending angle  $\alpha_b$  (in Eq. 1.21) through  $\alpha D_S = \alpha_b (D_S - D_L)$ , such that  $\alpha$  can be expressed as

$$\alpha = \alpha_b D_{LS} / D_S \quad (1.22)$$

where we have used the abbreviated form of writing  $D_{LS} = D_S - D_L$ . Substituting Eq. 1.21 into Eq. 1.22 and using that  $r = D_L \tan(\Theta) \approx D_L \Theta$  gives

$$\alpha = \frac{4GM_L}{c^2 r} \frac{D_{LS}}{D_S} = \frac{4GM_L}{\Theta c^2} \frac{D_{LS}}{D_L D_S} \quad (1.23)$$

which substituted into the simple lens equation (Eq. 1.20) gives the more interesting form of the lens equation

$$\beta = \Theta - \frac{4GM_L}{c^2} \frac{D_{LS}}{D_L D_S} \frac{1}{\Theta} \quad (1.24)$$

Multiplying with  $\Theta$  on both sides of Eq. 1.24 brings the equation into the form of a standard quadratic equation in  $\Theta$ , which has the two solutions

$$\Theta_{\pm} = \frac{\beta}{2} \pm \sqrt{\left(\frac{\beta}{2}\right)^2 + \frac{4GM_L}{c^2} \frac{D_{LS}}{D_L D_S}} \quad (1.25)$$

We now introduce the standard scaling (central in all microlensing calculations), of the angular radius of the Einstein ring, the *Einstein radius* or  $\Theta_E$ . It can be expressed as

$$\Theta_E = \sqrt{2R_{Sch} \frac{D_{LS}}{D_L D_S}} = \sqrt{\frac{4GM_L}{c^2} \frac{D_{LS}}{D_L D_S}} = \tilde{k} \sqrt{M_L} \sqrt{\frac{D_{LS}}{D_L D_S}} \quad (1.26)$$

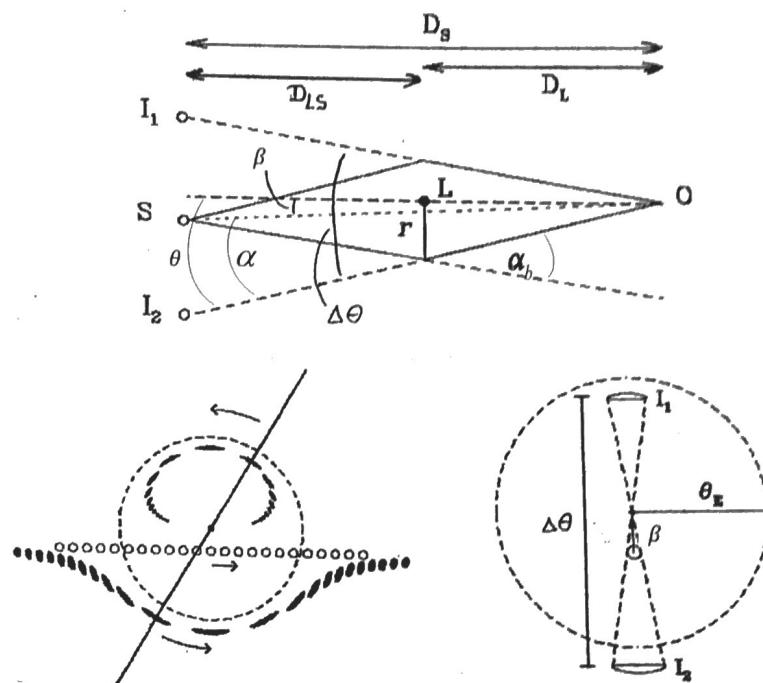
expressed in terms of the mass  $M_L$  of the lens, the distance  $D_{LS}$  between the lens and the source, the distance  $D_L$  between the lens and the observer, and the distance  $D_S$  between the source and the observer.  $R_{Sch}$  is the Schwarzschild radius of the lens. If  $M_L$  is expressed in units of  $M_\odot$ , and  $D_L$ ,  $D_S$ , and  $D_{LS}$  in kpc, then  $\tilde{k}=2.85$  will give  $\Theta_E$  in mas (milli-arc-seconds).

Normalizing  $\beta$  and  $\Theta$  in Eq. 1.25 with  $\Theta_E$  to

$$u = \beta / \Theta_E \quad \text{and} \quad y = \Theta / \Theta_E \quad (1.27)$$

rewrites Eq. 1.25 to become

$$y_{\pm} = \frac{u}{2} \pm \sqrt{\left(\frac{u}{2}\right)^2 + 1} \quad (1.28)$$



**Figure 1.20.** The geometry of a source star,  $S$ , passing behind a lens star,  $L$ , seen from “the side” (upper panel; drawn as it takes place in the source-lens-observer plane), and seen as projected onto the sky (lower panel; the projected lens-plane).  $I_1$  and  $I_2$  are the two images that an observer in  $O$  will see instead of the source, while it is magnified by the lens.  $\Theta$ ,  $\alpha$  and  $\beta$  are the angles given in Eq. 1.20 (the lens equation) and  $\alpha_b$  the bending angle given in Eq. 1.21.  $D_S$  is the distance between the observer and the source star,  $D_L$  is the distance between the observer and the lens star, while  $D_{LS}$  is the distance between the lens star and the source star. In the lower left panel, the small circles are the geometrical positions of the source star,  $S$ , while the full line connects the two images,  $I_1$  and  $I_2$ , of the source at a given time. When the source moves from left toward right in the figure, relative to the lens, the images move as indicated by the arrows. The dashed circle is the Einstein radius,  $\Theta_E$ .  $\Delta\theta$  is the angular distance between the two images of the source star.

which in a convenient way allows us to compute the amplification  $A_+$  and  $A_-$  of the two images  $I_+$  and  $I_-$  (at  $\Theta_+$  and  $\Theta_-$  or  $y_+$  and  $y_-$ ) relative to the unmagnified source image. A central result from general relativity is that the surface brightness (i.e. the flux density arriving to the telescope per angular unit of the sky) of the source star is unchanged, but the area of the image is magnified as illustrated in Fig. 1.20. The magnification is therefore simply the area of the distorted images  $I_+$  and  $I_-$  relative to the unmagnified angular area of the source star, which can be expressed as

$$A_{\pm} = \frac{y_{\pm}}{u} \frac{dy_{\pm}}{du} = \frac{u^2 + 2}{2u\sqrt{u^2 + 4}} \pm \frac{1}{2} \quad (1.29)$$

and summed to the total magnification,  $A$ , as

$$A = A_+ + A_- = \frac{u^2 + 2}{u\sqrt{u^2 + 4}} \quad (1.30)$$

We see from Eq. 1.28 that for  $u > 0$ ,  $y_-$  (the solution with “-” in front of the square root) always have  $y$  less than 1, and that  $y_+$  always is larger than 1, i.e. the image  $I_+$  at  $\Theta_+$  on the source side of the lens is always outside the Einstein ring and  $I_-$  always inside the Einstein ring, and the distance  $\Delta y$  between the two images is

$$\begin{aligned} \Delta y &= y_+ - y_- = 2\sqrt{(u/2)^2 + 1} = \sqrt{u^2 + 4} \geq 2 \quad \text{or} \\ \Delta\Theta &= \sqrt{u^2 + 4}\Theta_E \geq 2\Theta_E \end{aligned} \quad (1.31)$$

At  $u = 0$  Eq. 1.28 diverge into  $|y_+| = |y_-| = 1$  and the two images merge into forming a ring around the lens position (the Einstein ring).

Likewise we see from Eq. 1.29 that  $I_+$  is always brighter than  $I_-$ , and that for large values of  $u$ ,  $A_+ \rightarrow 1$  and  $A_- \rightarrow 0$  (and  $\Theta_+ \rightarrow \beta$  and  $\Theta_- \rightarrow 0$ ), i.e. the image (obviously) converge into the normal single unmagnified image of a source star unaffected by the lens. From Eq. 1.30 (and 1.29) we see that for  $u \rightarrow 0$ ,  $A \rightarrow \infty$ , i.e. in the adopted approximation of the lens being a point source, the amplification of the light from the source star goes toward infinity if the source moves exactly on line behind the lens. This phenomenon is called a caustic point, and it plays a central role in the identification of planets orbiting the lensing stars. The fact that the source brightness gets infinite is an artifact of the point lens approximation, but also in reality the brightness will increase rapidly and substantial (e.g. a factor several hundred in brightness) over very short timescales. This is what is meant with the word *caustic – a sharp transition*.

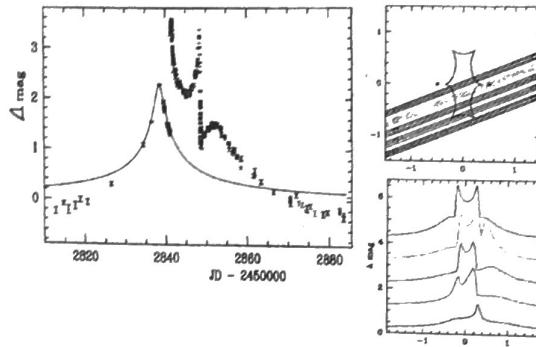
Since the source is moving relative to the lens,  $u$  is a function of time,  $u = u(t)$ . Let now  $\mu$  be the angular velocity of the source,  $t_0$  the time of closest approach between the source and the lens, and  $u_0$  the closest approach (angular distance in units of the Einstein radius) between the lens and the source ( $u_0 = u(t_0)$ ). Then

$$u(t) = \sqrt{u_0^2 + \mu^2(t - t_0)^2/\Theta_E^2} = \sqrt{u_0^2 + ((t - t_0)/t_E)^2} \quad (1.32)$$

where  $t_E = \Theta_E/\mu$  is the time it takes for the source to move one Einstein radius in the projected lens plane. By introducing Eq. 1.32 into Eq. 1.30 we can therefore now plot the amplification as function of time. The smaller  $u_0$  is the higher the maximum amplification. We can also from observations of the light curve determine from Eq. 1.32 and 1.30 how close the source and the lens passed one another on the sky, and how long time it will take for them to separate one pixel in our camera system, such that they can be observed as two individual stars in the sky.

Before proceeding to the discussion of planets, we will first examine the above equations with some examples.

A solar radius at a distance of 1 kpc is seen under an angle of 0.0045 mas. Typically the lens will be a star of a bit smaller radius than solar at a distance  $D_L$  of a few kpc, while the typical source will be a K-giant ( $R_* \approx 20R_\odot$ ) in the Galactic bulge ( $D_S \approx 8$  kpc). Hence, the angular size of the lens will typically be 0.001 mas and that of the source typically 0.01 mas. The typical size of the Einstein ring will (from Eq. 1.26) be  $\tilde{k}/\sqrt{10} \approx 1$  mas. We therefore see that it is a good approximation to assume that the lens and the source are



**Figure 1.21.** The observations in the left panel are obviously very different from the best fit model derived from Eq. 1.20 to 1.32 (the full drawn bell shaped curve). In particular one note an abrupt (i.e. *caustic*) deviation starting at day 2840 and lasting  $\sim 20$  days to  $\sim 2860$ . Attempt to model the observations with a binary star of two approximately equal mass stars is shown in the panels to the right, where the upper panel show the position of the two stars, the caustic region between them, and 5 different possible source star tracks through the system. The lower panel shows the corresponding model light curves. It is seen that the middle light curve resembles the observed curve best, and this could therefore be a starting model for finding the best fit model, through many more simulations with different parameters starting from this one.

point sources while the source is at the lensing Einstein ring, and it is still fine for  $u \approx 1/10 \sim A(u) \approx 10$ , but not good for  $u \approx 1/100$  where the angular distance between the lens and the source is of the same order as the angular size of the source. For high-magnification events it is therefore necessary to include a detailed description of the source star surface. While such inclusion makes the analysis more complicated, it also allows extra information about the lens system and the source to be extracted from the light curve data.

Note that for a typical ground based observation, the stellar angular image size on the detector (i.e., the full half width of the stellar point spread function) is around 1 arc second; i.e. ca 1000 times larger than the size of the Einstein ring. In contrast to the cosmological gravitational lenses, where the distorted image of a distant quasar is seen at several places along the Einstein ring of a cluster of galaxies, the stellar microlensing images are not resolvable even in our best telescopes. Even the smallest angular pixel size of any detector (the HST) is 0.04 arc second or about 50 times the size of a typical stellar Einstein ring. With a relative angular speed of  $\Theta_E/\text{month}$  it will therefore take a few years for the source and the lens to move a pixel apart, and probably a decade or more before they are separated enough that a direct photometric identification of the lens star, and maybe even a spectrum, can be obtained. In a few years it will be possible to make direct observations of the most interesting of the lensing stars that have proven to be orbited by exoplanets.

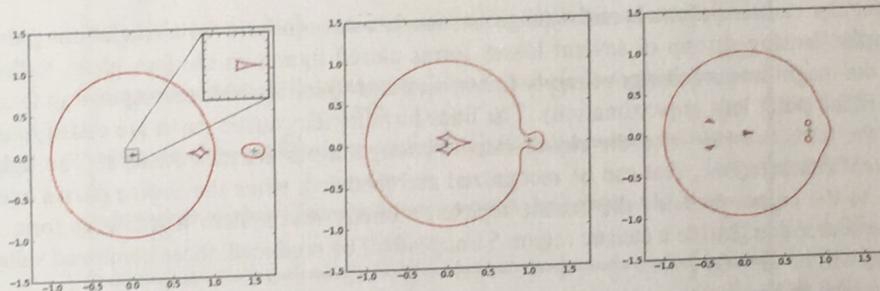
We note that when  $u = 1$ , then  $A(u) = 3/\sqrt{5} = 1.34$ . I.e., when the source passes the Einstein ring, the amplification of the light from the source is 34%, or  $\Delta\text{mag} = 0.32$ . Hence, the change in luminosity is easily observable, even with a modest telescope and for stars at quite large distances (compare e.g. with the obtained accuracy of 0.0053 mag in the HST observations of  $m_v = 18.4$  main sequence stars in 47 Tuc, 13,000 light years away, which is not far from a typical source star distance). Magnifications of a few full magnitudes when the source crosses inside the Einstein ring of the lens, are not uncommon.

#### 1.4.2 The planetary microlensing.

Typical relative angular velocities  $\mu$  between nearby dwarf stars and further away bulge giants are about 1 mas per month, and hence the passage over the Einstein ring around a solar type star takes around  $t_{E,\odot} = \Theta_E/\mu \approx 1$  month. We see from Eq. 1.26 that the size of the Einstein ring, and hence also the crossing time, scales with the square root of the lens mass. If we therefore substitute the solar mass in Eq. 1.26 with the mass of Jupiter, we see that the crossing time of the Einstein ring of a single Jupiter-like planet is  $t_{E,J_{up}} = t_{E,\odot} \cdot \sqrt{M_{J_{up}}/M_\odot} \approx 1 \text{ month} \cdot \sqrt{10^{-3}} \approx 1 \text{ day}$ . Similarly the crossing time over the Einstein ring around an Earth-mass planet is  $t_{E,\oplus} = t_{E,\odot} \cdot \sqrt{M_\oplus/M_\odot} \approx 1 \text{ month} \cdot \sqrt{3 \cdot 10^{-6}} \approx 1 \text{ hour}$ . Discovery of stellar lenses therefore just requires single observations with days interval, while characterization of Jupiter-mass and Earth-mass objects requires

an observation density of respectively few hours and few minutes, and obviously one does not know where on the light curve the planetary signal may appear before afterward. Therefore microlensing exoplanet searches is an observationally intensive effort with relatively few discoveries (compared to the transit method), but it covers a parameter space not easily accessible by any other method, including true analogues to most of the planets in our own solar system. For the same reason, microlensing surveys have typically consisted of wide field surveys that monitor a few hundred million stars in dense fields with a cadence of one or a few observations of each field per night, followed by dedicated observations in higher cadences of the selected best candidates for identifications of exoplanets. With the development of software and hardware, future microlensing surveys with a variety of telescopes, including possibly a dedicated space telescope, will be able to discover hundreds or even thousands of exoplanets in the solar system analogue regime.

The order of magnitude time estimates above were done by simply envisioning that the background star is an almost point like source passing independently through a stellar and planetary Einstein ring. This is good enough for estimating the timescales, but for predicting the combined light curve (and hence also for inverting the observed light curves to learn about the lensing system) this simplified assumption is not very useful. When the image of the source star has been distorted by the gravitational field of the lensing star, it is already two "banana-shaped" figures instead of one spherical symmetric star as illustrated in Fig. 1.20). It is one of these images that have to pass through the planetary lensing region in order that we will spot the influence of an orbiting planet as a deviation in the stellar microlensing light curve.



**Figure 1.22.** The critical curves and caustic regions for three configurations of a lensing star orbited by a planet of 100 times smaller mass than the star. The three panels show, from left to right, the situations where the planet is far from the star, at the stellar Einstein ring, and closer to the star than the stellar Einstein ring, respectively. The planet is marked with a pale blue dot, the caustics in purple colour, and the critical curves in orange.

In general we can think of a more complex lens system as consisting of  $N_L$  point mass lenses (stars and planets), each with relative mass  $\epsilon_i = m_i/M$  compared to the total mass  $M = \sum_{i=1}^{N_L} m_i$  of all the individual lensing masses  $m_i$  in the system. The total Einstein radius of the system is calculated from Eq. 1.26 (with mass  $M$ ), and the dimensionless source and image positions,  $u$  and  $y$ , could be computed as defined in Eq. 1.27, but the Einstein ring is now no longer a ring but rather a more general critical curve which for a star and a planet is not deviating too much from a ring. Fig. 1.22 show three examples for a planet that is 100 times less massive than the star (i.e. the analogue of a  $\sim 10$  Jupiter-mass planet) with the planet respectively outside the stellar Einstein ring, at the ring, and inside the ring, so-called far, resonant and near configurations of the star-planet lensing system. The critical curve of the system is the analogue to the Einstein ring in the single lens case, i.e. the mapping of the source star onto the lens plane, or the curve which a point source star will fill on the sky if the lens and the source are perfectly aligned. The caustic curves, to be discussed in detail below, are the regions a point-like source star has to cross in order to be magnified "infinite". As expected, the critical curves for a star and a planet well separated looks almost like two individual and independent Einstein rings (left panel of Fig. 1.22), whereas they deviate from this simple picture if the stellar and planetary lenses are closer to one another.

With more than one lensing object it is convenient to express  $u$  and  $y$  as complex numbers  $\zeta = u_1 + iu_2$

and  $z = y_1 + iy_2$ . The lens equation then becomes

$$\zeta = z - \sum_{i=1}^{N_L} \frac{\epsilon_i}{\bar{z} - \bar{z}_{m,i}} \quad (1.33)$$

The magnification  $A$  of the source star (the light curve) can now be computed in the same way as in Eq. 1.29 for the one-object lens, but with more objects in the lensing system, there will now be more than two images of the source. Although the surface brightness of each image is still conserved, each image may now be larger or smaller than the unlensed source, resulting in a complex combination of magnified and demagnified images, with  $A_j$  given by the inverse of the Jacobian evaluated at each image position,

$$A_j = \frac{1}{\det J}|_{z=z_j}; \quad \det J = 1 - \frac{\partial \zeta}{\partial \bar{z}} \frac{\partial \bar{\zeta}}{\partial z} \quad (1.34)$$

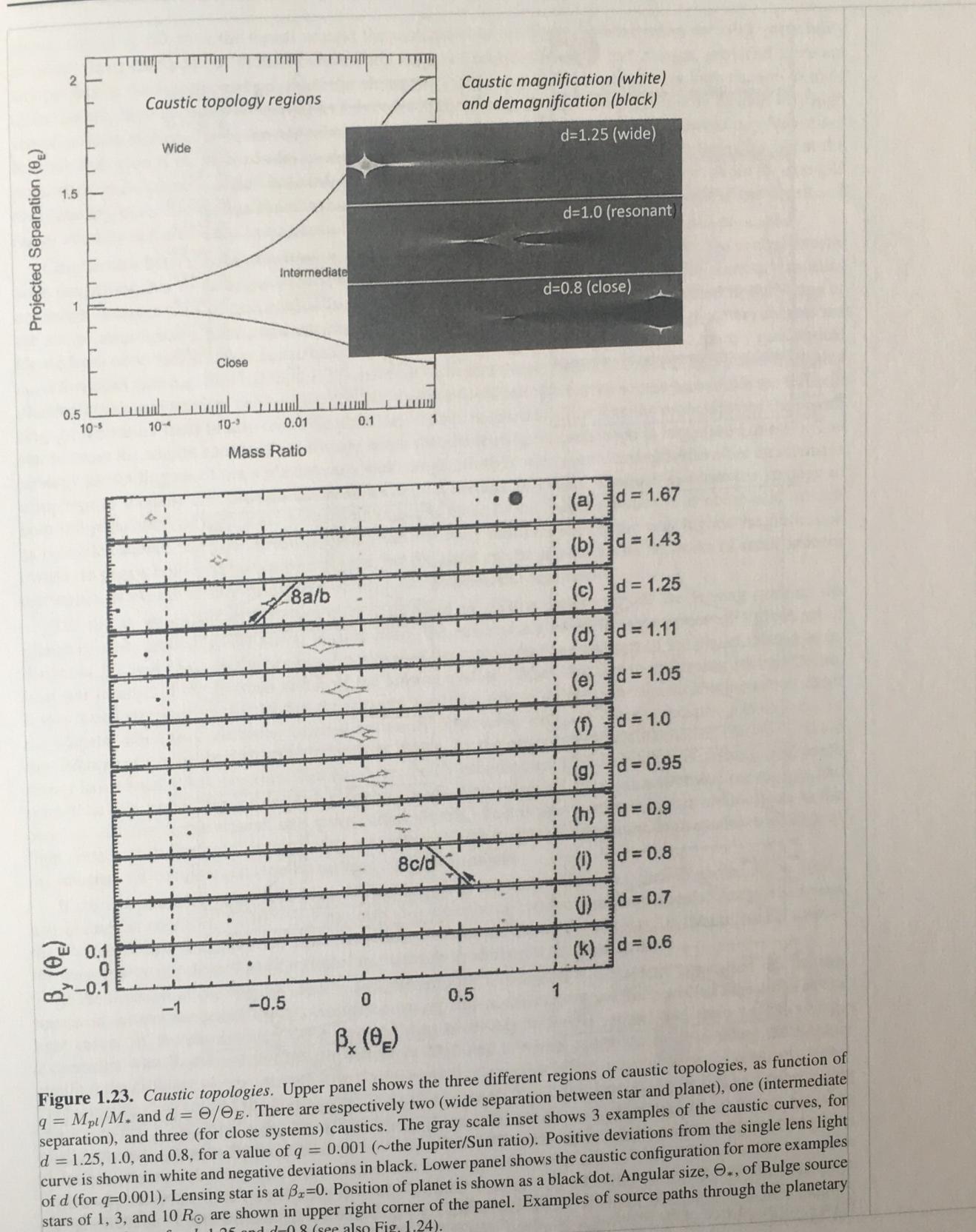
The total amplification at any given time is simply the sum of the amplification of the individual images,  $A = \sum_j A_j$ , but since some of the images are magnified and some are demagnified, the resulting deviations from the single lens light curve, due to lensing planets, can be both positive and negative, depending on the detailed path of the source through the lensing system. The total number of images,  $N_{im}$ , caused by a lensing system of  $N_L$  lenses (the star and its planets or a multiple stellar system or a combination of several stars and planets) is

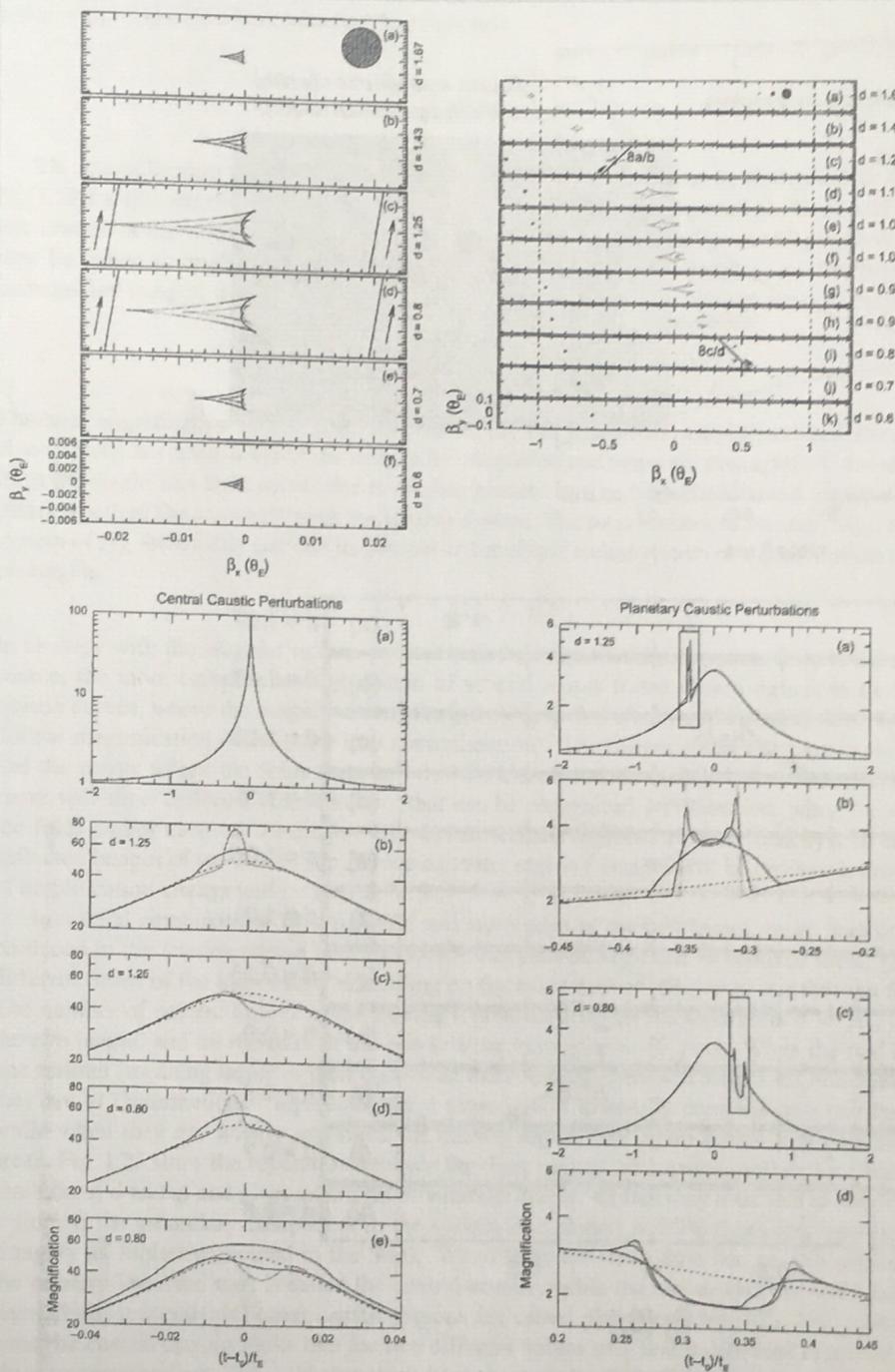
$$N_{im} \leq 5(N_L - 1) \quad (1.35)$$

In analogy with the singular caustic point where the magnification becomes infinite for a single-lens point source, the more complex lensing system of several lenses forms closed figures in the lens plane, called caustic curves, where the magnification changes sharply (i.e. where  $\det J=0$  in Eq. 1.34, corresponding to an infinite magnification in the point lens approximation). The lines forming the caustic curve are called *folds* and the points where the folds meet to form the closed caustic configurations are called *cusps*. The light curve will show different characteristic, that can be recognized and modeled, when the source passes over the folds and/or close to the cusps. Outside the caustic regions, a binary lens system will always form 3 deflected images of the source star. Inside a caustic region 5 images will be produced, those combined value of amplification always will be  $A \geq 3$ .

In several combinations of separation and mass ratio of the two lenses, more than one caustic curve is produced in the lensing region, and the source can pass through one of more of them, giving rise to many different forms of the light curve, depending on the exact path of the source star through the lensing system. The number of caustic figures in the lensing region depends on the separation  $d = |z_{m,1} - z_{m,2}|$  between the two lenses, and on the ratio of the two lensing masses,  $q = m_1/m_2$ . When the two lenses are close to one another (meaning inside  $\sim$  their combined Einstein ring) they will form 3 disjoint caustic regions. When they are at "intermediate" separation the 3 regions will gradually combine into one large single caustic, while when they are widely separated the caustic region will again divide into now two distinct caustic areas. Fig. 1.23 show the separation between the three regions of caustic topology for a binary lens system as function of  $d$  and  $q$ , and gives examples of what the caustic figures may look like in the different topological regions if the secondary component in the system (the planet) is 1000 times less massive than the primary (roughly as Jupiter compared to the Sun). When there are more than one caustic region, the one close to the primary lens (the star) is called the central caustic, while the one or two (for wide and close star-planet separations, respectively) outer caustic regions are called planetary caustic(s). Fig. 1.24 show examples of what the central caustic looks like for two different values of  $q$  and 6 different values of  $d$  in the wide and close caustic topology. Fig. 1.24 also show what the corresponding light curves would look like for examples of source paths through central and planetary caustics, respectively.

In order for the source to pass close to the region of the central caustic, the minimum angular distance  $u_0$  (or  $\zeta_0$ ) between the source and the lensing star has to be very small, as illustrated in Fig. 1.23. On the other hand, any such encounter, within say  $u_0 \leq 0.02$  (i.e.  $\beta \leq 0.02\Theta_E$ ) corresponding to  $A \geq 50$ , will give rise to a planetary signal, for any sufficiently large planets in orbital distance between say 0.7 and 1.5  $\Theta_E$  (corresponding to 1.5-3.0 AU for a typical M-dwarf and 2.5-6 AU for a solar type star, in a typical lensing distance of 4 kpc). Therefore we can a priori conclude that if a light curve brings the maximum amplification





**Figure 1.24. Caustics and amplification.** Upper left panel shows the central caustics ( $\beta \sim 0$  in right panel) and four examples of source paths through or close to the caustic regions for  $d=1.25$  and  $0.8$ . Lower left panel shows the corresponding light curves and caustic deviations. Dotted lines are for no planet, and the 3 solid lines are for a bulge source star of  $\Theta_s=0$ ,  $0.003$ , and  $0.01 \Theta_E$  (plus  $0.03$  in the lower most panel). Upper right panel shows two examples of source paths through the planetary caustic of a system with  $d=1.25$  and  $0.8$ , respectively. Lower right panel shows the corresponding deviations from the single lens light curve for the same choices of the source star as in the left panel. Panel (b) is a magnified view of the deviation in (a), and (d) correspondingly of (c). Larger source stars smooths the deviation.

## 1. EXOPLANETS

---

above, say,  $A = 50$ , then the signal around the maximum of the light curve has close to 100% probability to reveal any planets in the system over a wide range of orbital distances and masses, provided there are any planets in the system, and provided the photometric accuracy and the cadence are high enough to trace them. Dense, high-accuracy observation of the very highest magnification events therefore contain very high statistical information about planetary abundances around the important "snow-line" of planetary formation. Remark that even if no planet is detected, such light curves contain strong statistical information about the population of exoplanets, which has made it possible to give valuable statistical estimates about for example the planetary mass distribution function based on relatively few detected planets (compared to the transit and radial velocity detected exoplanets), but many observed light curves.

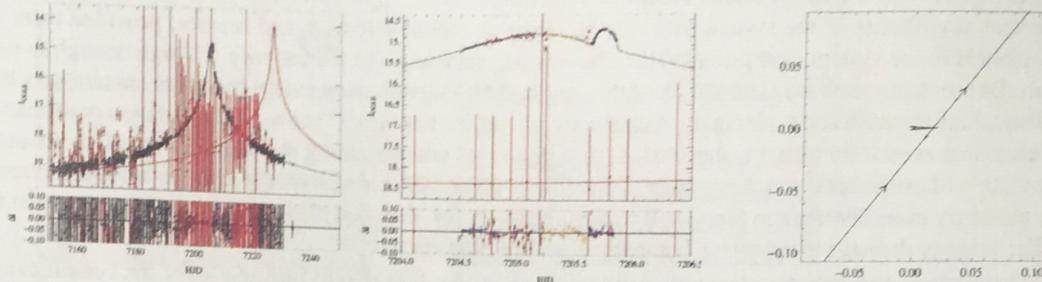
Comparison between the planetary magnifications in Fig. 1.24 due to passages over the central caustic with deviations due to passages over the planetary caustics in Fig. 1.24, show that the planetary caustics are much stronger than central ones. The strongest planetary signal is therefore expected in the wings of the stellar microlensing curve, and therefore the widest range in orbital distances and planetary masses are traced from observations away from the maximum of the light curve. On the other hand, purely geometrical considerations (see e.g. Fig. 1.20 and 1.23) makes it clear that the probability of revealing a planet through a planetary caustic is smaller, because the probability for a source star that cross somewhere inside the Einstein ring (of the stellar lens) to also cross the planetary caustic is much smaller than the probability of the source star to cross the central caustic if we already know that the lensing magnification is large (see Eq. 1.37). The strategy for finding small mass planets on a wide range of orbits with microlensing is therefore an intensive campaign to monitor the full light curve of as large a number of events as possible, while the strategy to map the population of high mass exoplanets in orbits close to the Einstein radius is to observe in as high as possible cadence and photometric accuracy the relatively small number of the very highest magnification events. In praxis both strategies are employed, but the latter can be achieved with networks of small amateur telescopes, while the former demands medium sized professional equipment.

The theory described so far gives only a snapshot of relative parameters of the lensing system: the planetary mass relative to the stellar lensing mass, the stellar mass under the assumption of a given set of distances  $D_L$  and  $D_S$ , and the projected position of the momentaneous position of the planet relative to its host star in units of the Einstein radius of the lensing system. When the first microlensing surveys began, it was therefore expected that all that the surveys would be able to deliver was statistical information about the planets, albeit in a parameter interval not easily observable with any other technique, and in areas of the Galaxy not reachable with other techniques either. As the observational technique has matured and the theory has refined, it has, however, fortunately gradually become clear that the method is capable to do much more than this, and in many cases can pinpoint the absolute values of the mass and orbital parameters, the absolute distance to the system, and several other details. This is due to subtle higher order effects in the light curve which, however, requires very accurate and high cadence photometry, high spatial resolution of the images, and complex and time consuming computer analysis.

If the source star is a giant star in the Bulge (which is most often the case), then its radius  $R_* \approx 10R_\odot$  and its angular radius  $\Theta_* \approx 10^{-2}\Theta_E$ , meaning that when  $A \approx 100$  the minimum angular distance between the lens and the star is of the same order as the source diameter. Already when  $A \approx 10$ , the size of the source, the *finite source effect*, need to be included for accurate modeling of the light curve.

If the duration of the lensing event is long enough the non-linearity of the lens movement relative the source movement due to the Earth's orbital movement causes measurable *orbital parallax* deviations in the light curve. If the observations are performed simultaneously from the ground and from space (as e.g. a campaign with Rosetta in 2013, with Spitzer in 2015 and a planned campaign in 2016 using the Kepler satellite) the different viewpoints will give rise to a *satellite parallax* effect, which due to the relatively large Earth-satellite distance and the high sensitivity to alignment of the source-lens, can give rise to a substantial effect (see Fig. 1.25). Finally the magnification during passages of the caustic curves are so abrupt that the small *terrestrial parallax* due to simultaneous observations from different locations on the Earth are traceable in the light curve.

The typical event time of a lensing is  $\approx$  one month. This is long enough that the planet-star axis will change direction, which will manifest itself in the light curve if observed accurately enough, and for a binary or multiple lens system even the caustic pattern will change, which is observable if the caustic passage is



**Figure 1.25.** A high-magnification event observed in 2015 simultaneously from the ground and from the Spitzer space telescope. The left panel show the light curve observed by several observatories from the ground (left curve) and from space (right curve) during 100 days from HJD 7150 to 7250. Remark the large space parallax shift of  $\approx 3$  weeks between the peak magnification seen from ground and from Spitzer (and the, in this case, much poorer photometric quality obtainable from space). The clear planetary deviation at HJD day  $\approx 7206$  (near the top, seen best in the middle panel) indicates a planet of  $15 M_{\oplus}$  in an orbital radius of 2 AU (assuming the lens is an M-dwarf at 4 kpc distance). The right panel show the central caustic (compare Fig. 1.23) and the computed best fit source path.

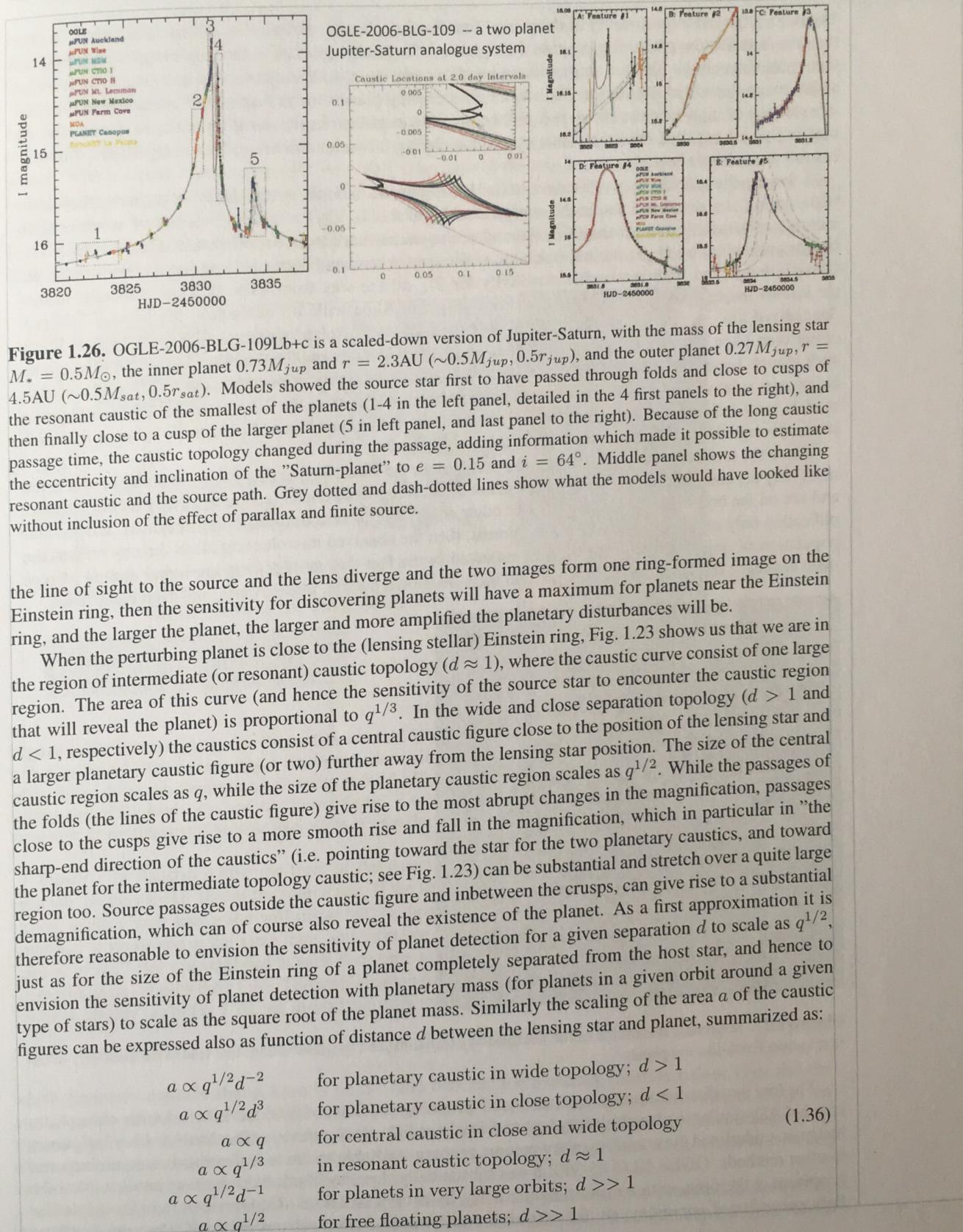
long enough or if the configuration of the source track is such that more than one caustic is crossed. Detailed analysis of these *orbital motion* effects can give information about the projected orbital velocity of the lensing planet(s), as illustrated in Fig. 1.26.

The more of these second order effects one is able to extract and model from the light curve, the better confined the system becomes. If, for example, the finite source deviation is well confined from the light curve, then one can determine  $\Theta_E$ .  $D_L$  can usually be assumed to be the distance to the Galactic centre (or be derived from photometry). With  $D_L$  and  $\Theta_E$  known, Eq. 1.26 gives a mass-distance relation for the lens. If further the parallax deviation can be characterized, then the physical size of the Einstein radius can be determined, which together with  $\Theta_E$  gives us  $D_L$ , and hence the absolute mass of the lens. For several of the observed microlensing planets all of these parameters have been derived, and for some even additional parameters for the orbit, like eccentricity and inclination, could be derived. In the future, with continued improved equipment and computer codes and capacity, it will be possible to routinely derive this full description for an increasing number of all the observed microlensing planets. If spectroscopy can be obtained during the passage of the caustic, additional information can be obtained about the source star, such as the limb darkening profile, or even such details as the velocity field of the upper stellar atmosphere.

An accurate quantitative analysis of the light curve is typically done by running simulations of thousands of different configurations of  $d$ ,  $q$  and source paths through the lensing system, and estimating which one results in a light curve most resembling the observed one (i.e. having the lowest chi-square). The more second order phenomena (and hence parameters to fit) are included in the analysis, the more complex and time consuming it obviously becomes, and the more accurate photometry it requires (to characterize the often slight deviations related to the second order effects). The development of the details of the theory, hence foster an increasing demand for higher observational accuracy, and visa versa. Only few computer codes in the world can handle the binary problem fast and accurate, and even fewer can model multiple systems such as a star with several planets with all the theoretically known second order effects included. However, also without this accurate full numerical characterization, one can obtain a good understanding of the qualitative aspects of the phenomenon, even from considering just the simple single-lens system.

In first approximation the binary system of a star and a planet can be well understood from the equations 1.20 to 1.32. The positions of the "banana-formed" images of the source star in Fig. 1.20 are roughly the same as in the single-lens approximation, because the star dominates the total lensing gravitational field of the star and the planet. Hence, the planet has to be close to one of the two images in order to perturb the light curve. However, the images of the source are not small and symmetric (as the source star itself), so the place of "infinite magnification" is no longer a point (as in Eq. 1.30), but a caustic curve, as shown in Fig. 1.21, 1.23, 1.24 and other figures. The source has to cross the caustic curve instead of the lens star to be "infinitely" magnified. Since the images are largest (and hence the source magnification strongest) when

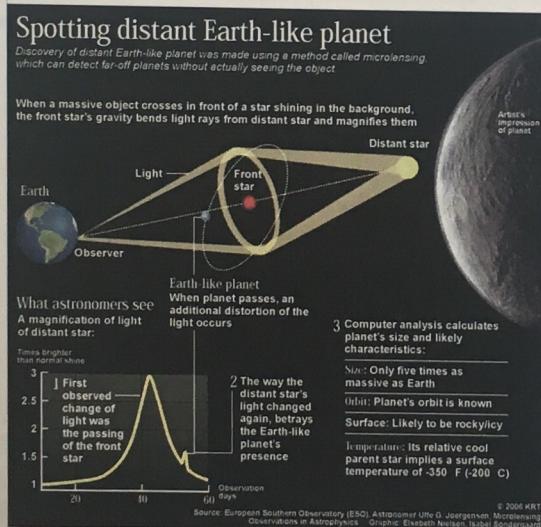
## 1. EXOPLANETS



The ratio between the probability of detecting "free floating planets" and planets in very large orbits, is just the ratio between the probability to pass through the (effectively separated) planetary magnification region (i.e. approximately its Einstein ring) with and without passing close enough to the (host) star to detect the stellar magnification of the source. With more microlensing planets in the future (and in particular with the introduction of high-cadence surveys) it will be possible to give strict statistical estimates of the number of planets that are expelled from the planetary systems during formation relative to how many statistically stay in orbit around their host star.

Likewise the probability for discover stellar lensing events (and therefore planets of a given type) scale with lensing star distance relative the source star distance ( $D_L/D_S$ ). Since the density of (source) stars increase substantially toward the Galactic centre, the source stars are most often near  $D_S \approx 8$  kpc. The largest probability of  $D_L$  is around half way to the source star, and hence  $\sim 4$  kpc for surveys toward the Galactic bulge, but with reasonable sensitivity for  $D_L$  all the way from 1 kpc to 8 kpc (reflected in that the known lensing planets have distances from 400 pc to 8 kpc with 1/3 of them in the interval 3-5 kpc). Specialized searches would in principle be able to discover microlensing exoplanets orbiting stars much closer than 1 kpc as well as as far away as in the Magellanic clouds and the Andromeda galaxy, although no such planets have been identified yet. The fact that the detection rate of microlensing planets associated to bulge stars relative to those of disk stars is considerable smaller than the 3/2 ratio expected from Galactic models, may indicate that bulge stars have a substantially lower abundance of (detectable) planets than disk stars, although the statistical significance of this is still small.

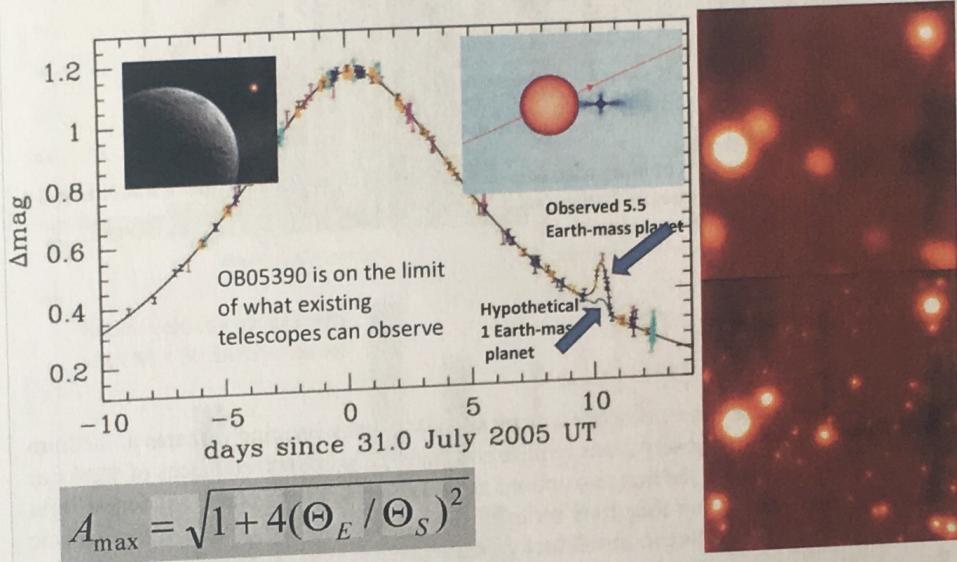
Since the probability of a source star to pass through the Einstein ring of a foreground object of a given mass only depends on the square root of the mass of the foreground object (and only weakly on the distance, and not on the brightness or activity level or other observable or selection-disturbing features), and the amplification only on the geometry of the alignment, then the observed microlensing alerts directly reflects the population of objects of different masses, reflected in the fact that most planets are found around M-type main sequence stars, but that the sample also includes brown dwarfs, white dwarfs, and even black holes.



**Figure 1.27.** Schematic illustration of the microlensing technique and the observations that lead to the discovery of the first known Earth-like exoplanet.

The first two discoveries of exoplanets by microlensing (in 2004 and 2005) revealed a Jupiter-mass planet at  $\sim 2\text{-}4$  AU, similar to planets already known from the radial velocity surveys. The third microlensing planet, however, underlined the complementary parameter space reachable by the microlensing technique compared to other methods. OGLE-BLG-05-390Lb was substantially closer to "Earth-like" than any previously known exoplanet at the time, with a mass of only  $5.5 M_{\oplus}$  in an orbit of 2.6 AU. The deviation from the single star light curve lasted approximately 6 hours, and was observed from a network of small telescopes in Chile,

New Zealand and Australia. A combination of finite source effect (remark the relative size of the source and the planetary caustic in the simulated source path in Fig. 1.28), photometric analysis of the source star, and bayesian analysis, estimated the source star to be a G-type giant with  $R \approx 10R_{\odot}$  situated close to the Galactic centre, and the lensing star to be an M-dwarf with a mass of only  $0.2M_{\odot}$ , at a distance of 6.6 kpc. The observations do not reveal anything about the composition, but the fact that its mass is well below the theoretical limit of  $\sim 10M_{\oplus}$  for the formation of gas-planets, tells us that it is very likely to be a solid planet. With a surface temperature of  $T \approx -200^{\circ}\text{C}$  because of the low luminosity of the host star, it might be rocky like the terrestrial planets or icy like Kuiper belt objects in our solar system. The light curve can qualitatively be well understood from Eq. 1.20 - 1.32 of the single lens analytical solution as illustrated in Fig. 1.27 from a popular announcement of the discovery, with the main difference that the planetary deviation lasts considerably longer than the simple prediction, due to the smearing out caused by the finite source effect.



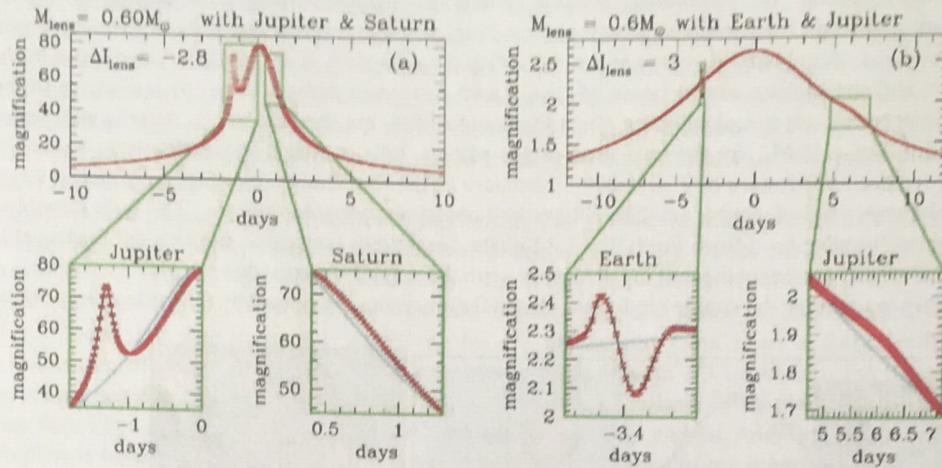
**Figure 1.28.** The two right panels show a crowded field of the central region of the globular cluster M3, photographed with a traditional CCD camera (upper panel) and a lucky imaging camera (lower panel), respectively. The left panel of the figure shows the discovering light curve of the  $5.5M_{\oplus}$  planet OB05390Lb together with (upper right inset) the relative size of the source star and the “Einstein ring” (i.e., the caustic region) of the planet, and the source star path through it. Upper left inset is an artist impression of what the planet could look like.

Figure 1.24 illustrated how the form of the caustic deviation from the single star light curve depends critically on the angular size of the source star. If the source is “point-like” the deviation becomes short and strong, while a large source star will smear out the deviation – the different regions of the source star is not amplified to maximum at the same time, and therefore effectively the deviation consist of many less bright “pieces of star” that are amplified at each their time and thereby contributing to the total amplification (or deamplification) at each their time. The amplification (or deamplification) therefore does not reach anything “close to infinity” as in Eq. 1.30, but instead can only reach a maximum amplification

$$A_{\max} = \sqrt{1 + 4(\Theta_E / \Theta_S)^2} \quad (1.37)$$

which obviously depends almost linearly on the ratio between the planetary Einstein radius,  $\Theta_E$ , and the angular radius of the source star,  $\Theta_S$  (reflecting how large a fraction of the source star which at any given time can be “inside” the planetary Einstein ring). Since  $(\Theta_E)^2 \propto M_E$  and  $(\Theta_S)^2 \propto L_S$ , then finding say 10 times smaller mass planets requires identifying 10 times less luminous background stars, which in crowded fields is determined by the resolution of the images.

Therefore high spatial resolution in the images from which the photometry is extracted is central for detecting lower mass planets around the lensing star. This can be reached either from space (like HST),



**Figure 1.29.** A simulation of two planet systems orbiting a  $0.6 M_{\odot}$  star. Left panel show the situation of a Jupiter and a Saturn system with the same orbital radii and masses as Jupiter and Saturn in our solar system, and the right panel show the same for a system with a Jupiter and an Earth analogue.

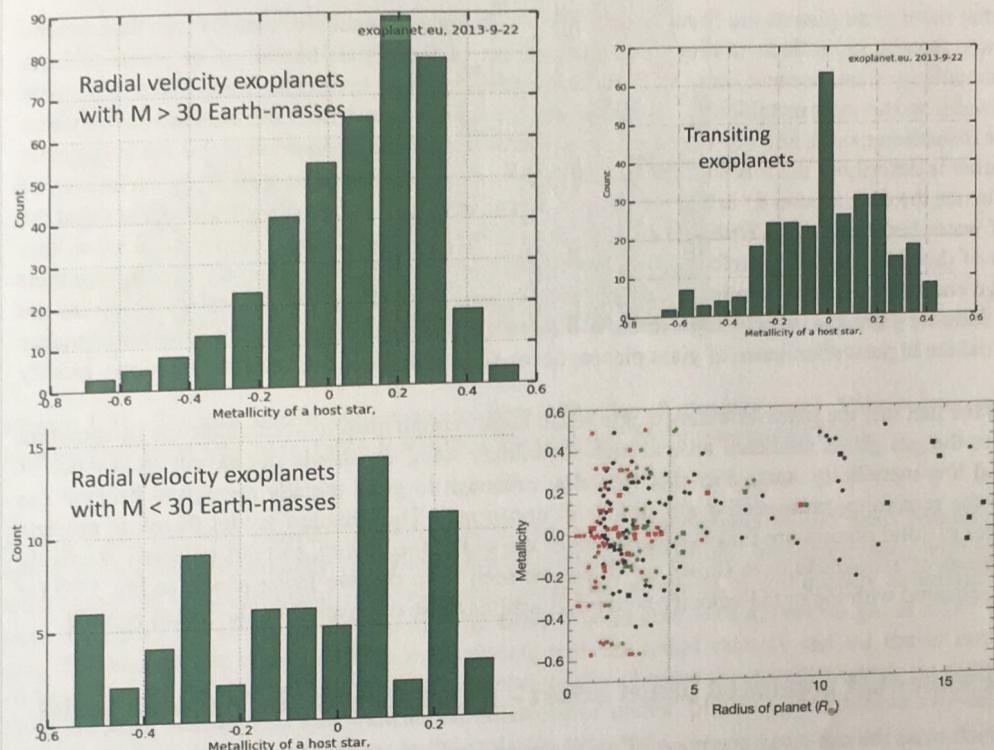
from instruments with adaptive optics at large telescopes, or by so-called lucky imaging cameras at medium sized telescopes. Given that microlensing observations require continuous observing sequences of weeks or months every year in order to be successful, the first two options are very costly, and in praxis not doable right now. For the medium sized telescopes where long time series of observations can be obtained, the realistic option to reach small mass planets, is therefore to install lucky imaging cameras. These are fast cameras that take advantage of the fact that the timescale of the atmospheric turbulence is of the order of one tenth of a second. By stacking many images of 0.1 second exposures instead of obtaining one long exposure of say 2 minutes, one can therefore reach a resolution close to the theoretical diffraction limit, just as from space based telescopes. as illustrated Fig. 1.28.

From space a dedicated satellite for microlensing would be able to obtain the same kind of advantage as the Kepler satellite for transit observations, namely a very high accuracy in the photometry and continuous observations uninterrupted by the Earth's weather system and potentially uninterrupted by the day-night cycle. Figure 1.29 show a simulation of what a Jupiter-Saturn two planet system analogue and a Jupiter-Earth analogue system orbiting a  $0.6 M_{\odot}$  star could look like (depending on the source path) as seen from a proposed dedicated microlensing satellite.

The  $\sim 40$  planets listed in exoplanet.eu (as of mid-2015) have orbits anywhere between Mercury and Neptune sized, and masses from below  $1 M_{\oplus}$  to several Jupiter-masses. One fourth are in the Earth-Neptune mass regime and half of the planets are in Venus-Mars sized orbits, indicating that such planets are abundant. The ability to draw relatively strong statistical conclusions from relatively few detected microlensing planets are well illustrated by the analysis of OGLE-2005-BLG-169Lb. This is a Uranus-like planet in terms of mass and surface temperature, situated in a 3 AU orbit around a  $0.5 M_{\odot}$  star. Analysis of the light curve at the same time excluded the existence of Jupiter-mass planets in the 0.5 to 15 AU interval and Saturn-mass planets in the 0.8-9.5 AU interval. By combining such exclusion analysis from all the observed microlensing events (those with planets detected as well as those without observed planets) it has been possible to conclude that there are roughly 50 times more Earth-Neptune sized planets than there are Neptune-Jupiter sized planets orbiting main sequence stars in the 1-5 AU region. This is in qualitative agreement with the Kepler results for smaller orbits, and with theoretical predictions from the core accretion model.

## 1.5 Some major results from the first years of exoplanet research

### 1.5.1 Giant planet metallicity dependence – a strong support for the standard model



**Figure 1.30.** The giant planets discovered with the radial velocity technique (upper left panel;  $M_{\text{planet}} > 30 M_\oplus$ ) are much more abundant around high-metallicity stars, while for the corresponding low mass planets (lower left panel;  $M_{\text{planet}} < 30 M_\oplus$ ) there is no dependence between stellar metallicity and number of detected planets. The same effect is seen for the Kepler transiting exoplanet candidates where good stellar metallicity is known (lower right panel), where planets with  $R < 4 R_\oplus$  show no metallicity dependence, while larger planets are found mainly orbiting high metallicity stars. For the transiting exoplanets observed from the ground (upper right panel) there is no clear metallicity dependence, apart for a tendency that most of the planets are around solar metallicity stars, which almost certainly is due to poor metallicity estimates.

The standard model for the formation of solar systems, as described in previous chapter, explains that one can draw a virtual line in the proto-stellar disk out of which the planets form, where the temperature was such that water ice could condense to solid particles (snow and ice) outside this line, but not inside. Inside the snow line only high temperature condensates such as metal and minerals that make up rocks on present day Earth could participate in the build up of the planets. Such elements this material is made of is rare from a universal point of view. Therefore the inner, terrestrial planets in our solar system became small and solid. Outside the snow-line similar material could of course also condense out of the nebula, and one can envision that the outer planets contain an inner core very similar in size and composition to our own Earth. Out there, however, also water could condense to solid material, and water accounts for several times more mass than all the other condensable material combined. Therefore the solid cores of the outer planets therefore quickly grew much larger than the Earth. When the size of these objects exceeded somewhere around maybe  $10 M_\oplus$ , their gravitational fields became strong enough to make the surrounding hydrogen and helium gas in the nebula collapse onto them, and a gas planet was formed. One of the most important conclusions from

the first decades of discovery of exoplanets, is that the giant planets that can be discovered with the radial velocity method, are much more abundant around high metallicity stars than around low metallicity stars. Fig. 1.30 shows the number of known exoplanets as function of the metallicity of their host stars, discovered from radial velocity studies and from transit surveys, and for various planetary mass groups. There is a marked peak in number of giant planets for stars of  $[Fe/H] \approx -0.25$ ; i.e. for stars slightly less than twice the solar metallicity.

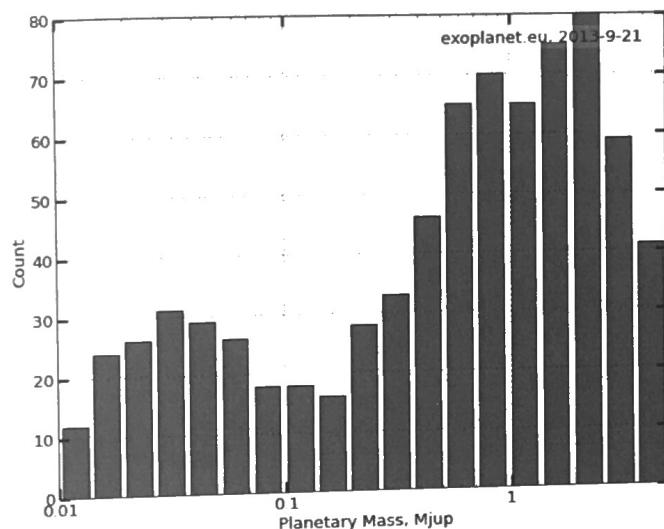
The fact that more giant planets are found around stars with higher metallicity than the sun than around stars with lower metallicity, reflects a very pronounced effect of the higher likelihood of giants to form around high metallicity stars, because these stars are themselves very few in number compared to stars with metallicity lower than the solar metallicity. It is a tendency we should expect from the standard solar system core accretion formation model, and one of the strongest evidences that the standard model is not completely wrong, but rather is describing the formation of planetary systems in general quite well. If the metallicity of the star (and hence the disk around it) is high, it includes a high abundance of oxygen, which means that the abundance of water becomes high. With a large amount of water, it "snows wildly" outside the snow line, and the cores of the giant planets therefore grows rapidly, making them quicker reach the limit where their gravity is large enough for the surrounding hydrogen and helium gas to collapse onto them. This makes the likelihood of forming giant gas planets large, and it is therefore around these stars the standard model makes us expect to find the highest abundance of giant planets observable with the radial velocity technique, exactly as seen.

Since it is the fact that the planetesimals quickly reach large enough mass to accumulate the surrounding gas that makes the gas giants abundant around high-metallicity stars, we should not expect an absence of giants around low-metallicity stars, only that they don't manage to grow quickly enough to become gas giants before the remaining proto-stellar gas disk is disappearing. The standard model therefore predicts that the number of solid planets are fairly independent of the stellar metallicity, and this tendency is exactly what observations of exoplanets have shown us, as can be seen from the two lower panels of Fig. 1.30 for exoplanets discovered with the radial velocity method as well as those discovered by the transit method.

### 1.5.2 The somewhat populated planet desert – a challenge for the standard model

Exactly at which mass the run-away accretion of hydrogen and helium gas sets in, is still an unsolved question. Some models predict the gas collapse to begin already at  $5 M_{\oplus}$ , while some models predict a number as large as  $30 M_{\oplus}$ . In our own solar system, the total mass of Uranus and Neptune are  $13 M_{\oplus}$  and  $17 M_{\oplus}$ , respectively, so the gas accretion must have begun substantially before planets grew to  $30 M_{\oplus}$ ; at least in our solar system. Both Uranus and Neptune have a substantial gas envelope, but also a substantial fraction of the total mass in their solid metal-rock-ice core. Inspired by the models of a transition mass below which the planets will be of pure metal-rock composition (like the terrestrial planets) and above which a gas collapse would quickly transform them into Jupiter-Saturn like gas planets, many scientists have a tendency to now call Uranus and Neptune for failed ice-giants. This is meant to indicate that the cores of Uranus and Neptune became large enough to accrete some gas from the surrounding nebula envelope, but never grew large enough to reach the size when rapid run-away gas accretion can take place, before the nebula was gone.

The idea that Uranus and Neptune began accreting surrounding gas exactly when it was almost gone may seems an unlikely coincidence, but on the other hand, also Jupiter and Saturn must have reached approximately the mass they have now at more or less exactly the time the remaining nebula disappeared. Had there still been a substantial nebula left after Jupiter and Saturn formed, friction of the new-formed planets and the remaining nebula gas would quickly had made them spiral into the inner solar system, just as it seems to have happened in most exoplanetary systems with observed giant exoplanets. Had the nebula disappeared before Jupiter and Saturn had reached the limiting mass that allowed them to make the surrounding nebula gas collapse onto them, they would have stayed solid metal-rock-ice planets (as may be the most common scenario of planet formation since so few stars seems to have any gas giant planets at all). Therefore it is not impossible that Uranus and Neptune reached their gas accretion mass exactly at the moment when the nebula disappeared from the system, and hence are truly failed ice-giants, that almost managed to become large enough to accrete substantial amounts of gas from the surrounding nebula, but just failed. In any case,



**Figure 1.31.** The total number of known exoplanets as function of planetary mass. There is some indication of a bimodal distribution with fewer planets (a so-called “planet desert”) for masses around  $10M_{\oplus}$ .

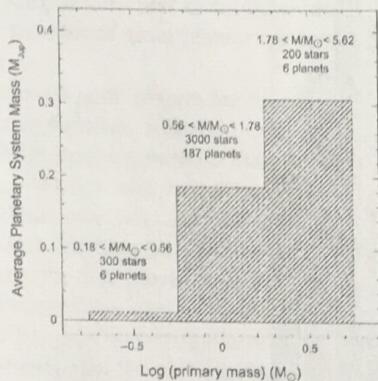
if such a gas-accretion planetary transition mass exist, we should expect to find many planets with masses below it, very few just around it (but among them Uranus and Neptune), and again many with masses somewhat above the transition mass. If the run-away gas accretion sets in, it will quickly accumulate lots of gas once it begins. The result should therefore be either a solid core with almost no gas at all or with a lot of gas. The bias toward observing high-mass planets with the radial velocity and the transit methods is still so large that it is hard to say if we are able to statistically observe the transition mass (in particular if it is close to  $5M_{\oplus}$ ), but Fig. 1.31 may indicate the existence of such a “planet-desert” around a transition mass of  $\sim 10 M_{\oplus}$ , but it is far from as obvious as theoretically expected.

### 1.5.3 The disk and planet mass as function of the host star mass

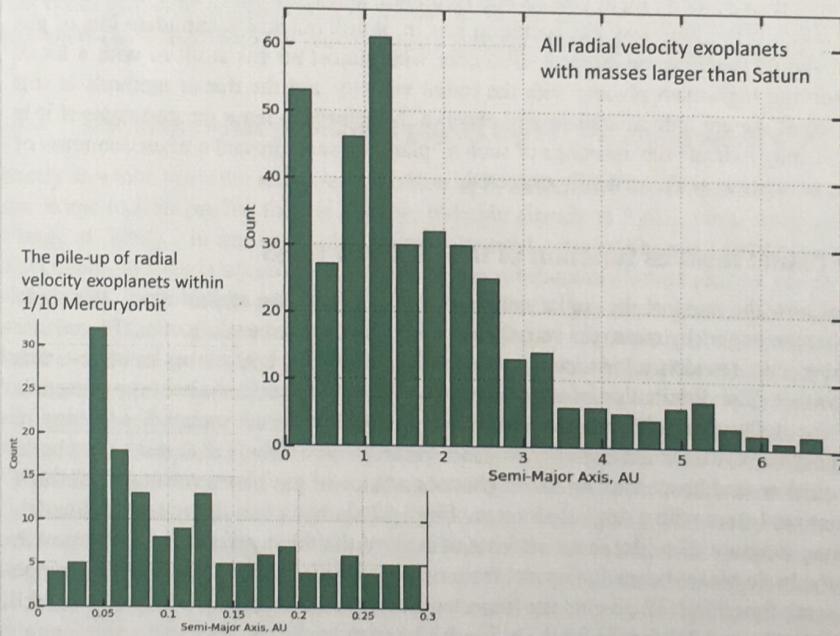
Fig. 1.32 gives us a hint of how the mass of the stellar nebula disk depends on the stellar mass. We could intuitively have guessed that the larger (by mass) the star, the larger is also the surrounding disk. There is no obvious observational or theoretical evidence for such a correspondence, but Fig. 1.32 makes us believe that the intuition is probably correct. The distribution of exoplanets show that the average mass of the planets is an increasing function of the stellar mass. In analogue with the argument for higher metallicity leading to more likely formation of gas giants, we should expect that also a more massive nebula disk makes the build up to gas giant planet go quicker, and hence leading to a higher abundance of gas giants (or higher mass of the individual gas giants) around stars with a large disk mass. Fig. 1.32 show us that the average planetary mass is a strongly increasing function of stellar mass, making us believe that high mass stars are born with high mass disks, and that the basic planet formation model from our solar system is correct for a wide range of stellar masses. At the same time Fig 1.32 give us the important information that only  $\sim 5\%$  of stars are orbited by a giant planet (2%, 6%, and 3% for the 3 bins in Fig 1.32, respectively).

### 1.5.4 Inward migration to the pile-up distance

Our own solar system harbours 4 giant planets at orbital radii  $\sim 5, 10, 20$  and  $30$  AU. A simple first glance at Fig. 1.33 immediately reveals that a standard planetary system doesn’t look like that. There is an almost flat distribution of giant exoplanets from what corresponds to  $\sim 1/10$  of Mercury’s orbit ( $\sim 0.03$  AU) to somewhere in the middle of the asteroid belt ( $\sim 3$  AU), and then a sharp drop off to basically nothing at the



**Figure 1.32.** The planetary mass (i.e. number of giant planets) as function of stellar mass.



**Figure 1.33.** The number of radial velocity exoplanets larger than Saturn, as function of orbital radius. Most of the planets lie in the terrestrial region, while also a substantial number lies inside Mercury's orbit, and very few in the Jupiter-Saturn region. Lower left panel illustrate the pile-up region at  $\sim 1/10^{\text{th}}$  of Mercury's orbital distance.

distances where all the giant planets are in our solar system are, and where we should expect all giant planets to form. We will return to this distribution and comparison of it with our solar system in a little while, but here first focus on the excess of giant planets at  $\sim 0.03$  AU, as seen in the lower left panel of Fig. 1.33.

If we believe in the standard core accretion model for the formation of planets, then the gas giants will always form outside the snow-line (i.e. for solar type stars in the region where our gas giants are today). Then Fig. 1.33 shows us that in an overwhelmingly large fraction of planetary systems, the gas giants will after their formation migrate inward. There are several different mechanisms that can account for such migrations, but basically they all rely on some kind of friction between the newly formed gas planets and the remnant of the nebula. The planet will lose angular momentum to the gas because of the Keplerian movement of the planet and the sub-Keplerian velocity of the gas. For those systems where strong inward migration of gas-giants do take place, theory predicts that they will migrate into very small orbits, but not so far in that they generally fall down onto the star. The reason for this stopping of the inward migration is that the expected strong magnetic field of the young star will be able to empty the inner part of the nebula gas disk, as described in detail in connection with the loss of the early solar angular momentum in the chapter on the formation of solar systems. From this theory we should therefore expect a pile up of planets very close to the star, where the original magnetic field was able to clean the nebula for gas such that the friction between the planets and the gas would have stopped. Qualitatively this is exactly what we see in Fig. 1.33. Here the number of planets in orbits considerably smaller than Mercury's orbit are shown, and we see that into approximately one third (i.e. 0.1 AU) of Mercury's orbit, the distribution follows the same flat pattern as in Fig. 1.33, but inward of that, the number of planets increase strongly as function of decreasing distance, with a maximum number at distances around a tenth ( $\approx 0.04$  AU) of Mercury's orbit.

Further in than 0.04 AU the abundance of exoplanets drops steeply again. These observations therefore indicate that a standard protostar cleans the inner part of the nebula for gas in a region out to  $\sim 1/10$  of Mercury's orbit. If young planets come into that region there will be no more friction against the nebula, and the migration will therefore stop. The density of exoplanets on this boundary where we should expect their inward migration to stop is  $\sim 10$  times higher than further out toward Mercury-like orbits. The fact that we see markedly fewer planets even further in (where they should be even easier to detect) than this limit, shows that the pile up at  $\approx 0.04$  AU is not an observational bias, but a real physical phenomenon.

### 1.5.5 The extremely close-in planets

It is not obvious why there are planets further in than the pile-up distance, but they may have migrated in there because of some other migration mechanism than friction against the nebula gas, or they may belong to a type of stars that were unable to clean up an inner void in the nebula. In any case they are rare, as is seen in Fig. 1.33.

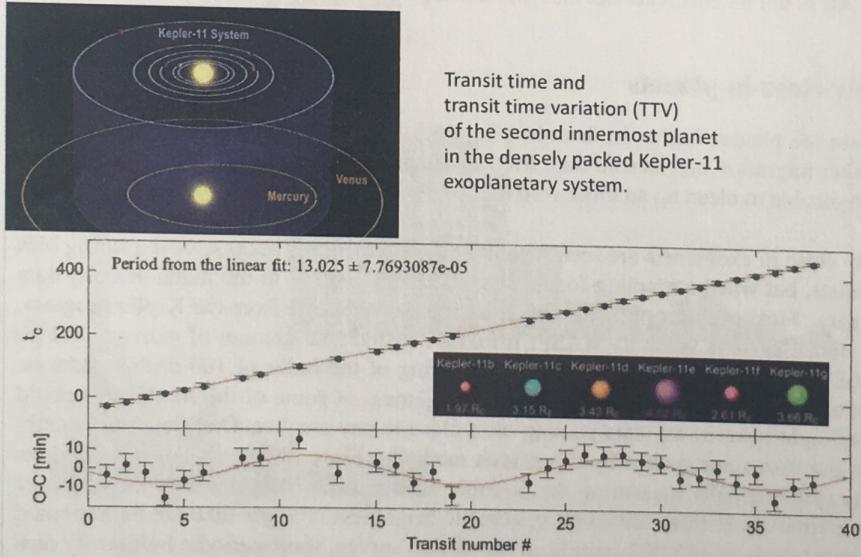
Many more extremely close in exoplanets are seen with the transit technique, because of the strong bias of transits to close in planets, but when correcting for the bias the same trend as in the radial velocity data is obvious in the transit data. Most of the confirmed very close in exoplanets are from the Kepler program, but also the other large transit surveys (such as WASP) reveal a considerable amount of extremely close in exoplanets. The OGLE microlensing survey has been monitoring of the order of 100 million stars on a nightly basis over several years to look for variations in the brightness of some of the stars which could be attributed to the line-of-sight toward two stars passing ultra-close to one another. One can conveniently find planets like those in our own solar system by this search method (as described in detail in a previous section), but since one is systematically measuring the possible change in the brightness of the observed stars, one will at the same time see if there are small systematic brightness changes that can be attributed to transiting planets. In an analysis of 155,000 stars from the OGLE survey, short-periodic luminosity dips were identified for 137 objects. Doppler investigations were able to confirm a planetary radial velocity signal in 5 of these, OGLE-TR-10,56,111,113,132. Since these stars are typically much further away than the stars on the radial velocity monitoring programs, only relatively low precision spectra can be obtained; in this case spectra that could monitor the radial velocity with an accuracy of 100 m/s. With such accuracy it is only possible to identify large exoplanets in extremely small orbits. Some of the most close-in planets were therefore identified from the OGLE transit survey and from the second HST survey (described above) which

each analyzed the light curves from more than 100,000 individual stars.



**Figure 1.34.** Artist impression of the Sun-like ( $G0$ ,  $M=1.35M_{\odot}$ ,  $R=1.57R_{\odot}$ ,  $T_{\text{eff}}=6300\text{K}$ ,  $[\text{Fe}/\text{H}]=0.3$ ) star WASP12 with its close in ( $r=0.023\text{AU}$ ,  $e=0.049$ ,  $P=1.09\text{days}$ ) giant planet WASP12b ( $M=1.39M_{\text{jup}}$ ,  $R=1.83R_{\text{jup}}$ ,  $\rho=0.326\text{ g/cm}^3$ ,  $T=2500\text{K}$ ) slowly being disrupted by tidal forces and “eaten up” by the star.

Other ground-based programs are more focused toward transiting exoplanets than the OGLE program, and the existence of several exotic planets have been captured by these large-scale surveys. One of the most extreme is the recently discovered 1.1 day period transiting exoplanet WASP-12b (see Fig. 1.34) which orbits well inside the pile-up distance, and must have got there recently. It is so close to its Solar-like star that tidal forces make it loose its energy and mass to its star at a rate of several million billion tons per second. This is so rapid that it will last only of the order of 10 million years before it is completely swallowed by its star. Monitoring the transit time with accuracies on the few-seconds level (as in Fig. 1.11) makes it possible to watch in real time how the orbital period gets systematically shorter and shorter while the planet speeds up as it gets closer and closer to its final disruption.



**Figure 1.35.** Upper left panel: Artist impression of the Sun-like ( $G6V$ ,  $M=0.95M_{\odot}$ ,  $R=1.1R_{\odot}$ ,  $T_{\text{eff}}=5700\text{K}$ ,  $[\text{Fe}/\text{H}]=0$ , age=8Gyr) star Kepler-11 with its 6 close in ( $r=0.09$  to  $0.46$  AU) super-Earth ( $M=2$  to  $14 M_{\oplus}$ ) exoplanets, compared to the orbits of Mercury and Venus in our own solar system. Middle panel shows the transit time of Kepler-11c as function of transit number, while the lower panel show the transit time variation (TTV) compared to the computed mean transit time, again as function of running transit number. The sinusoidal too late and too early arrival of the transit compared to the average, is due to the periodic pull on Kepler 11c by the other planets in the system.

Monitoring the variation in the transit time allow us to model how the orbit change over time. This can be gradually and in one direction (as for Wasp-12b where the orbital time continuously get systematically shorter due to the spiraling in of the planet), but it can also be periodic toward longer and shorter periods. By studying such transit timing variations (TTV), one can monitor the gravitational interaction by additional known or unknown exoplanets in the system. Four new exoplanets have been discovered (all from the Kepler data set) solely from the TTV effect they have on other planets in the system. If the timing accuracy gets fine enough, one would eventually be able to infer the existence of orbiting large moons, since the center of mass of the planet and its moon would orbit the star with a constant speed, allowing the planet (and hence the transit) to be periodically a small amount respectively before and after the passage of the centre of mass.

Fig. 1.35 shows the transit time as function of time for the compact Kepler 11 system of the second to innermost planet (Kepler-11c) in the system, together with the sinusoidal variation in the exact time of the midpoint of the transit. The TTV signal of Kepler-11c is a consequence of the gravitational influence of the 4 other planets in the system (mainly the nearby larger Kepler-11d) and it allows us to compute the mass of all the 5 planets in the system based on the transit data alone (without the usually required mass determination from radial velocity measurements).

It is now custom to call all large exoplanets in orbits with period shorter than 3 days, such as WASP-12b, for very-hot-jupiters. They account for only a few % of all known exoplanets. Those with periods between 3 and 10 days are called hot-jupiters. Remark that it has already become a custom mis-description to claim that the radial velocity technique mainly identifies hot-jupiters, because this was what the technique revealed during the first few years. Hot-jupiters in the definition given here, accounts for only  $\sim 10\%$  of the known population. The known giant exoplanets with periods larger than 10 days have no specific name, but could conveniently be called Jupiter-like planets in terrestrial-like orbits. They account for about 3/4 of all known exoplanets discovered with the radial velocity technique.

### 1.5.6 The standard planetary system in our Galaxy

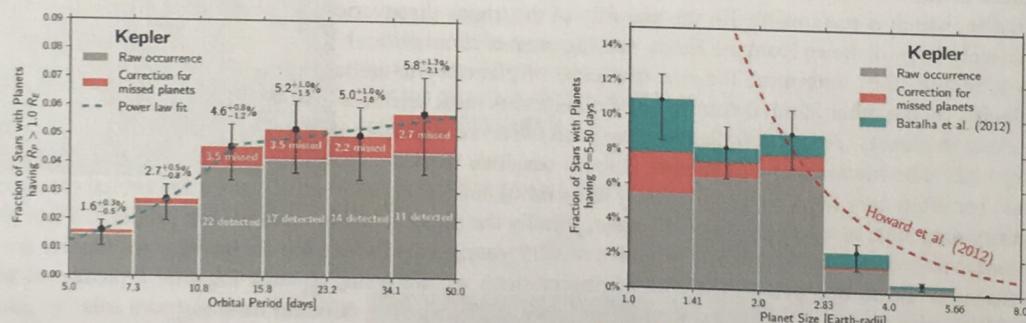
The two most successful methods in terms of number of detected exoplanets (the radial velocity and the transit methods) are both highly biased toward discovering planets in small orbits and/or of large size. The transit method is the one mostly biased toward small orbits, and due to the selection criteria of the Kepler satellite, which is responsible for the majority of the transit discoveries, the observed transit host stars are at relatively large distance from the Earth, making most of them difficult to follow up by radial velocity studies. As a default, they determine the size (the ratio of planetary to stellar radii) and not the mass. The radial velocity, on the other hand, determines the mass and is most sensitive to giant planets, with some preference to close in planets. About 200 planets have been observed with both techniques, giving rise to the important Fig. 1.12. The microlensing technique is more sensitive to planets in larger orbits. It reveals the planetary mass (or often only the stellar to planetary mass ratio) and the instantaneously projected orbital distance, and mainly map stars at relatively large distance. Finally the direct imaging method map primarily giant planets at larger distances, but it is mainly sensitive to very young stars (where the contracting gas planets are most luminous). From this patchy material of observations we are trying to stick together a picture of what a standard solar system in our Galaxy looks like. Is it similar or very different from our own solar system, and if so then why?

The radial velocity method is able to trace planets of smaller masses if they are in small orbits, where many evolutions can be measured within available observing time. If we therefore disregard the small planets we get a good picture of the abundance and radial distribution of giant planets. We already saw in Fig. 1.33 that the distribution of planets with mass larger than Saturn is relatively flat in the region 0.05 to 3 AU, drops markedly off outside 3 AU, has a peak around 0.04 AU, drops off inside this, and maybe has a peak around 1 AU. The drop outside 3 AU is somewhat biased due to lack of long term homogeneous surveys, so one would need to correct for this bias. The estimates range from believing that there in reality are as many giant planets outside 3 AU as inside, to a relatively sharp drop in the real abundance. The pile up around 1 AU is by some interpreted as an indication of the existence of two distinct populations of planets inside and outside 1 AU, but one might also speculate whether the peak could have a more psychological reason than a physical one, as for the metallicity distribution among transit host stars shown in Fig. 1.30. As a working

hypothesis we could assume that the giant planets have a flat distribution as function of radial distance from 0.03 to 3 AU (apart from the peak around 0.04 AU) and a linearly declining distribution from 3 to 30 AU, with half as many planets outside 3 AU as inside. Figure 1.32 indicate that  $\sim 6\%$  of solar type stars in the solar neighbourhood have a giant planet of the mass of Saturn or larger, within 3 AU. If these number are more or less correct, then a standard solar system have  $\sim 0.1$  gas giants, distributed as 0.03 within 1.5 AU, 0.03 between 1.5 and 3 AU, and 0.03 outside 3 AU (with 10 times more of these close to 3 AU than close to 30 AU).

With a strong development during recent years in the accuracy of existing spectrographs, the radial velocity method is now able to discover planets in small orbits with masses as small as  $\sim 1 M_{\oplus}$ . In a systematic monitoring of 166 G and K dwarfs with the Keck spectrograph over a 5 years period, it was found that for  $a < 0.25$  AU (i.e. well inside Mercury's orbit) the planetary mass distribution could be expressed as  $dN/dM = M^{-1.48}$ , with an average of 0.2 planets detected per star, which after correcting for bias was translated into 0.5 planets per star in the mass interval  $1 M_{\oplus}$  to  $13 M_{Jup}$ . Such a mass distribution corresponds to approximately 0.2 planets per star in the mass interval  $1-3 M_{\oplus}$ , 0.2 from  $3-30 M_{\oplus}$  and 0.1 from  $30 M_{\oplus}-13 M_{Jup}$ , or a ratio of 10 between "small and large" planets (for example the ratio  $M_{Mars-Neptune}/M_{Neptune-BD}$  or  $M_{Earth-Neptune}/M_{Saturn-BD}$ ), *BD* referring to the brown dwarf limit of  $13 M_{Jup}$ . Analysis of the larger HARPS survey found slightly higher numbers of small planets, but within the uncertainty the two surveys were in agreement with one another.

With more than 3000 exoplanet candidates, mainly in very small orbits, the Kepler data base has an even better statistical coverage of the size distribution of planets in small orbits than the radial velocity surveys. The data base is now public and very valuable for statistical studies because of its high homogeneous photometric accuracy, long continuous time series, well calibrated correction for sensitivity difference toward large and small planets, and huge amount of planets. Figure 1.36 show the results of a recent study that systematically picked the 1200 Kepler candidates of lowest noise level such that it was possible to extract a large and statistically significant number of planets with radii as low as  $1 R_{\oplus}$  (and even several of  $R$  below  $1 R_{\oplus}$ ) in small orbits. Figure 1.36 show the number of planets with radii between 1 and 8  $R_{\oplus}$  orbiting solar-like stars with period  $P$  in the interval 5 to 50 days (i.e. semi major axis  $a < 0.25$  AU). One notice the almost flat distribution in  $\log(P)$  for period larger than 10 days (left panel), and the plateau in number of planets as function of size for planets smaller than  $2.8 R_{\oplus}$  (right panel).



**Figure 1.36.** The number of exoplanets in the Kepler sample with orbits smaller than 0.25 AU, shown as function of orbital semi major axis (left panel) and as function of planetary radius (right panel). Grey columns are planets reported by the work of Petigura et al (2013), red the planets added to the detected ones to compensate for incompleteness in the study, and blue additional planets from the study of Batalha et al (2012), while red dotted line is the analytical distribution suggested in the work of Howard et al (2012).

In a study from 2012 (shown as dotted line in Figure 1.36) it was concluded that there are  $\sim 0.2$  planets with  $R > 2R_{\oplus}$  per star in the interval  $P < 50$  days. The number of identified planets as function of the planetary radius  $R$  were approximated by a power law  $dN/d(\log(R)) = 2.9R^{-1.92}$ , which, if extrapolated downward to  $1 M_{\oplus}$ , predicted that each solar-like star in the Kepler sample would be orbited by 0.7 planets, and extrapolated down to Mars-size would correspond to 2.5 planets per star (with  $P < 50$  days), and a ratio between "small and large" planets (as defined above) to be close to 50 (rather than 10 as derived

above from the radial velocity measurements). One of the new discoveries in the work from 2013 (shown as histograms in Fig. 1.36) was the quite unexpected identification of a plateau in the number of exoplanets below  $\sim 2.8 R_{\oplus}$ . Including the plateau from 1 to  $2.8 M_{\oplus}$  would result in a small-to-large planet ratio of  $\sim 25$ , somewhat mid-way in-between the RV indication of 10 and the earlier indication from Kepler of  $\sim 50$ .

The radial distribution in Fig. 1.36 can be approximated by a relatively flat distribution in  $\log(a)$ , where  $a$  is the semi major axis. However, Fig. 1.33 indicates that at least for the giant planets, the distribution grows faster with  $a$  than indicated by the constant number in  $\log(a)$  spacing, quickly approaching a flat distribution in  $a$  instead, from inside Mercury-like orbits to  $\sim 3$  AU, after which it again grows less steep with distance.

Using the numbers from Fig. 1.32 and assuming 50% more giant planets in total outside 3 AU compared to the numbers in Fig. 1.32 (as discussed above), leave us with a total of approximately one giant planet (with mass as Saturn or larger) per 10 solar-like stars. The distribution in Fig. 1.33 then tell us that this one giant planet per 10 stars is distributed with 3 planets outside 3 AU per 100 stars, 3 in the region 1.5 - 3 AU, and 3 inside 1.5 AU, of which one will be "hot" (i.e. within 0.25 AU). The standard solar system will therefore have no giant planets, and the few that do have giant planets will most often have them well inside the frost line.

If the ratio of small-to-large sized planets is a constant ratio independent of distance, and there are  $\sim 25$  times more Earth-Neptune like exoplanets than Jupiter-Saturn like exoplanets (as indicated by radial velocity and transit detections in the inner 0.25 AU of the exoplanetary systems, as discussed above) then "a standard solar system" has  $0.1 \times 25 = 2.5$  small planets distributed with  $\sim 0.5$  inside Mercury's orbit, 1.5 in the terrestrial region, and 0.5 in the region corresponding to where the 4 giant planets are in our solar system (although most of these a bit inside Jupiter's orbit).

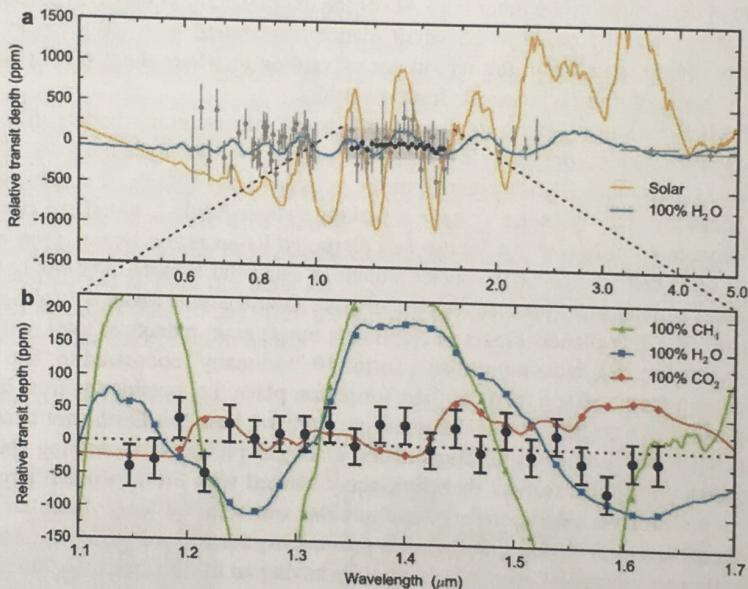
This distribution of "Earth to super-Earth sized" planets is not widely different from our solar system, although spread a bit more (both inward and outward). Neither is the mass distribution of the small planets in our solar system markedly different from the standard solar system. It is sometimes discussed as a surprise in the literature that our solar system is lacking "super-Earths", but if we consider that Mars most likely would have had  $\sim 5 M_{\oplus}$  if not Jupiter had disrupted its accretion region, then we have a quite normal both size and orbit distribution of the rocky planets in our solar system. The really unusual feature of our solar system is the many giant planets. Statistically we should expect about 0.1 (of Saturn-mass and above) to 0.2 (if considering also planets closer to Neptune's mass) giant planets in total, with at least 2/3 of them in the terrestrial region. We have more than a factor 10 "too many" compared to "the standard planetary system", and at least a factor 50 too many at their formation place, i.e. outside the frost line. From an astrobiological point of view this ought to raise the question: "are we here (on Earth) not because the Earth is something special, but because the surroundings of the giant gas planets is something special?" – if this was not the case I guess we should suspect that the Galaxy teamed with life to a much larger degree than immediately seems to be the case. We therefore conclude that our solar system is peculiar by having 4 giants in large orbits (approximately where giant planets presumably form), and by having exceptionally many planets (8 orbits) compared to an average of only 2.5), but not by having an Earth-sized planet in an Earth-sized orbit. We may expect from the statistics of exoplanetary distributions that there are many exoplanets in the habitable zone, but few habitated.

## 1.6 Modelling the atmosphere – understanding the spectra

At the time of writing this chapter it is the general expectation that we will not be able to resolve in direct imaging an earthlike exoplanet from a sunlike star before the 39 m large E-ELT telescope will be operational, in the mid-2020's, with its full capacity of adaptive optics and basic instruments. This gigantic telescope, far larger than any existing telescope today, is designed such that it can zoom in on an earthlike planet in the habitable zone around nearby stars. When this fantastic facility becomes available, we will be able to obtain a spectrum of the nearest Earth-like exoplanets in the habitable zone around solar type stars, the same way as we today obtain spectra of Mars and the other planets in our solar system. It will tell us whether there are living organisms modifying the chemical composition of their planet's atmosphere, just as the biology on Earth does today in large scale. Until then, however, we are not without spectra of exoplanets, and exciting new results are rolling in almost every month. It is indeed already possibly to obtain direct spectra now, as

will be done in the 2020's by the E-ELT, but today only of very large gas planets orbiting in large orbits around very young stars. Such planets shine relatively bright due to the energy from contraction, and if they are in large enough orbits around small enough stars, they look like a component in a binary star. The bulk of the exoplanet spectra we have today are, however, obtained by taking a spectrum when we see the planet and the star together, and then subtracting a similar spectrum for situations when we see the star alone. This can obviously only be done for transiting planets, where we can obtain the stellar spectrum alone when the planet hides behind the star, and it obviously requires very high accuracy and excellent spectra, spectrophotometry or photometry.

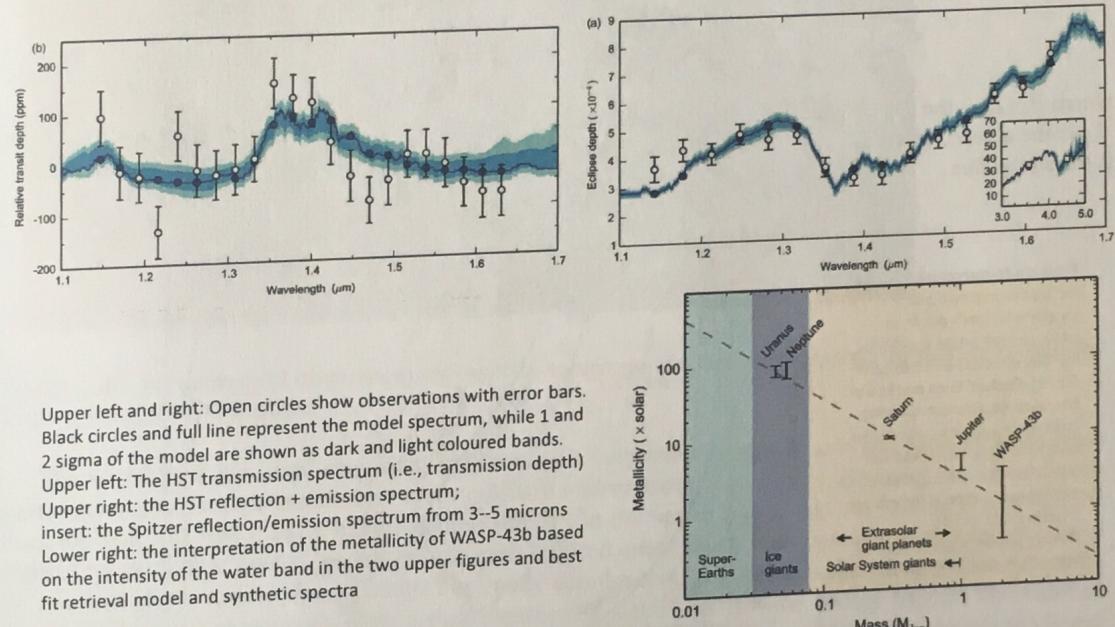
There are basically two types of combined spectra than can be obtained; a transmission spectrum when the planet passes in front of the star, and a combined reflection plus thermal emission spectrum observable just before the planet enters the occultation phase. The transmission spectrum can be obtained either as a changing radius (i.e., how much stellar light is blocked) as function of wavelength, or as a regular spectrum manifested as spectral lines Doppler shifted relative to the stellar spectral lines. One can in principle obtain even more information about the distribution of the planetary atmosphere as function of longitude by carefully observing the development of the combined stellar and planetary spectrum along with the predicted changes in phase of the planet while it orbits the star.



**Figure 1.37.** One of the longest exposure spectra ever taken with HST – maybe even the longest – was the 60 full orbits exposure spectrum of GJ1214b. This is a super-Earth exoplanet orbiting a late M-type dwarf star, only 3 solar radii above its surface, with an orbital period of only 1.6 days. The gray dots in the upper panel are previous observations, while the high-accuracy black dots are the HST observations. The yellow line is a normalized solar spectrum (transformed into relative transmission depth, i.e. how much a clear solar gas composition planet would absorb the starlight as function of wavelength). The blue, green, and brown "spectra" (relative transit depth as function of wavelength) show models of what the measurements would have looked like if the planetary atmosphere was made of 100% methane, water, and carbon dioxide, respectively. It is concluded that the HST spectrum cannot be explained with any clear gas atmosphere, but probably represent a thick layer of clouds in the upper atmosphere of the planet.

An impressive spectrum of the 7 Earth-mass exoplanet GJ1214b orbiting an M4.5 dwarf star only 13 pc away, was obtained in 2010 from 60 full cycles of HST observations where the transit transmission spectrum was subtracted the stellar spectrum, to obtain the planetary spectrum. The extremely well exposed spectrum made it possible to conclude that one sees absolutely nothing in the spectrum, most likely meaning that the planet is covered in a thick gray (i.e. featureless) cloud layer.

Fig. 1.37 was one of the first and most extensive attempts to obtain a spectrum of an exoplanet atmosphere, an ironically enough a pick of a planet with a thick cloud cover like Venus, with no spectral features to reveal. Many of the later transit spectra of other exoplanets ended with the same negative results, but some of them do have a more clear atmosphere, where more or less clear molecular bands of H<sub>2</sub>O, CO, CO<sub>2</sub>, CH<sub>4</sub>, and TiO might have been detected, as well as strongly broadened atomic lines of Na and K, and raleigh scattering from dust. One of the clearest identification is the 1.4 μm band of water seen in the spectrum of WASP-43b shown in Fig. 1.38. Left panel in the figure shows the transmision spectrum (just as for GJ1214b in Fig. 1.37), while the right panel shows the corresponding occultation spectrum (i.e., the combined stellar and planetary spectrum just before occultation subtracted by the stellar spectrum during the occultation). In both cases we clearly see the 1.4 μm band of H<sub>2</sub>O.



**Figure 1.38.** The HST WFC3 spectrum of WASP-43b in the 1.1 – 1.8 μm region, showing the strong 1.4 μm band of water, plus the Spitzer spectrum from 3 – 5 μm, and the extracted metallicity estimate.

In planetary atmosphere theory (as opposed to stellar atmosphere theory) it is the tradition to operate with the concept of scale height,  $H$ . This is the height one would have to go upward in the atmosphere in order to see the gas pressure drop with a factor  $e$ , if the temperature  $T$  in the atmosphere was constant. Under this assumption

$$H = \frac{k_B T}{\mu_m g} = \frac{RT}{g} \quad (1.38)$$

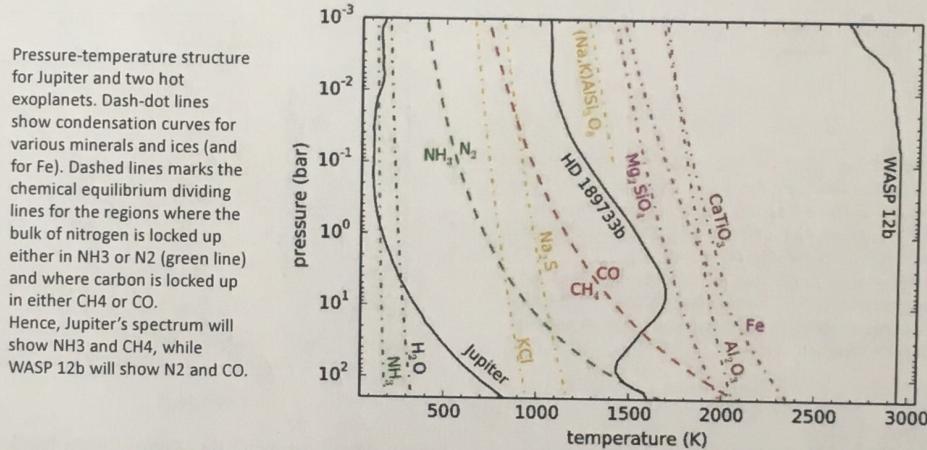
where  $k_B$  is Boltzman's constant,  $\mu_m$  is the mean molecular mass of the atmospheric gas (assumed to be constant),  $R$  is the molar gas constant, and  $g$  is the planetary surface gravitational acceleration. The scale height of the Earth's atmosphere is ~8.5 km and Jupiter's atmosphere is ~27 km. For a strong absorption line, the same optical depth is typically reached about one scale height further out in the atmosphere than for the surrounding spectral continuum. This means that if one is determining the planetary radius from observing the light curve in a continuum point and in a nearby strong line, then the derived planetary radius typically should differ by approximately a scale height. If one assumes  $T$  to be the effective temperature, one would have gained some information about the amount of atmosphere the planet has. If one is able to repeat the exercise for many spectral lines, one would have learned something about the amount of gas of the species that gave rise to the individual lines, and hence some crude information about the composition of the atmosphere. In praxis the ratio between the scale height and the planetary radius is of the order of 10<sup>-4</sup>,

and one should be able to solve for the planetary radius as function of wavelength with at least this accuracy in order to extract some information. Most often one would determine the light curve not from an observed spectrum, but through one or more broadband filters, and the signal is then much weaker still, because each filter samples both the continuum and a number of different spectral lines.

The occultation signal is usually much weaker than the transit signal, but if it can be measured accurately it gives complementary information to the transit signal. Right before the occultation, the phase of the planet is seen from us as (almost) full, and the flux density,  $f_{occ}$  from the planet is therefore a combination of the thermal emission,  $f_{occ,thermal}$ , from the day-side of the planet (modulated by the overlaying planetary atmosphere) plus the reflected starlight,  $f_{occ,reflected}$ ,

$$\Rightarrow \frac{f_{occ}}{f_*} \approx \frac{R_p^2 B_\lambda(T_p)}{R_*^2 B_\lambda(T_*)} + A_\lambda \frac{R_p^2}{4a^2} \approx \frac{R_p^2 T_p}{R_*^2 T_*} + A_\lambda \frac{R_p^2}{4a^2} \quad (1.39)$$

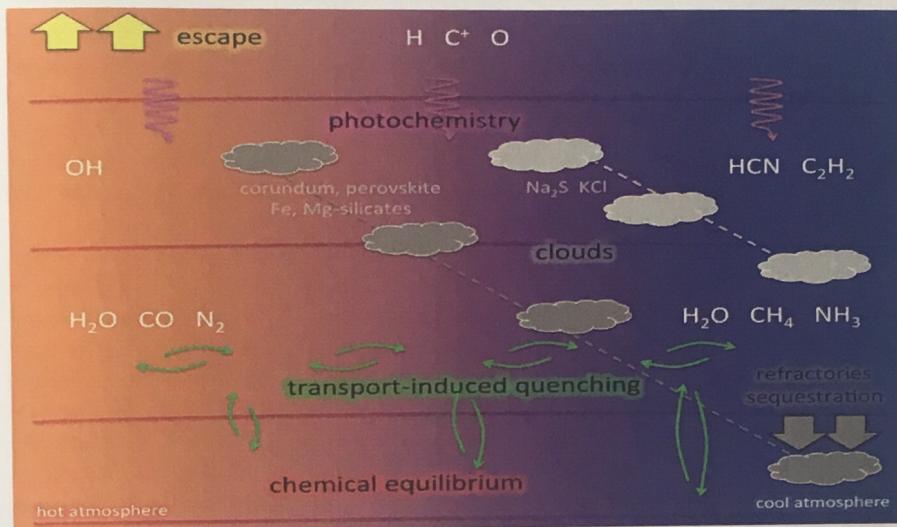
where  $B_\lambda(T)$  is the Planck function (which in the long wavelength Rayleigh-Jeans approximation is  $2k_B T / \lambda^2$ ),  $A_\lambda$  is the albedo, and  $a$  is the orbital semi-major axis,  $R_p$  and  $R_*$  are the planetary and stellar radii, and  $f_*$  is the stellar flux.



**Figure 1.39.** Model structure and chemical equilibrium results for Jupiter and two exoplanets, assuming solar metallicity.

If the elemental abundances and the temperature-pressure structure are known, then one can compute a synthetic spectrum and compare it with observations. In analysis of stellar spectra, it is possible to compute the temperature-pressure structure ("the model atmosphere") self consistently by varying the fundamental stellar parameters, and a best fit to the observed spectrum would usually be the iteratively fitted fundamental parameters ( $T_{eff}$ ,  $logg$ , elemental abundances) that fit the spectrum best. In this way one would claim that the best fit to the spectrum would have determined the stellar fundamental parameters plus the abundances of the observed species. The exoplanetary atmospheres are still not sufficiently understood to allow such a fundamental approach, mainly because of more dynamically active processes than in typical stellar atmospheres. It is therefore normal to fit the spectrum by a model with very many free fitting parameters, often including the pressure-temperature structure and the partial pressures. Such a fit is in some sense not really a determination of the atmospheric structure and composition, but a fit of many parameters that under the given assumptions will lead to a spectrum resembling the observed one. One usually call this process the spectral retrieval. If the retrieval looks like the observed spectrum over a large spectral range, one would have confidence that at least qualitatively we have gained knowledge about the atmospheric structure and composition. Often, however, the retrieval would have several solutions, and in the long term one should aim at developing fully self-consistent models with no free parameters, as those that are used for stellar

atmospheres. For example, it has been discussed in the literature whether observed weak H<sub>2</sub>O bands should be interpreted as depleted metallicity or due to cloud layers that block some of the water absorption. Fig 1.40 illustrates some of the complexity.



**Figure 1.40.** Ingrediences of exoplanetary atmospheres as function of effective atmospheric temperature (x-axis) and height (y-axis) in the atmosphere.

The very hottest gas giant exoplanets may resemble very much brown dwarfs and cool main sequence stars, with daytime temperatures close to the effective temperatures of cool stars. However, strong irradiation from the host star will usually induce photochemistry in the upper layers, and the middle layers may be affected by vertical as well as horizontal transport-induced quenching – a process not considered as part of the standard stellar atmosphere solution. For cooler exoplanetary atmospheres the situation becomes even more complex, and clouds will often play a completely central role for the energy balance and radiated spectrum, just as is the case for the Earth and in much larger extend for Venus and the giant gasplanets in our solar system and for the moon Titan.

## 1.7 Are anybody like us living on exoplanets like our's ?

There are approximately  $10^{11}$  stars in our Galaxy, and more than  $10^{11}$  galaxies in the visible Universe. A common layman conclusion is that because of this overwhelmingly large number there is almost certainly many places in the Universe with life, and probably many places with life much more intelligent than us. This is qualitatively the same argument that lead Anaximander (c. 610 – c. 546 BC), Democritus (c. 460 – c. 370 BC), Giordano Bruno (1548 – 1600), Christiaan Hyugens (1629 – 1695), William Herschel (1738 – 1822) and many others believe that there must be many planets like the Earth and many intelligent beings like us in the universe. However, the problem with this argument is that we do not know the probability,  $P_{life}$ , of life originating in the vicinity of any given star (nor the probability of life becoming intelligent). Suppose  $N_* = 10^{22}$  is the number of stars in the Universe, and that the probability of life to originate on a planet around a star is  $P_{life} = 10^{-22}$ . Then the most likely estimate for the number of planets with life in the Universe is 1. We have absolutely no clue to a realistic estimate of  $P_{life}$  other than that it is above zero and less or equal to one. Even the possible first identification of a radio signal from an intelligent civilization on a planet around another star, would only tell us that there has been two civilizations in the Universe, and identification of possible life forms on Mars would only tell us that life exist two places in the Universe.

In order to get some understanding of whether life is a common or an unusual phenomenon in the Universe, we need to split the problem into some sub-problems we can figure out ways to work at in a scientific

way. Frank Drake formulated in 1961 his famous equation to do exactly this,

$$N = R_* \cdot f_p \cdot n_e \cdot f_l \cdot f_i \cdot f_c \cdot L \quad (1.40)$$

$N$  is the number of civilizations in our Galaxy whom we in principle can communicate with.  $R_*$  is the number of stars born in the Galaxy per time unit,  $f_p$  is the fraction of these which will have planetary systems,  $n_e$  is the number of Earth-like planets (planets suitable for life) per planetary system,  $f_l$  is the fraction of Earth-like planets where life actually originates (i.e., the probability that life will develop once the physical conditions are as on Earth),  $f_i$  is the probability of life to develop into intelligent beings once it has originated,  $f_c$  is the probability of intelligent life to develop equipment which allows interstellar communication, and finally  $L$  is the life-time of such communicating civilizations.

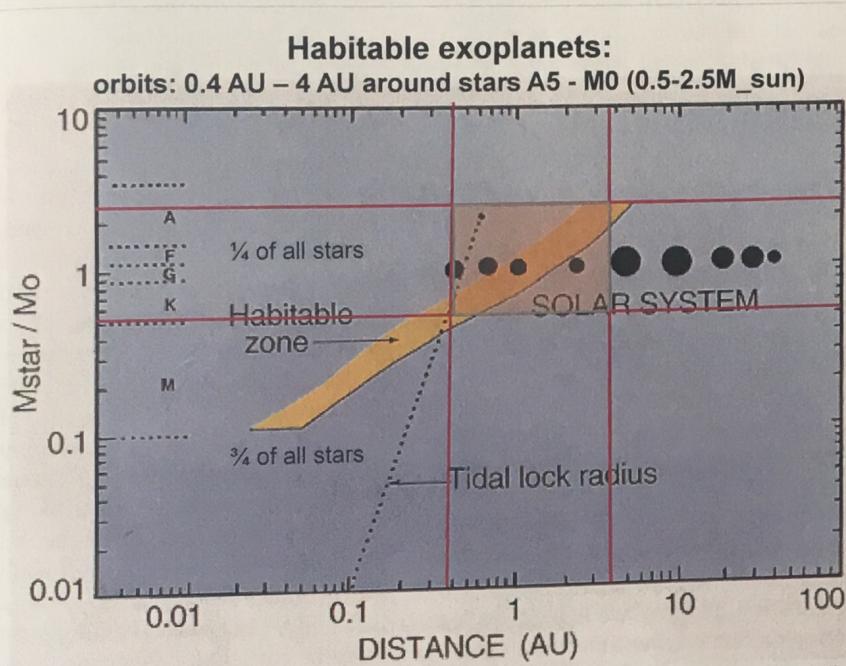
The difficulty is obviously not in solving the equation, but in estimating the factors. Only  $R_*$  is well known. If  $10^{11}$  stars have originated during  $10^{10}$  years (almost all stars ever born in our Galaxy still exist as stars), obviously on average 10 new stars were born per year in our Galaxy. We have discussed above how our knowledge is rapidly increasing about the fraction of stars that form planetary systems ( $f_p$ ). We have been able to formulate extrapolation formulas that with some uncertainty can give us an estimate of the exoplanet mass distribution function, and hence the number  $n_e$  of Earth-like exoplanets, and telescopes and instruments already under development holds the potential of giving us information about how many of the Earth-like planets harbours life ( $f_l$ ). These three factors are somehow what I would tend to call the core of *astrobiology*. The remaining three factors in Drakes equation are still mainly part of other disciplines (medicine, computer sciences, engineering, sociology, and maybe even environmental studies, military studies, and psychology), although a possible break through in SETI (the Search for ExtraTerrestrial Intelligence) obviously would change this picture completely.

## 1.8 Habitability of exoplanets

The ultimate challenge of astrobiology, and its most central question to answer, is undoubtedly how human beings relate to the development of the universe; are we as an intelligent species, that seems to understand the basic principles of how nature and we ourselves came about, alone in the entire universe, or are there myriads similar to us, who is just waiting to be contacted? The questions we need to solve concerning exoplanets to approach this astrobiological goal, include: are there solar systems like our own elsewhere in the Galaxy?" and "how abundant are planets where life can exist?"". We don't know, of course, the answer to these two question, but we are also forced to admit that in reality we have a long and challenging way to proceed first, even to just being able to specify what we really mean by the questions. What are the physical requirements for life to exist, and how similar to our own solar system does an extrasolar system have to be in order to fulfill the conditions that led to life in our solar system? In this section we will develop some considerations about what a planet should fulfill in order that life can exist on it; i.e. in order that it is habitable, and then see which of the exoplanets we have discovered so far looks as the most promising candidates.

Several different opinions have been presented in the scientific literature, as to what should be meant by habitability. While some authors think of places where humans could live, most authors have converged toward defining the habitable zone (HZ) as the orbital region where a solid Earth-sized planet with an atmosphere like ours, an albedo like Earth's (and often a rotational period more or less as the Earth's as well), can sustain a liquid water-ocean at its surface. In its simplest approximation this region can be defined as the distance interval  $r_{\text{HZin}}$  to  $r_{\text{HZout}}$  from a star of luminosity  $L_*$  where the planetary surface temperature,  $T_{\text{surf}}$ , is between 0°C and 100°C.

While the surface temperature depends on details of the atmospheric composition (which could be strongly influenced by whether the planet is actually already habitated, such as the atmospheric composition of our own planet is), the equilibrium temperature,  $T_{\text{eq}}$ , is straightforward to calculate from requiring that the incoming energy,  $F_{\text{in}}$ , that is absorbed by the entire planet, should equal the outgoing energy,  $F_{\text{out}}$ . If we let the albedo,  $A$ , be the average percentage of the incoming radiation that is reflected by the planet, and  $R$  be the planetary radius, then the energy balance of a planet at a distance  $r$  from the star, can be expressed



**Figure 1.41.** The habitable zone (yellow band) as function of stellar mass and orbital distance from the star. Also shown are the planets in our own solar system, and the orbital radius inside which planets will be tidally locked to their star. The region inside which planets are in the habitable zone and not tidally locked to their host star, is marked with red hatching. 1/4 of all stars are within the corresponding stellar mass range, while for 3/4<sup>th</sup> of all stars the formally habitable zone is inside the tidal lock zone.

as

$$\begin{aligned} F_{\text{in}} &= F_{\text{out}} \Leftrightarrow \\ (1 - A) \frac{L_*}{4\pi r^2} \pi R^2 &= 4\pi R^2 \sigma T_{\text{eq}}^4 \Leftrightarrow \\ T_{\text{eq}} &= \sqrt[4]{\frac{(1 - A)L_*}{16\pi r^2 \sigma}} \end{aligned} \quad (1.41)$$

For the Earth, with  $L_* = L_\odot$  and  $A = 0.39$ , we get  $T_{\text{eq}} = 246$  K. One can therefore express  $r$  in astronomical units as

$$r[\text{AU}] = \sqrt{\frac{1 - A_{\text{pl}}}{1 - A_{\text{E}}}} \sqrt{\frac{L_*}{L_\odot}} \left( \frac{246}{T_{\text{eq}}^{pl}} \right)^2 \quad (1.42)$$

Had  $T_{\text{eq}}$  been the actual surface temperature (which would be the case on a planet with no atmosphere), then today's HZ in our solar system would just simply had been from  $(246/373)^2$  to  $(246/273)^2$  AU, or from  $r_{\text{HZin}} = 0.44$  AU to  $r_{\text{HZout}} = 0.81$  AU, with one planet only, namely Venus, being in the habitable zone. As discussed elsewhere, there are several indications that the solar luminosity at the time of the Earth's formation was around 70% of its present value. Inserting  $L_* = 0.7 L_\odot$  in Eq. 1.42, gives the early habitable zone in the solar system to be from 0.32 AU to 0.59 AU. We see that without atmospheres, none of the planets in our solar system would have been in the habitable zone throughout the lifetime of the solar system. At early times Mercury would have been habitable, and today only Venus would have had the right temperature to be habitable. It is the effect of the atmosphere to heat the planetary surface above  $T_{\text{eq}}$ , which has brought the Earth into the habitable temperature regime (and Venus out of it). With an average surface temperature of  $13^\circ\text{C}$ , the "greenhouse effect" of the Earth's atmosphere keeps the surface temperature  $\Delta T = 286 - 246 = 40^\circ\text{C}$  above its equilibrium temperature.

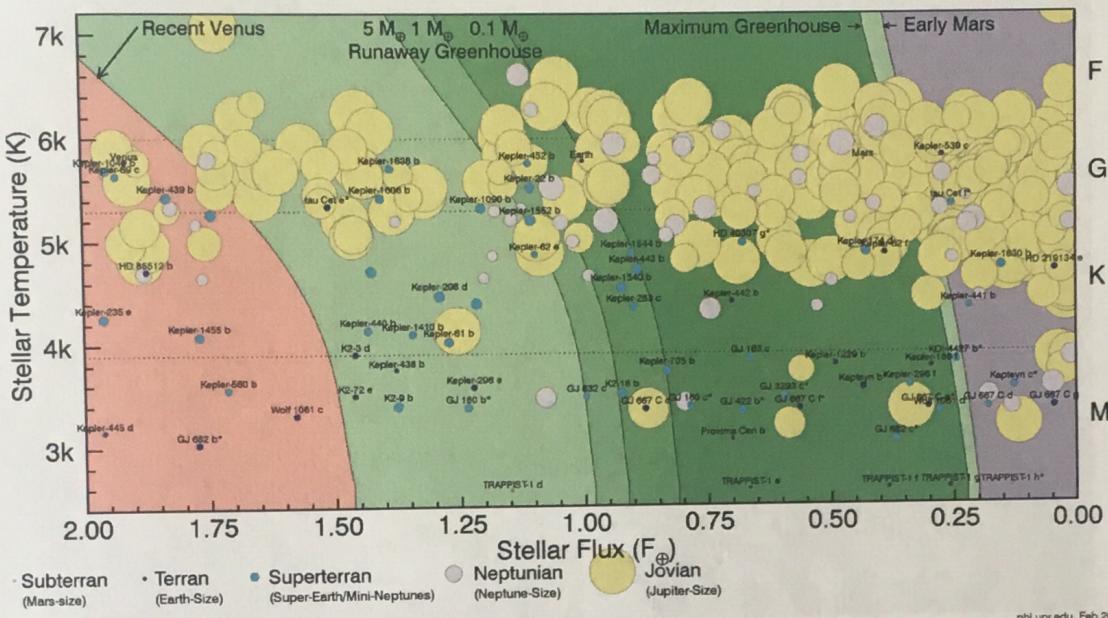
The numbers also reveal the well known faint-sun paradox, which stress the fact that not only is a greenhouse heating atmosphere necessary in order to keep present day Earth warm enough for life, but the greenhouse effect needs to have been much larger at early Earth, and developed in "just the right way" for life, in order to have kept the surface temperature of the Earth constant and "biofriendly" during all of its 4.5 Gyr long evolution.

It is seen that the actual existence of an atmosphere is crucial for where the HZ is, but at the same time its effect cannot be calculated (with any reasonable accuracy), since it depends sensitively on the atmospheric density and composition (and on the existence of life forms), which is far beyond our knowledge about how planetary atmospheres form and develop. Even for the Earth, it is largely unknown why the atmosphere has the composition it has today, and what it was at early times. One could make a reasonable rule of thumb about the amount of greenhouse heating for habitable exoplanets,  $\Delta T(t)$ , as function of time, by adopting the Earth's average temperature today ( $T_{\text{surf}} = 287 \text{ K}$ , implying  $\Delta T = 40 \text{ K}$  today), and requiring that  $T_{\text{surf}}$  should have been always approximately the same. If the solar luminosity 4.5 Gyr ago ( $t = 0$ ) was  $0.7L_{\odot}$ , then this implies from Eq. 1.41 that  $\Delta T(t = 0)$  was 60 K on the Earth soon after its formation. The process that gave rise to an initial greenhouse heating on Earth of  $\approx 60 \text{ K}$  and controlled its response to the increasing sunlight by gradually reducing it to 40 K today, was the process that was responsible for making the Earth continuously habitable. Adopting these numbers as a basis, would allow us to calculate the continuous habitable zone around any type of star. One could claim (or define) that a planet is biofriendly if it is continuously in the habitable zone throughout biological timescales (which may be a few hundred million years for complex evolution, and maybe a few billion years for development of intelligent life, based on our only one known example). Some authors reach the conclusion that solar type stars are the most likely stars to be orbited by biofriendly planets, because they have an optimal combination of a reasonably large habitable zone, and a relatively long and stable lifetime. Whether this is correct, obviously requires much more knowledge about life than we have at present, but it is certainly true that the habitable zone around very small stars is narrow (while the stellar lifetime is long) and that the habitable zone is broader for higher mass stars, but that the highest mass stars have a very short lifetime compared to what we consider biological timescales (based on our knowledge about life around one solar type star). Therefore there must necessarily be an optimum somewhere in-between the highest and the lowest mass stars – maybe it is  $1 M_{\odot}$ .

### 1.8.1 The potentially habitated known exoplanets

In the standard theory about the solar system formation presented in a previous chapter, planets form by accumulation of solid particles, until they reach a mass of about  $10 M_{\oplus}$ . If the surrounding gas nebula is still present at this stage, the planets will then enter a run-away phase where they quickly (i.e., within a Kelvin-Helmholz time scale) will accumulate gas enough to become gas planets of  $100 M_{\oplus}$  or more. We should therefore expect that any planet with a mass below  $10 M_{\oplus}$  is a solid rocky-metallic planet like the Earth and that any planet above  $100 M_{\oplus}$  is a gas planet like Jupiter or Saturn. We should expect only very few exoplanets with masses in-between  $10$  and  $100 M_{\oplus}$  because of the short time it takes to grow from  $10$  to  $100 M_{\oplus}$ . Some authors use the term "failed ice-giants" for the group of these predicted rare planets, and the term "super-Earths" for planets in the mass-range from a few  $M_{\oplus}$  up to around  $10 M_{\oplus}$ . In this definition, Uranus and Neptune are "failed ice-giants" in agreement with the fact that their ratio of ice-rock-metal to gas is quite high (compared to Jupiter and Saturn), indicating that they grew considerably slower than Jupiter (and Saturn), and failed to reach the necessary  $\sim 10 M_{\oplus}$  of solid material before most of the solar gas nebula was gone. There are no super-Earths in our solar system, but Mars could have been super-Earth sized if not Jupiter had formed early and perturbed many planetesimals from the Martian accretion region away.

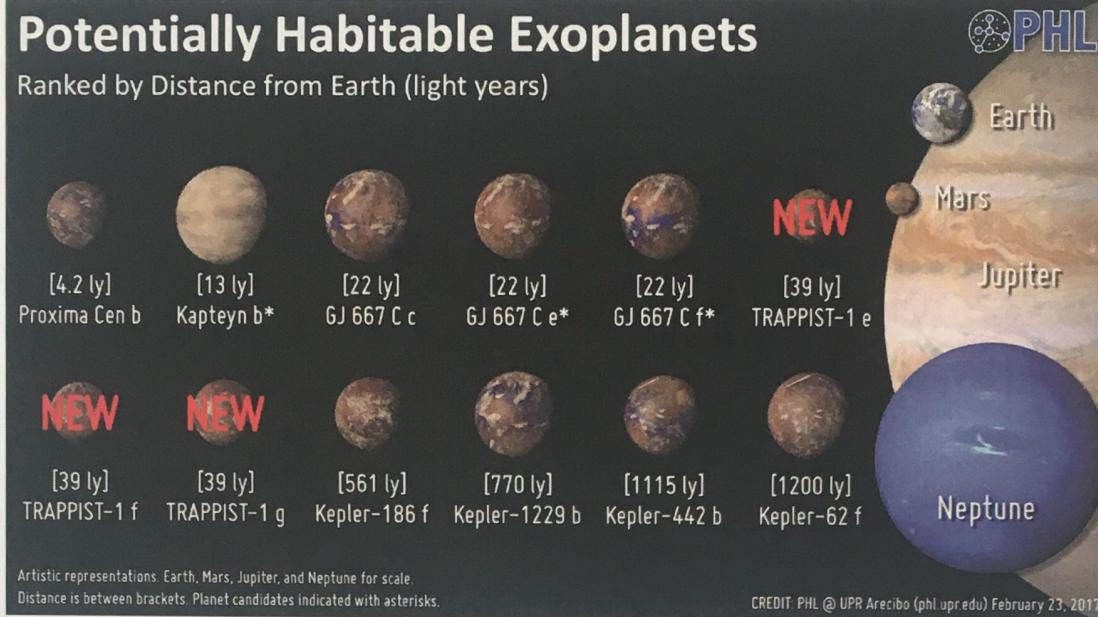
There is no indication of a strong division between the small rocky planets and the large gas planet population in the exoplanet data catalogs, at least for the small orbits where it has so far been possible to identify a statistically significant sample of small planets. However we will still tend to believe that planets have to not only be in the habitable zone to be habitated, but also be sufficiently small to be likely to have a solid surface where liquid water can exist in the form of lakes, rivers, and/or oceans. Fig 1.42 illustrate all the identified exoplanets known to be in or near to the habitable zone, and Fig. 1.43 highlights the smallest (and therefore maybe potentially most habitable) of them.



**Figure 1.42.** The known exoplanets estimated to be within or close to the habitable zone. Most of these are giant jupiter-like planets (large yellow circles) because they are easiest to detect, but few are Neptune-sized (light blue) and even Earth-sized ("Terran"; dark blue).

Independent of whether such a natural division between small potentially habitable-sized planets (say Mars-Neptune) and large un-habitable planets (say Saturn-masses and upward) exist, we can of course for simplicity classify our planets in those two groups, and we have argued above that there are up to 50 times more exoplanets in the former group than in the latter. This would be good news for the possibility of finding habitable exoplanets. M-dwarfs would be the most likely place to find them with the present technique, because for the radial velocity method as well as the transit method, low mass planets would be most easy to identify around M-dwarf star, which are the smallest and lowest-mass stars that exist ( $0.08M_{\odot} < M_* < 0.6M_{\odot}$ ,  $2100K(M9) < T_{\text{eff}} < 3800K(M0)$ ,  $0.1R_{\odot} < R_* < 0.5R_{\odot}$ ,  $7.5 < M_V < 20.0$ ). These stars are also the most abundant type of stars in our Galaxy, which makes them the most common lens stars. For the radial velocity method they are not as abundant in the samples as one might had expect, which is because their low luminosity makes it possible to observe them in sufficiently high spectroscopic resolution only in a relatively small volume around the Sun. It may therefore be around stars with lower mass than the Sun (K-dwarfs, and maybe M-dwarfs) that we may identify the first truly habitable exoplanets, and perhaps the first life, and maybe such planets are already in our samples.

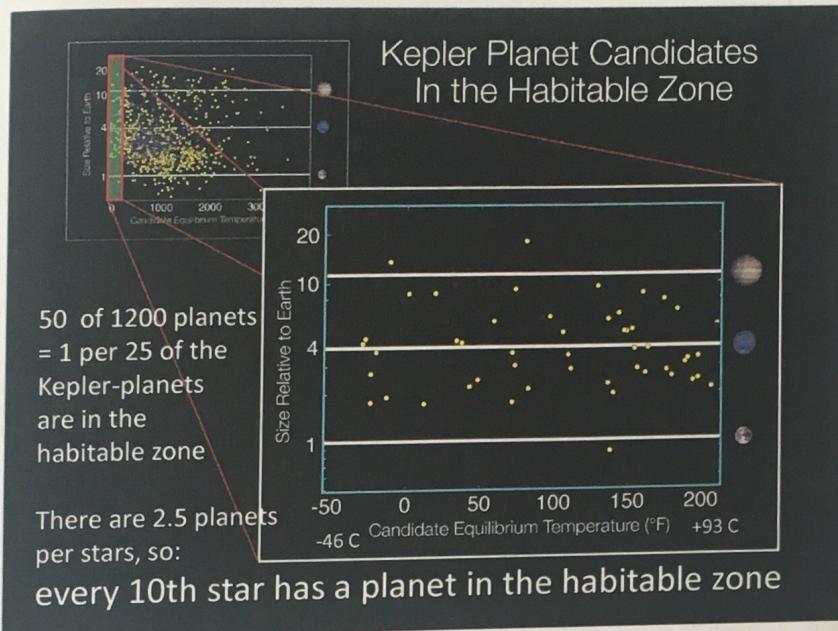
To get a rough estimate of how many rocky planets there are in total in our Galaxy inside the habitable zone of its host star, we can use the number we estimated above of an average of 2.5 exoplanets per star in our Galaxy. Considerations about the type of stars in the Kepler sample and the corresponding orbital distance of the identified planets, place  $\sim 50$  of the 1200 first Kepler candidates identified by February 2011, within or close to the habitable zone as defined in Eq. 1.42 and the associated text. If we use the ratio  $50/1200 = 1/25 = 4\%$  together with the 2.5 planets per star, then  $4\% \times 2.5 = 1/10$ , or one out of every 10 stars in our Galaxy has a small (i.e.,  $\approx$  Mars to Neptune sized) planet in its habitable zone. Since there are of the order of 100 billion stars in our Galaxy, this implies that there are 10 billion planets in our Galaxy inside the habitable zone of the stars they orbit. This is the answer to the first 3 factors in the Drake equation (Eq. 1.40). It is remarkable that the question about the number of potentially life bearing planets in our Galaxy has been debated for at least 3000 years, with sometimes very dramatic consequences and strong emotional involvement, and now in our times the number coming out of the present exoplanet data bases, combined with theoretical models, for the



**Figure 1.43.** The 12 most habitable exoplanets as of February 2017, according to one web page (<http://phl.upr.edu/projects/habitable-exoplanets-catalog>), i.e. the smallest ones most centrally in the habitable zone from Fig. 1.42.

Name	$m_{pl}$ $M_{\oplus}$	r AU	P days	$T_{sf1}$ °C	$T_{sf2}$ °C	$T_{sf3}$ °C	$R_{in}$ AU	$R_{out}$ AU	dist pc	$M_*$ $M_{\odot}$	$L_*$ $L_{\odot}$	spt.
MOA07-192	3.2	0.62	727	-170	-213	-258	0.00	0.01	1300	0.06	0.00	BD
Gliese 370 b	3.6	0.26	54.4	-130	-47	18	0.25	0.47	11.1	0.69	0.33	K5V
HD 40307 b	4.2	0.05	4.3	661	522	312	0.33	0.67	12.8	0.73	0.36	K2.5V
CoRoT-7 b	4.8	0.02	0.9	1507	1269	891	0.52	1.06	150.0	0.93	0.91	K0V
OGLE05-390	5.4	2.10	3500	-151	-196	-244	0.03	0.07	6500	0.22	0.00	M5V
Gliese 876 d	5.7	0.02	1.9	424	312	150	0.07	0.14	4.7	0.32	0.02	M4V
HD 40307 c	6.9	0.08	9.6	482	364	190	0.33	0.67	12.8	0.73	0.36	K2.5V
HD 181433 b	7.6	0.08	9.4	526	403	220	0.37	0.76	26.1	0.78	0.47	K3IV
CoRoT-7 c	8.4	0.05	3.7	882	717	463	0.52	1.06	150.0	0.93	0.91	K0V
HD 285968 b	8.4	0.07	8.8	301	204	66	0.15	0.32	9.4	0.49	0.08	M2.5V
HD 40307 d	9.1	0.13	20.5	337	235	90	0.33	0.67	12.8	0.73	0.36	K2.5V
HD 7924 b	9.2	0.06	5.5	688	546	330	0.42	0.86	16.8	0.83	0.59	K0V
HD 69830 b	10.5	0.08	8.7	596	464	267	0.45	0.92	12.6	0.86	0.68	K0V
HD 160691 c	10.6	0.09	9.6	731	583	360	0.69	1.41	15.3	1.08	1.61	G3 V
55 Cnc e	10.8	0.04	2.8	1132	938	634	0.63	1.29	13.0	1.03	1.35	G8 V
GJ 674 b	11.8	0.04	4.7	282	187	52	0.08	0.17	4.5	0.35	0.02	M2.5V
HD 69830 c	12.1	0.18	31.6	333	232	87	0.45	0.92	12.6	0.86	0.68	K0V

**Table 1.1.** Some of the smaller exoplanets ( $M_{\oplus}$ ; column 2). Column 5 to 7 give possible temperatures at the surface of the planets in °C. Column 5 is the most extreme hot I could think of (a greenhouse warming  $\Delta T$  of 80°C and an albedo of  $A = 0$ ) while column 7 list the most extreme cold I could think of ( $A = 0.79$  as for Venus and  $\Delta T = 0^{\circ}\text{C}$ ). Column 6 gives the surface temperature the planet would have if it had  $A = 0.39$  and  $\Delta T = 40^{\circ}\text{C}$  as the Earth has. Column 8 and 9 gives the inner ( $T=373\text{K}$ ) and outer ( $T=273\text{K}$ ) boundaries of the habitable zone in AU for  $A$  and  $\Delta T$  as on Earth. Column 10 gives the distance from us to the star (and planet) in pc, column 11 the stellar mass, column 12 the stellar luminosity, and column 13 the stellar spectral type.



**Figure 1.44.** The ~50 of the 1200 Kepler planet candidates from the 2011 release which lies in or close to the habitable zone.

first time ever gives us a qualified guess of the answer. The answer is an overwhelmingly large number, but one has to remember that there are 4 more factors in the Drake equation which defines the number of (radio communicating) civilizations in the Milky Way, and remember that the Kepler data monitor only small orbits, and remember that a planet in the habitable zone is not the same as a habitated planet. But we are today, for the first time in history, able to give a qualified guess, based on observations, of how many potentially habitable planets there exist in the Galaxy.

We will return to the other 4 Drake-factors in the following chapters, but first allow ourselves to speculate whether ET is already hiding in our existing data – on one of the exoplanets we have identified and given a name (or a number).

The 3 microlensing planets at 0.6, 2.1, and 2.8 AU, in the table are all close to 1 AU as the Earth's orbit. However, they all orbit very low-luminosity stars. This is because the microlensing technique is most sensitive to the statistically dominant type of stars. Two of the lens stars in Tab. 1.1 are M-dwarfs, and one is on the limit between M-type and brown dwarfs, most likely to be a brown dwarf. Even though they are in terrestrial-like orbits, these planets are much too cold for biology, with surface temperatures likely to be somewhere between  $-150$  and  $-200$  °C. However, an advantage of the microlensing identification of small exoplanets, is that its main sensitivity is in the orbital regime where Earth-sized planets can have formed (according to our understanding of solar system formation), and therefore conditions at their surface can have been stable over biological timescales, such as we think is important for development of life. A larger sample of identified microlensing exoplanets will eventually reveal also solar type stars with Earth-mass planets in Earth-like orbits (if they exist). Statistically, approximately 1/4 of all stars are solar type stars.

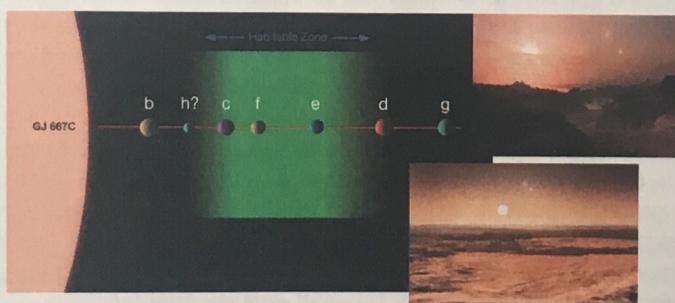
Most of the other planets have the opposite problem of the microlensing planets – they are far too hot for life. Six of the planets in Tab. 1.1, however, stand out by looking reasonable (as concerns temperature) for life to exist at their surface. However, they have another problem which they share with almost all "habitable planets" announced to date. They orbit their star at distances between one tenth and half the distance of Mercury in our solar system. This close to the star tidal forces will have locked one side of the planet's surface permanently toward the star and one side permanently toward the deep freezer of empty space. Most likely the star-facing side will have been heated to atmospheric escape temperatures, and these planets will have been airless and dry since shortly after they formed. Should they miraculously have been spared this

faith, another difficult phenomenon would face potential living organism on these 6 planets. Late type dwarfs have strong magnetic activity, including flares that regularly would reach extremely close to planets like these.

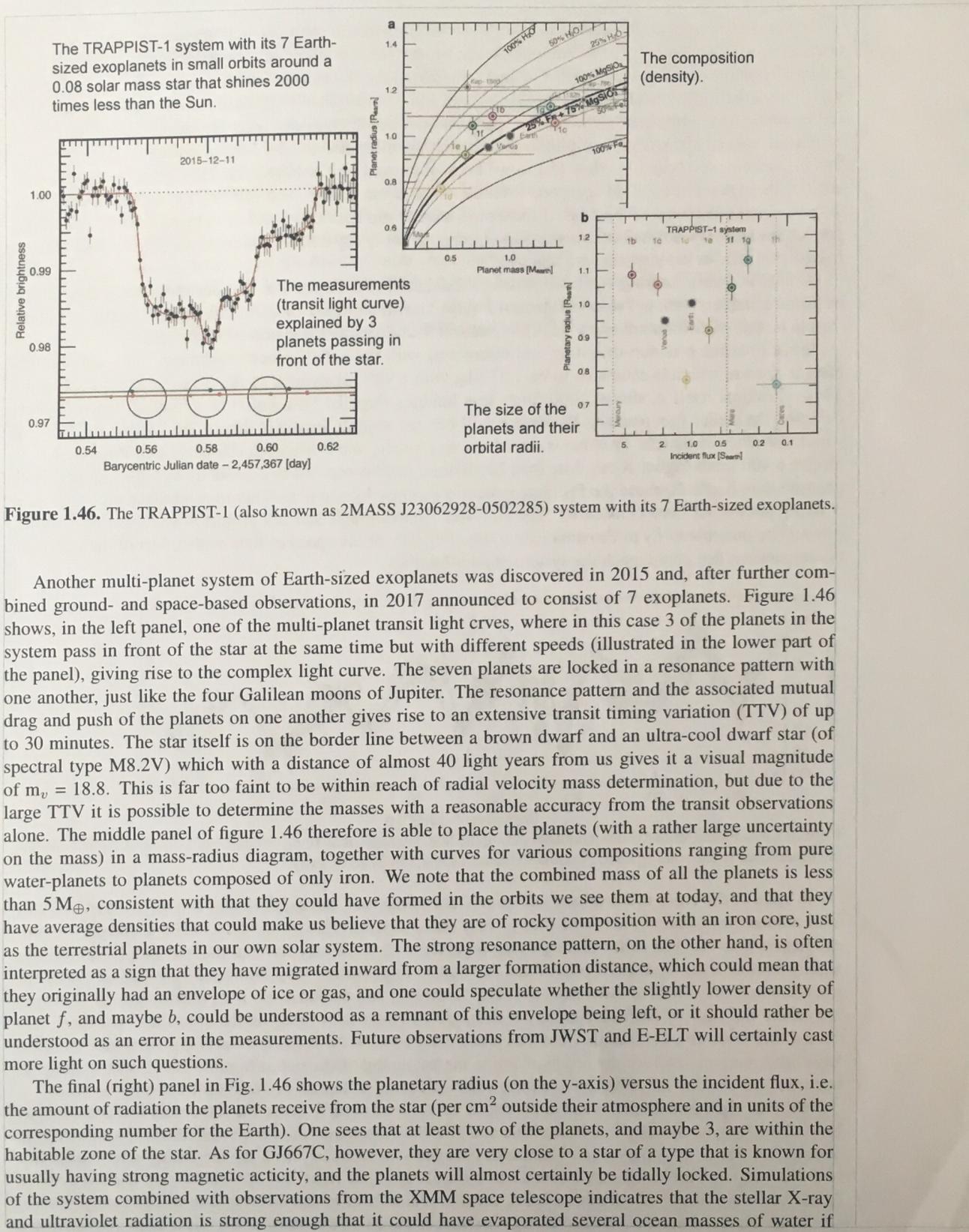
The last concern we would have for the potential inhabitants of the six planets with “biofriendly temperatures” in Tab. 1.1 is that the planets cannot have formed where we see them today. In our own solar system, Mercury became much smaller than Venus and Earth because the proto-solar nebula was too hot for any common minerals to condense out of the cloud to form solid dust grains that could serve to build up the planet. At distances between one tenth and one half Mercury’s distance the situation will not have been better (even around an M-dwarf), and it is hard to envision how e.g. a 5-Earth mass planet could have formed there. The most likely scenario is therefore that all the six planets have formed further out in the nebula and migrated into the position where we see them today. Unless the migration took place in the very beginning of the evolution of the system and then stopped (which is absolutely possible), the conditions will therefore not have been biofriendly for this reason either. There have not been stable conditions during evolutionary timescales, which we would think is a prerequisite for life to exist.

When all the precautions above are mentioned, which one of the systems in Tab. 1.1 or in Fig 1.44 or anywhere in our data bases is then the most likely to house lifeforms that we maybe even potentially could communicate with and/or visit?

GJ 667 is a triple star system with the two brightest components, GJ667A and GJ667B, orbiting only 12 AU from one another, with the third dimm component GJ667C in a 220 AU orbit around the A/B pair. In 2013 it was suggested that GJ667C is orbited by no less than 3 super-Earth planets within the habitable zone of GJ 667C, as illustrated in Fig. 1.45. Two of the planets seem “just right” (GJ667C f and e), one is on the warm end of the biofriendly temperatures (GJ667Cc), and one is right outside the cold end of the habitable zone (GJ667Cd) like Mars in our solar system, but with a sufficient mass to potentially adjust to life with an extreme greenhouse heating atmosphere. The four planets are not far from one another – within moving distances for an advanced civilization that stellar evolution could have pressed out of its cradle once it got too hot at their first home. Combined the 7 proposed planets, however, consist of  $25 M_{\oplus}$  – far more solid material than condensed inside the snow line in our solar system, and mysteriously much for a tiny M1.5 dwarf star of  $0.3 M_{\odot}$ . In 2014 it was proposed that 5 of the 7 planets rather were noise in the data than real exoplanets. Only GJ667Cb and GJ667Cc seems to have survived further tests. The surviving GJ667Cc is by some considered one of the most likely candidates for a known habitable exoplanet (e.g. Fig. 1.43). If GJ667Cc has an atmosphere and albedo identical to the Earth’s, it will have an average surface temperature only  $25^{\circ}\text{C}$  warmer than the Earth. However, the star GJ667C itself is an M1.5 dwarf, which usually has strong magnetic activity, and GJ667Cc is within 1/3 of Mercury’s distance from the Sun and will continuously be bombarded with huge amounts of x-ray radiation and ionized gas-masses, at the same time as it is tidally locked to the star. Could life have arisen and evolved under such conditions?



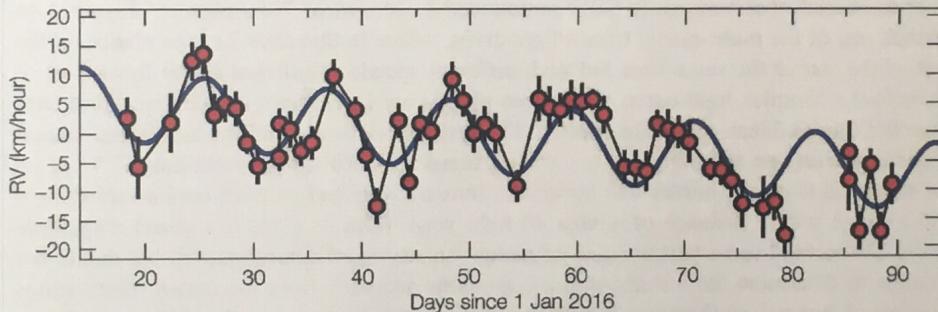
**Figure 1.45.** The proposed GJ667C planetary system with 3 super-Earth sized planets in the habitable zone and additionally 2 right outside and 2 more further out of the HZ in this 7-planet system of the M-dwarf component of a triple star system . Insets are artists impression of what GJ667Cc on the inner edge of the habitable zone could look like (upper inset) with the red dwarf host star GJ667C bright on the sky at a diameter 2.5 times the way the Sun looks seen from the Earth, and with the two companion stars GJ667A and GJ667B bright as full moons, and what GJ667Cd (right inset) could look like, envisioned as with a cool Mars-like surface and the 3 suns at a somewhat greater distance. Today it seems that 5 of the planets were just noise in the data.



water was ever present on these planets. While the odds for life to exist on such planets are small, it is still fascinating to speculate what it would look like to be an inhabitant of one of these exoplanets, bathed in soft red light from the cool dwarf star, with 6 planets wandering the sky at a speed that would make it possible for you to track their movement with the naked eye, and some of the planets appearing larger in the sky than the full moon seen from the Earth. Maybe the system would be a convenient target for consideration for future humans to move to in a couple of billion years when the magnetic activity has diminished, when our own Sun has become too bright to allow life as we know it to exist on Earth, and the TRAPPIST-1 system still has 6 trillion years of its main sequence lifetime left to shine as now, and offering plenty of opportunities to move from one planet to the other if the temperature is too hot or too cold for a pleasant stay. The SETI project has already performed extensive radio searches to see if any radiocommunicating civilisation already lives there, without having heard anything so far.

A final and very exciting candidate among the faint M-dwarf stars that harbour Earth-sized exoplanets in their habitable zones, is Proxima Centauri b (also known as Alpha Centauri Cb, or just Proxima-b). The system is only 40 trillion km away, 4.2 light years. In fact Proxima is the nearest known star.

Since Proxima b is non-transiting and discovered with the radial velocity technique, only a minimum mass is known, which is estimated to be  $1.27 M_{\oplus}$  with a 90% likelihood of the mass being no larger than  $3 M_{\oplus}$ . Proxima itself is almost 1000 times less luminous than the Sun, but at an orbital distance of only 0.05 AU, the stellar flux reaching Proxima b will be only slightly less than that of the Earth ( $\sim 65\%$ ). As with most other M-dwarfs, Proxima is, however, very active, and it has been estimated that Proxima b will receive a 400 times higher X-ray dose than Earth does, and be exposed to a 2,000 times higher stellar wind pressure than Earth. Because the Proxima system is so close by, the possibility to send space probes to take close up photos has been considered. The so called Breakthrough Starshot project is investigating whether it would be possible to fly to Proxima with a speed of 20% of the speed of light with a fleet of thousands of micro-cameras that could reach the system in mid-2060's.



**Figure 1.47.** Radial velocity measurements during the first 3 months of 2016 revealed an Earth-mass exoplanet orbiting the nearest star to the solar system, Proxima Centauri, only 4.2 light years away.

In spite of the close to Earth-size and convenient stellar flux, the small orbital distance of the habitable zone around the magnetically active M-dwarfs, may make both the GJ667C, the TRAPPIST-1, and the Proxima exoplanets poor candidates for finding living organisms. If so, it may turn out that the exoplanets orbiting slightly warmer stars, such as HD 40307 and HD 69830, might be the most biofriendly candidates among the known exoplanets. Both of these stars are early K-dwarfs, and the habitable zone planets around both stars are on the hot edge of biofriendliness, which might be better than being on the cold side. The stars have started their lives dimmer than they are today, and conditions on the surface of the two planets may therefore have been more optimal for the origin of life in the beginning of the stellar lifetime than it seems today. In our solar system, the Sun is believed to have been increasing its luminosity with 40% since it was born.

With the right initial albedo and atmospheric conditions, at early times when the stars had considerably lower luminosity than today, the HD69830c and HD40307d planets could have started out at habitable conditions. On Earth, the atmospheric conditions changed synchronously with the increasing solar luminosity, in such a way that the temperature conditions were always “just right” for life. As we will discuss in detail in the following chapter about life on Earth, some scientists believe that this response was an emergent reaction of the Earth’s total biosphere to changes in the cosmic environment, and that this is even a fundamental

## 1. EXOPLANETS

---

principle of life as a concept. If this has any grain of truth in it, we could then well ask whether possible life forms on HD69830c and HD40307d could have found their own appropriate response to the increasing luminosity of their stars, in order to have kept their planet continuously habitable. If life on HD69830c and HD40307d has managed to increase the albedo of the planets to  $A_{pl} = 0.8$  or more, and make the atmosphere transparent (e.g. a pure N<sub>2</sub> and O<sub>2</sub> mixture), as a response to the increasing luminosity of HD69830 and HD40307, then life on these two planets could in principle have managed to keep the planet habitable for an extensive evolutionary timescale.

The K0 dwarf star HD69830 is 7 billion years old. We argued above that it is not impossible that HD69830c has been habitable for most of this time, and 7 billion years is almost twice the timespan of our own planet's habitability. Maybe it is here we should send the questions about how one solves the pollution crisis and about the meaning of life (the two most common questions layman wants to ask ET, according to opinion polls). At least HD69830c (and HD40307d) ought to be prime candidate for SETI and for E-ELT; Both of these two stars are only 40 light years away! -- already our grandchildren could receive the answer to the questions we could think about sending to HD69830c and HD40307d, if we send them today.