

Chapter 2

Detection Methods

Surprising though it may seem, practically all of the exoplanets that have been currently detected have never been ‘seen’ directly, in the sense that no images obtained with a telescope exist for these objects. The reason is very simple, and lies in the extreme difficulty of detecting these objects. This also explains why it was necessary to wait for the beginning of the 1990s for the first systems to be discovered, when far more exotic astrophysical objects (quasars, pulsars, etc.) had been imaged for several decades. In this chapter, we discuss the different techniques employed to detect and determine the properties of exoplanets and their environment.

2.1 The Extent of the Problem

The direct observation of an exoplanet – in the sense of being able to separate physically the photons from the planet from those of the central star, sufficient to obtain an image of the two objects – is a problem that is as simple to state as it is difficult to achieve in practice. It may be largely summarized by three critical points which strongly influence the detection methods that may be envisaged:

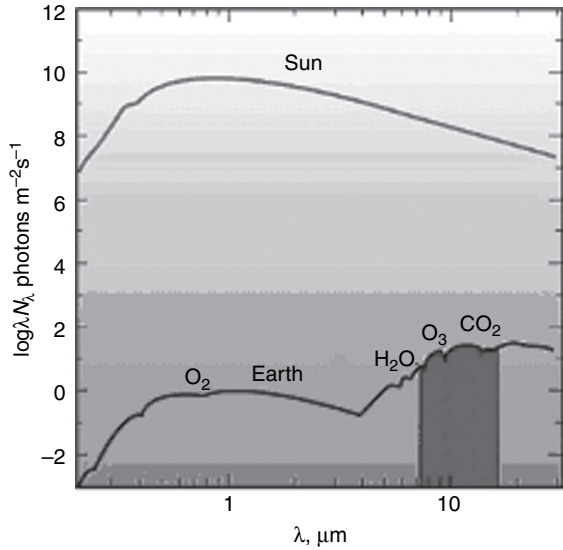
- the contrast between the star and the planet,
- the angular distance between the two objects,
- the environment of the Earth and of the exoplanet (and more generally, of the exosystem as a whole).

2.1.1 Contrast Between Star and Planet

The contrast between an exoplanet and its parent star depends on several factors:

(a) the spectral region in which observations are made: The spectrum of an exoplanet primarily consists of two components (see Chap. 7): one component arising from the reflection of the star’s light by the planet, and the other being its own emission component. In the case of a planet like the Earth, orbiting a solar-type

Fig. 2.1 Comparative spectra of the Sun and the Earth as they would be seen from a distance of 10 parsecs. The Earth's spectrum clearly shows the component consisting of reflected sunlight (0.1–4 μm) and the component linked to the planet's own emission (beyond 4 μm)



star (Fig. 2.1), the reflected component is dominant in the visible and near-infrared regions, while the planet's own emission component dominates in the thermal infrared (beyond 4 μm).

(b) the nature of the stellar and planetary objects: The two spectral components just mentioned obviously depend on the nature of the various objects forming the planetary system. The spectral distribution and the amplitude of the reflected component depend on the star's effective temperature (and thus on the type of star), but also on the albedo, the size, and the distance of the planet from its star. The planetary emission itself depends exclusively on the nature of the planet – whether terrestrial or giant – and primarily on its effective temperature, but also on the atmospheric composition, which is indicated by the presence of absorption lines in the emission spectrum (Fig. 2.1).

(c) the distance between the objects: The distance between the star and the planet is one of the essential factors involved in the calculation of the radiative equilibrium temperature of the planet. For example, an object of the size of Jupiter has an effective temperature of about 1500 K at 0.05 AU (i.e., as is the case with a hot Jupiter), whereas it is only about 110 K at 5.2 AU (Jupiter's location in the Solar System). The contrast is directly affected: at a wavelength of 10 μm it is about 10^4 for a hot Jupiter, but it rises to 10^9 for a true Jupiter-like planet.

2.1.2 Angular Separation Between the Objects

Because of the significant distance between the Sun and the nearest stars – the closest stars are several parsecs from us – exoplanetary systems subtend very small angular distances. A Sun-Earth system, for example, at a distance of 10 parsecs

would subtend an angle of one tenth of a second of arc. (One second of arc, abbreviated ‘arcsec’, is 1/3600th of a degree, one milliarcsec, abbreviated ‘mas’, is 1/1000th of an arcsec, one microsecond of arc abbreviated ‘ μ as’, is 1/1000th of a mas.) This tiny angular separation means that methods with high angular resolution must be used to carry out direct observation (see Sect. 2.3). An illustration of these problems of contrast and angular separation may be framed as follows: Trying to observe a planet like the Earth orbiting a star like the Sun, the whole system being at a distance of 10 parsecs from us, resembles trying to observe a glow-worm 30 cm from a lighthouse in Marseille, when the observer is in Paris (at a distance of about 700 km).

2.1.3 Environment of the Earth and Exoplanets

Apart from its suite of planets, the Sun is surrounded by a disk of dust resulting from collisions between asteroids as well as cometary dust. This disk of dust lies in the plane of the ecliptic. The dust, both illuminated and heated by the Sun, has its own emission, which (in the visible region) is known as the zodiacal light. An observer in the Solar System who is examining the sky in the infrared therefore records a signal from this zodiacal emission. This emission is not negligible because, at a wavelength of 10 μ m, the spatially integrated emission from the whole disk of dust is 300 times the Earth’s emission.

In the same manner, it is not unreasonable to assume that exoplanetary systems also contain a disk of debris (see Chap. 5), whose emission is certainly not negligible, and may even be greater than that of the Solar System. The presence of this disk therefore is a source of a parasitic signal in any radiation that is detected. To illustrate this last point, we may return to our earlier analogy by imagining that this time our glow-worm and our observer are both bathed in the light from Marseille and Paris, which restricts the visibility of faint objects. If you need to be persuaded of this, all you need do is compare the number of stars visible from the centre of Paris with those visible on a clear night in the wide-open country.

Table 2.1 Properties of certain typical objects which are assumed to be orbiting a solar-type star, lying at a distance of 10 parsecs

Object	Radius (R_{\oplus})	Mass (M_{\oplus})	Dist. (AU)	Angular separation (mas)	Contrast (Vis)	Contrast (IR)
Jupiter	11.18	317.83	5.2	520	5×10^8	5×10^7
Saturn	9.42	95.15	9.54	954	2×10^9	5×10^8
Neptune	3.94	17.23	30.06	3006	10^{11}	10^{10}
Earth	1	1	1	100	5×10^9	7×10^6
Hot Jupiter	10	300	0.05	5	4×10^4	10^3

Table 2.1 summarizes the main factors for different planetary objects, assuming that they are in orbit around a star like the Sun, lying at a distance of 10 parsecs from Earth.

2.2 The Indirect Detection of Exoplanets

We have seen that the direct detection of an exoplanet orbiting a star, even one close to the Solar System, is extremely difficult. It is not, however, absolutely essential to be able to obtain an image of the planet to reveal its presence. In certain cases it is possible to detect the planet by observing the effect it has on its parent star, thus by observing the star itself. The techniques that are based on this principle are known as ‘indirect’ methods, in contrast to direct imaging. In this section we discuss the principal indirect methods of detection. Here, we should explain that these methods are by no means theoretical, but that, with a few exceptions, all the exoplanets currently detected have been found by indirect methods.

2.2.1 The Effect of a Planet on the Motion of Its Star

It is commonly said that a planet orbits a star. This assertion is true if one neglects the mass of the planet relative to that of the star, or if the motion of the planet is considered in very rough terms. In reality, the star and its planet (or planets) are bound by gravitation, and each of the bodies in the system (star and planet or planets) has a motion about the centre of mass (the centre of gravity) of the system. The position of the centre of mass G , in a system with N bodies, each of mass m_i , the centre of which is at O_i , is defined by the following vector relationship:

$$\sum_i m_i \cdot \overrightarrow{GO_i} = \vec{0} \quad (2.1)$$

In particular, for a two-body system (the star being denoted by ‘*’ and the single planet by ‘p’), the centre of mass lies between the star and the planet, and the vectorial expression just given becomes an algebraic one:

$$\overline{GO_p} = \frac{m_*}{m_* + m_p} \overline{O_*O_p} \quad (2.2)$$

where $\overline{GO_p}$ denotes the distance between G and O_p and $\overline{O_*O_p}$ the distance between O_* and O_p .

Each of the bodies thus follows an elliptical orbit, with the centre of mass of the system at one of the foci. If a is the semi-major axis of the planet’s orbit around the star, the semi-major axes of the stellar and planetary orbits relative to the barycentre of the system at one of the foci may be written:

$$a_* = \frac{m_p}{m_* + m_p}a \text{ and } a_p = \frac{m_*}{m_* + m_p}a \quad (2.3)$$

In the general case, the ‘central’ star also describes an orbit that is more or less complex depending on the number of planets in the system, certain properties of which may be described by observing:

- stellar motion projected on the plane of the sky. Here, we observe a variation in the position of the star relative to a fixed reference frame (consisting of very distant bodies, such as quasars). This method is known as ‘astrometry’.
- stellar motion along the line of sight (radial motion). Here, we measure the star’s velocity of approach or recession as a result of the motion around the centre of mass. This measurement method is known as ‘radial velocimetry’ or ‘Doppler velocimetry’ (after the experimental method used to measure the radial velocity).

In certain instances, when the star emits a periodic signal (as with a pulsar, for example), the motion of the source may be deduced from the changes in the pulsar’s period as it is measured here on Earth. We shall return to this technique at the end of this section.

2.2.1.1 Astrometry

As the planet or planets move around the centre of mass, the central star also describes an orbit, the complexity of which depends on the number of planets in the system. In the case of our Solar System, the motion of the Sun is primarily caused by the presence of Jupiter, but also reveals the presence of other, less-massive planets (Fig. 2.2).

In the case of a system with a single planet (or where one of the objects dominates the others by a wide margin to such an extent that its effect masks those of all the other planets), the motion of the star is an ellipse. We shall be particularly concerned with such orbits, the general equation for which may be expressed, using polar coordinates having their origin at the focus, and in the orbital plane, by:

$$r(v) = \frac{a(1 - e^2)}{1 + e \cdot \cos(v)} \quad (2.4)$$

where a is the semi-major axis, e the eccentricity, and v the angular position of the object relative to an origin given by the orbit’s periastron.

The motion, projected onto the plane of the sky is deduced from the preceding expression by a change of reference frame that takes into account the orientation of the orbital plane relative to that of the sky, and also the conventions regarding the notation of the associated angles.

To describe the orientation of the orbit relative to the plane of the sky, it is normal to use the astronomical equivalent of Euler angles. The following are thus defined successively:

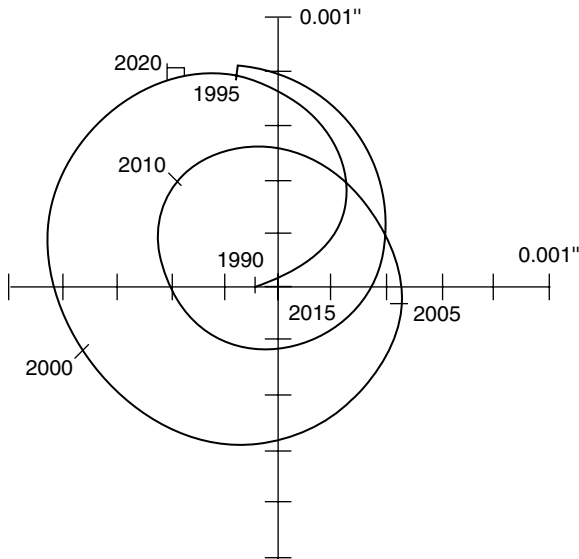


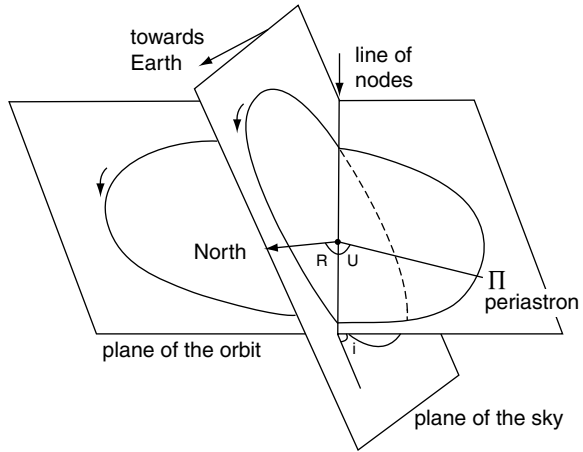
Fig. 2.2 Motion of the Sun as a function of time (in years) on the plane of the sky as it would appear from a location 10 parsecs away and perpendicular to the plane of the ecliptic. This movement is mainly dominated by the giant planets (Jupiter, Saturn, Uranus, and Neptune). Note the scale: the variations in positions are approximately 1 mas (0.001 arcsec noted 0.001'')

- *the line of nodes*: this is the straight line given by the intersection of the orbital plane with the plane of the sky
- *the ascending and descending nodes*: these are the intersections of the line of nodes and the object's orbit. By convention, the ascending node is crossed when the object recedes from the observer, and the descending node when the object approaches the same observer
- *the position angle of the ascending node*, denoted Ω : it describes the orientation of the line of nodes on the plane of the sky relative to celestial north. This angle lies between 0 and 180°
- *the inclination*, denoted i : this is the angle between the object's orbit and the plane of the sky. It is equal to 0 (or 180°) when the orbital plane and the plane of the sky coincide; i lies between 0 and 90° when the apparent motion of the object on the plane of the sky takes place in the trigonometrically direct sense, and lies between 90 and 180° in the opposite case (Fig. 2.3).

It can be shown (Bordé, 2003) that the position of the star in the reference frame based on the plane of the sky may be deduced from that in the plane of the orbit by 3 successive rotations:

- a rotation by the angle $-\Omega$ around the line of sight
- a rotation by angle i around the line of sight
- a rotation by angle ω around the normal to the plane of the orbit

Fig. 2.3 Definition of the geometrical parameters describing the orientation of the orbit of a star relative to the plane of the sky



The expression for the path on the plane of the sky is thus obtained in polar coordinates, whose origin is the centre of mass of the system:

$$\begin{aligned}\rho &= r\sqrt{1 - \sin^2 i \sin^2(\nu + \omega)} \\ \theta &= \arctan[\cos i \tan(\nu + \omega)] + \Omega\end{aligned}\quad (2.5)$$

which is still an elliptical path, and where the parameters defining it are the parameters of the initial orbit (in the reference frame of the orbit) and the observational geometry.

To detect an exoplanet by the astrometric method, it is thus necessary to be able to reconstruct the star's apparent path on the plane of the sky. By using Eq. (2.4), and by making the approximation that the star's true path is circular, and in the plane of the sky, it is possible to calculate the maximum value of the amplitude of the motion, which may be written as:

$$\delta\theta = \frac{a_*}{D} = \frac{m_p}{m_*} \cdot \frac{a_p}{D} \quad (2.6)$$

where m_p and m_* are the masses of the planet and the star, respectively; a_p is the semi-major axis of the planet's orbit, and D is the distance of the exoplanetary system from Earth. The value of $\delta\theta$ may be calculated for different cases (Table 2.2).

This table and Eq. (2.6) show clearly that the astrometric method is more sensitive (i.e., the amplitude of the motion of the star is the greater), if the planet is:

- massive
- located at a great distance from its star.

However, the greater the distance of the planet from the star, the longer the planet's orbital period – and consequently that of the star as well – (in accordance with Kepler's Third Law). This means that the time required to detect the planet

Table 2.2 The amplitude of the apparent motion produced by the planets in the Solar System acting on the Sun, as if the latter were being observed from 5, 10, or 15 parsecs

Object	$a_*(r_\odot)$	$\delta\theta$ at 5 pc (mas)	$\delta\theta$ at 10 pc (mas)	$\delta\theta$ at 15 pc (mas)
Jupiter	1.07	1.00	0.50	0.33
Saturn	0.59	0.55	0.27	0.18
Neptune	0.33	0.31	0.15	0.10
Uranus	0.18	0.17	0.08	0.06
Earth	6.5×10^{-4}	6.0×10^{-4}	3.0×10^{-4}	2.0×10^{-4}

(a significant fraction of the planet's orbital period around the star) also increases drastically with the planet's distance from the star. Nevertheless, the astrometric method remains one of the rare methods sensitive to low-mass objects, orbiting far from their star. In the case where the system is multiple (in practice, as soon as there are two planets with comparable effects), the path of the star rapidly becomes complex (see Fig. 2.2), and the accuracy of the reconstruction becomes limited by the resolution of the measurements.

Several techniques are possible to determine accurately the position of stars in the sky, and consequently their proper motion. The common factor with all these techniques is the necessity of being able to use a fixed reference frame, or one that is considered as such, on the sky. The principal difficulty is obviously that of finding the position of a fixed object in the sky to a greater degree of accuracy than the astrometric accuracy of the measurements (typically, to 1 μ arcsec for the most accurate instruments). The positions of stars are generally referred to distant point objects, if possible, extragalactic ones. Nowadays, the ultimate reference frames consist of those based on quasars (quasi-stellar objects), which are actually the active nuclei of extremely distant galaxies. In the past, the position of nearby stars (their distances being determined by measurements of their parallax) was measured relative to more distant stars lying within the same field of view.

The first astrometric measurements aimed at detecting the presence of low-mass companions coincided with the systematic use of photographic plates for astrophysical observations. Peter Van de Kamp, who, in 1963, announced the detection of two planets orbiting Barnard's Star (the closest star to the Sun after the Alpha Centauri triple system, one component of which, Proxima Centauri, is the closest star to the Sun), based his measurements on the analysis of more than 2400 photographic plates taken between 1916 and 1963 by the various successive astronomers who used the 80-cm telescope at Sproul Observatory. In this case, a measurement consisted of locating on the photographic plate the coordinates of the photocentre of the reference objects (stars considered to be 'fixed'), and of the star for which one wanted to determine the proper motion. All that remained was to reconstruct the star's path, subtract the proper motion caused by parallax (the motion of the Earth in the Solar System), and the motion of the star in the Galaxy, to obtain the motion

of the star that reflected the possible presence of a planet. The principal difficulty with this method is its low precision. The size of objects on a plate is, at best, 1 arcsec (that is the size of the physical spot forming the image with low atmospheric turbulence), and it is difficult to obtain a measurement on a photographic plate with a non-linear response, to better than one tenth of the size of the spot (i.e., 0.1 arcsec), to compare with the values in Table 2.2. Even with systematic measurements over several years, and using electronic cameras (CCD detectors), it is difficult, working from the ground and with classical telescopes, to obtain an astrometric accuracy better than 10 mas, which is too great for us to hope to detect planets – even giant ones.

The planets announced by Peter Van de Kamp remain, to this day, unconfirmed, despite the arrival of radial-velocity techniques that are far more accurate than astrometry with photographic plates. It is highly likely that Van de Kamp did not know how to overcome all the bias involved with the method.

Apart from the systematic use of electronic cameras (CCD devices), which have appreciably increased the accuracy and reproducibility of astrometric methods, two techniques have completely revolutionized this discipline and these are:

- observation from space
- interferometry.

Observation from space allows us to avoid all the problems arising from the atmosphere. In particular, turbulence, with its accompanying differential refraction, produces fluctuations in the position of an object against the plane of the sky. The degree of these fluctuations depends on the meteorological conditions under which the observations are made, and thus on the site's astronomical quality. The European whole-sky astrometric satellite, Hipparcos, was therefore able to obtain, with an accuracy of a few mas, the position and the proper motion of more than 120 000 stars in the Galaxy. This precision is at the upper limit for the detection of giant planets (see Table 2.3). The European GAIA project should, in 2012, allow us to attain an accuracy of a few μ as. It will then be possible to search for giant planets using astrometry.

Interferometry is a technique which uses two or more telescopes, the light from which is recombined in pairs to obtain interference fringes. From measurements of these fringes, one can deduce information about the spatial structure of the object with an extremely high angular resolution (as would be obtained with a telescope having a diameter equivalent to the distance between the interferometric telescopes). In its astrometric mode, interferometry allows an extremely precise measurement of the position of an object against the sky. A diagram showing the principles behind the method is given in Fig. 2.4.

The light coming from the star of which one wants to measure the position, arrives at telescope 2 with a delay relative to the light arriving at telescope 1. This delay, caused by the angle θ between the direction of the star and the baseline, B , between the telescopes, creates an external path difference $B \cdot \cos(\theta)$ between the two rays. This path difference may be compensated by the delay line which enables the optical path followed by the light from telescope 1 to be increased or decreased. Zero

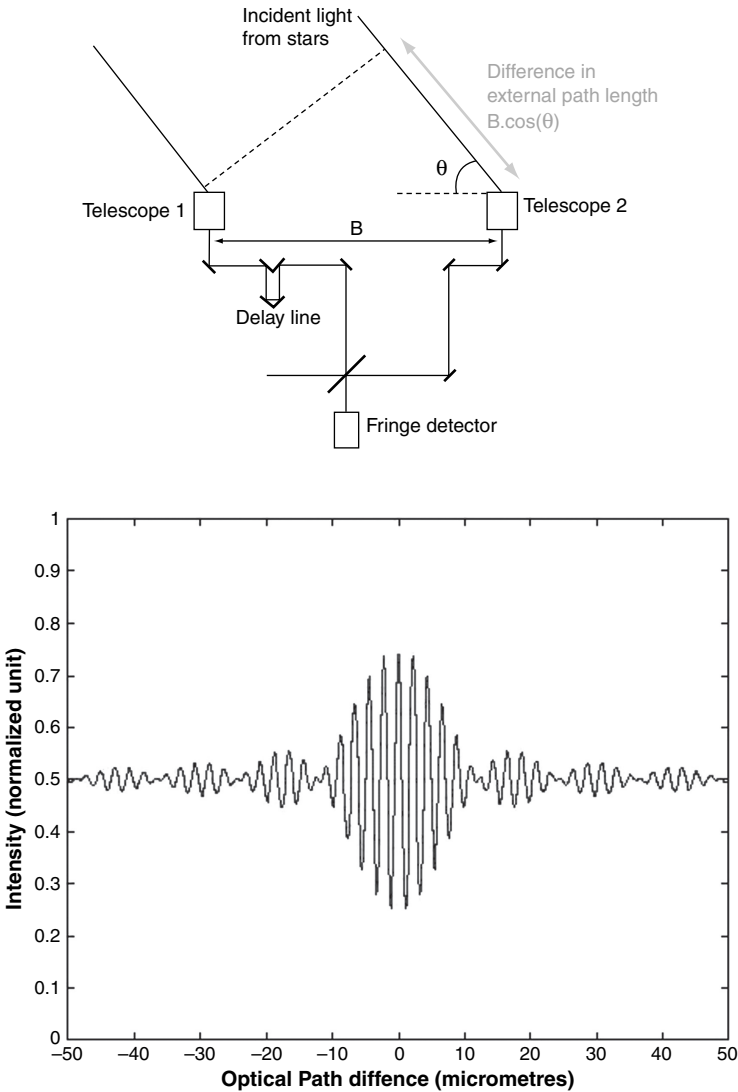


Fig. 2.4 The principle behind interferometry (described fully in the text)

path difference (the optical paths followed by the light gathered by both telescopes 1 and 2 being equal) is determined by observation of the interference pattern (on the bottom of Fig. 2.4). It corresponds to the point at which the envelope over the interference fringes is a maximum. The precise determination of B , by optical metrology (with an accuracy of a few nanometres), and measurement of the difference in the external path length ($B \cdot \cos \theta$) by measurement of the displacement of the delay line relative to zero path difference (again, with an accuracy of some few nanometres),

provides an accurate determination of the angle θ between the telescope's baseline and the star on the sky.

In practice, interferometry, like the other methods, functions in a differential mode to avoid the effects of diurnal rotation and of the variation in the baseline during the observation. Here, the astrometric movement of the star is measured relative to a reference frame that is assumed to be fixed. The interferometer should therefore examine at least two objects simultaneously. This standard of performance should be obtained with the planned PRIMA (Phase Reference Imaging and Micro-arcsecond Astrometry) instrument on the European Very Large Telescope Interferometer (VLTI), under commissioning at the time this book was written.

The best astrometric measurements are currently obtained by radio interferometry with continental baselines (VLBI), where astrometric accuracy close to $100 \mu\text{as}$ is being attained. The optical interferometers that will shortly enter service should enable us to obtain astrometric accuracies that are at least 10 times better.

An additional stage in the search for the ultimate astrometric accuracy consists of combining interferometry with observations from space. The American SIM mission is a space-borne optical interferometer, which, around 2015, should be able to obtain an astrometric accuracy of $1 \mu\text{as}$. To obtain such accuracy, it is necessary to know the length of the baseline to an accuracy of a few tens of picometres (in other words, the thickness on one atomic layer). Research and development efforts are currently under way to reach this objective, and allow the mission to be developed.

2.2.1.2 Radial Velocimetry

The astrometric method that we have just discussed, should enable us to measure and reconstruct the motion of the star in the plane of the sky. It omits, however, measurement of the star's motion along the line of sight, recession or approach relative to the observer. It is, however, impossible to observe, from Earth, the radial location of the star along the line of sight, because, by definition, the motion projected on the sky is non-existent. It is, however, possible to measure the radial velocity of the star by spectroscopy (see later). The expression for the radial velocity of the star may be derived from Eq. (2.4) and from the change of reference frame introduced in the preceding paragraph, which enables a conversion from the orbital reference frame to that of the sky, namely by rotations by the angles $-\Omega$, i , and ω . The equation for the position of the star on the line of sight is then written as:

$$z(t) = r \sin(v(t) + \omega) \cdot \sin(i) \quad (2.7)$$

By using Kepler's Second Law (the law of areas), which is expressed, for an elliptical orbit of period P , semi-major axis a , and eccentricity e , by the equation:

$$r^2 \frac{dv}{dt} = 2\pi P^{-1} a^2 \sqrt{1 - e^2} \quad (2.8)$$

and deriving the expression of the position of the star on the line of sight with respect to time, we obtain the equation for the star's radial velocity:

$$V_r = \frac{2\pi}{P} \frac{a_* \sin(i)}{\sqrt{1-e^2}} \cdot [\cos(v + \omega) + e \cdot \cos(\omega)] \quad (2.9)$$

In what follows, we will set:

$$K_* = \frac{2\pi}{P} \frac{a_* \sin(i)}{\sqrt{1-e^2}} \quad (2.10)$$

In the two-body case that we are considering, from the preceding equation, it is thus possible to derive an equation that is a function of the mass of the planet and the mass of the star. This equation is obtained using Kepler's Third Law, applied to the system consisting of the star and the planet:

$$a^3 (m_* + m_p)^{-1} = P^2 G (4\pi^2)^{-1} \quad (2.11)$$

where $a = a_* + a_p$ and $m_* \cdot a_* = m_p \cdot a_p$. As a result we have:

$$a_*^3 = P^2 G (4\pi^2)^{-1} \cdot m_p^3 (m_* + m_p)^{-2} \quad (2.12)$$

and we then obtain, using Eqs. (2.9) and (2.12) the equation:

$$\frac{(m_p \sin(i))^3}{(m_* + m_p)^2} = \frac{P}{2\pi G} K_*^3 (1 - e^2)^{3/2} \quad (2.13)$$

The above equation shows that measurement of the radial velocity (Eq. 2.9) does not allow us to simultaneously determine m_* and m_p . However, if we assume that the mass of the planet is negligible relative to the mass of the star, and also that the mass of the star may be estimated by some other means (for example from the position of the star on the Hertzsprung-Russell diagram, see Fig 1.12), then we may obtain an estimate of the product $m_p \cdot \sin(i)$:

$$m_p \sin(i) \approx \left(\frac{P}{2\pi G} \right)^{1/3} K_* \cdot m_*^{2/3} \sqrt{1 - e^2} \quad (2.14)$$

It will be noted that, whatever we do, we cannot derive equations that omit the system's angle of inclination (i). So measurement of radial velocities only allows us to derive a minimum mass for the planet. Only in the specific case where the system is seen edge on, and where the planet may also be detected by its transit of the star, is it possible to determine the individual masses of the system's planets. However, by assuming that planetary systems have a random orientation relative to the plane of the sky, it is possible to ignore the orientation statistically, and obtain a mass-distribution for the planets as a whole.

In the case of a circular orbit ($e = 0$), Eq. (2.14) may be expressed numerically as:

$$m_p \cdot \sin(i) \approx 3.5 \cdot 10^{-2} K_* \cdot P^{1/3} \text{ and } K_* = \frac{2\pi}{P} \frac{m_p}{m_*} \cdot \sin(i) \cdot a_p \quad (2.15)$$

Table 2.3 Properties and radial-velocity changes induced on the star by various planetary objects

	$m_p(m_\oplus)$	$a_p(\text{AU})$	$a_*(r_\odot)$	P (days)	$K_*(\text{m.s}^{-1})$
Jupiter	317.83	5.2	1.07	4332.6	12.5
Saturn	95.15	9.54	0.59	10759.2	2.8
Uranus	14.54	19.18	0.18	30685.4	0.3
Neptune	17.23	30.06	0.33	60189	0.28
Earth	1	1	6.5×10^{-4}	365.25	0.09
51 Peg b	130	0.05	0.004	4.23	50.2

where K_* is in m.s^{-1} , P in Earth years, and m_p in Jupiter masses. Table 2.3 gives the value of K_* for various specific instances.

If we continue to neglect the mass of the planet relative to the mass of the star, we may also deduce, from Kepler's Third Law, the semi-major axis of the planet's orbit around the centre of gravity (and thus around the star, because in this approximation, they are assumed to coincide) as a function of the mass of the star and of the orbital period:

$$a_p = \left(\frac{G}{4\pi^2} \right)^{1/3} P^{2/3} m_*^{1/3} \quad (2.16)$$

As Table 2.3 clearly shows, measurement of the radial velocity of a star to determine the presence of possible planets is a method that is biased towards:

- massive objects (K_* is proportional to the mass of the planet),
- objects close to their parent star (K is proportional to $P^{-1/3}$, and therefore greater, the smaller the value of P).

For massive objects, the radial-velocity method is therefore complementary to the astrometric method, which is more sensitive to distant objects. As with astrometry, to detect long-period objects the method requires regular observations and an instrumental stability that is monitored over time. We shall return to this point in the discussion of the equipment that enables these measurements to be carried out (Chap. 8).

Unlike the astrometric method, and as indicated earlier, measurement of the radial velocity does not allow us to determine the mass to better than the sine of the angle of inclination. So it does not allow us to observe systems that are viewed face-on. Finally, we should mention that in the case of multiple systems, the radial velocity, taken overall, consists of the sum of the different contributions of the planets, and of their periods of revolution. The detection of multiple systems (such as 55 Cancri, Upsilon Andromedae, etc.), is therefore carried out by analyzing the components of the radial velocity by subtracting the principal component, followed by analysis of the residuals until the measurement noise is reached.

Measurement of the radial velocity of a star is based on the Doppler-Fizeau effect:¹ any observer who receives a wave (of whatever nature) emitted at the frequency ν by a source in motion, detects it at the frequency $\nu + \delta\nu$, where $\delta\nu$ is positive (greater frequency and thus shorter wavelength) if the object is approaching, and negative (lower frequency and greater wavelength) if the object is receding.

A star is a source of electromagnetic waves: every object at a temperature T , has a thermal emission spectrum described, in the black-body model, by the Planck function. The star, as a source of electromagnetic waves may therefore equally be the source of a Doppler-Fizeau effect. The analogies with sound waves ceases here, however, because electromagnetic waves are transverse waves which propagate in a vacuum, whereas sound waves are longitudinal compression waves in a medium (air) that they require to be propagated (sound does not propagate in a vacuum).

In the relativistic expression of the Doppler-Fizeau effect, the wavelength observed is given as a function of the wavelength of the source by the equation:

$$\lambda_{\text{obs}} = \lambda_{\text{source}} \sqrt{\frac{1 + V_r/c}{1 - V_r/c}} \quad (2.17)$$

where V_r is the radial velocity of the source, positive when the object is receding, and c is the velocity of light in vacuum. From this expression we can derive the relative shift in wavelength (or frequency) as:

$$\frac{\Delta\lambda}{\lambda} = \frac{V_r}{c} = -\frac{\Delta\nu}{\nu} \quad (2.18)$$

So, for a hot Jupiter ($V_r \approx 50 \text{ ms}^{-1}$), the relative shift in wavelength is about 1.5×10^{-7} .

Measurement of the radial velocity of a star is therefore carried out by high-resolution spectroscopy. The spectrograph resolves the radiation from the star into its different components (corresponding to the ‘notes’ in our analogy with sound). Measurement of the shift in wavelength is made by observing the overall shift in the heavy-element absorption lines in the spectrum. Given the low amplitude of the shift ($\Delta\lambda/\lambda = 5 \times 10^{-7}$ – 10^{-8} , depending on the object), two conditions are required to obtain sufficient accuracy:

- using numerous lines to analyze the shift
- having a wavelength reference to calibrate the spectrum and obtain an absolute measurement of the radial velocity, allowing comparison of results obtained over several years (which is the time required to detect long-period planets).

The first point results in constraints on the spectral resolution of the spectrograph (the value $\lambda/\Delta\lambda$ where $\Delta\lambda$ is the spectral range covered by one element of the spectrum). The spectral resolution should be several tens of thousands for the best

¹ Fizeau’s name should not be dissociated from that of Doppler. In fact, although we owe the observation of the effect to the latter, the true understanding of the phenomena in terms of the actual physics is the work of Fizeau.

modern instruments. There is also a constraint on the choice of targets. Stars that are too young, or are too hot, are difficult to observe (or have a reduced accuracy of measurement), because their spectra do not exhibit sufficient narrow lines for the amount of shift to be measured. The second point requires the use of a reference source, generally a spectral lamp (thorium–argon), or a cell containing a gas, whose absorption spectrum is known extremely accurately (generally a cell containing iodine vapour is used). In either of the solutions, the reference standard should be essentially stable, and thus requires a thermal environment, which itself needs to be stabilized. Measurements are made by exposing the spectrograph's detector to the stellar spectrum and the reference spectrum simultaneously. The best current instruments (such as HARPS on the 3.6-m telescope at La Silla) allow us to obtain a radial velocity to an accuracy better than 1 ms^{-1} , and stable over a period of several years. The principal instruments used for the measurement of radial velocities and their method of operation are described later (Chap. 8).

2.2.1.3 The Timing of Pulsars

The detection of radial motion (along the line of sight for an observer on Earth) in a celestial body that is orbited by one or more planets may prove to be simplified if the body periodically emits a signal – electromagnetic waves, for example.

This is particularly the case with pulsars, which are neutron stars that have resulted from the explosion of a supernova, and which have the specific feature of emitting electromagnetic waves in a cone, which sweeps the sky in time with the rapid rotation of the star (with periods between a few milliseconds to several seconds), rather like the beam from a lighthouse. Some of these pulsars are visible from Earth, because the emission cone passes across the Earth at each rotation. They are then detectable as a periodic signal, which is easy to time. If the pulsar exists on its own, the Earth–pulsar distance does not vary and the period of the signal is absolutely constant. If the pulsar has planets, just like the other stars we have been discussing, then it will revolve around the centre of mass of the system, so that during the course of its revolution, the Earth–pulsar distance will increase and decrease, thus increasing or decreasing the travel time of the pulsar's signal. This variation in the travel distance, and thus in the travel time, is indicated by a variation in the period of the pulsar over the course of time.

Assuming, as previously, that the pulsar describes an orbit with a semi-major axis a_* about the centre of mass, inclined at angle i relative to the plane of the sky (where $i = 0$ when the orbit lies in the plane of the sky), the variation in the pulsar's period is given by:

$$\delta T_* = \frac{a_* \sin(i)}{c} \quad (2.19)$$

where c is the propagation velocity of the wave from the pulsar in a vacuum (i.e., the velocity of light).

Table 2.4 Amplitude of the period variations of a pulsar of one solar mass as a function of the planets in orbit

	$a_*(r_\odot)$	δT_* (ms)
Jupiter	1.07	250
Saturn	0.59	137
Uranus	0.18	42
Neptune	0.33	76
Earth	6.5×10^{-4}	0.15

Just as in the case of the variations in the radial velocity, these variations of the pulsar's period have a period of their own, which is set by the orbital characteristics of the body orbiting the pulsar.

Table 2.4 gives the value of δT_* for a pulsar of one solar mass accompanied by the bodies in our Solar System at their respective distances.

The accuracy of measurements of the variations in a pulsar's period caused by the presence of one or more companions increases with the intrinsic stability of the pulsar. This is the case with the old millisecond pulsars with a stable internal structure,² but which are revitalized by the accretion of material from a neighbouring red giant (Davis et al., 1985). So this applies to a small number of objects. Currently, the lowest-mass companions that are detectable are found in these systems, which we are able to search thanks to the stability of these pulsars.

The chronometric accuracy currently attained with these objects is far better than a millisecond (which is more than enough to detect a planet, even a terrestrial one). The method has the potential to detect planets with the mass of the Moon. In addition, the simplicity and accuracy of the method explain why it was the first to provide results. The first extrasolar planets detected, in 1992, were those around PSR 1257+12 (Wolszczan and Frail, 1992). But the discovery did not attract a lot of attention: the electromagnetic environment of the pulsar is, on the face of it, very hostile, and so leaves little hope that these planets could be habitable.

2.2.2 *The Effect a Planet has on Photometry of Its Star*

We have seen in the preceding section that the presence of one or more planets, orbiting a star, may be revealed indirectly by observing the movement of the star that is caused by the planet. In certain cases, we may equally hope to detect the presence of a planet by the effects produced on the luminosity of the star itself. Two principal effects may be involved: the passage of a planet in front of the star (a transit), or gravitationally induced amplification of the brightness of a background star by a multiple target.

² Unlike young pulsars, where deformation and changes in structure lead to irregularities in the period.

2.2.2.1 Planetary Transits

A planetary system observed edge-on (where the plane of the orbit is perpendicular to the plane of the sky), is a specific, and very interesting, case. The planets may, in fact, transit the star and cause micro-eclipses, which are detectable by continuously measuring the star's flux (Fig. 2.5). The amplitude of the extinction enables us to obtain the diameter of the planet, to a close approximation, and the duration of the phenomenon, and its periodicity, enables us to derive the planet's orbital period and hence its distance from the star.

Observation of a planetary transit is possible **ONLY** if the system is seen very close to edge-on. To calculate the likely probability of seeing a transit, consider Fig. 2.6

The transit will be visible only if the line of sight intercepts the cylinder, constructed on the orbit, of radius a_p and height $2r_*$. Assuming that the orientation of planetary systems with respect to the plane of the sky is essentially uniform, the probability p_T of actually observing a transit is thus expressed as the ratio between the surface of the cylinder that we have just described (the total number of favourable orientations of the line of sight) to the sphere of radius a_p (the total number of possible orientations of the line of sight.) For a circular orbit, the following relationship may be derived:

$$p_T = \frac{2\pi a_p \cdot 2r_*}{4\pi a_p^2} = \frac{r_*}{a_p} \quad (2.20)$$

where r_* is the radius of the star, and a_p is the semi-major axis of the planet's orbit. This relationship may equally be expressed as a function of the planet's orbital period P , by the application of Kepler's Third Law, and becomes:

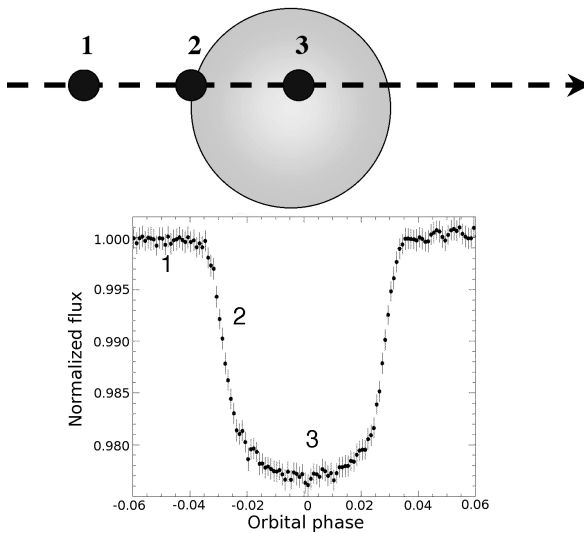


Fig. 2.5 Diagram illustrating the principles of a planetary transit in front of a star, and the associated photometric light-curve

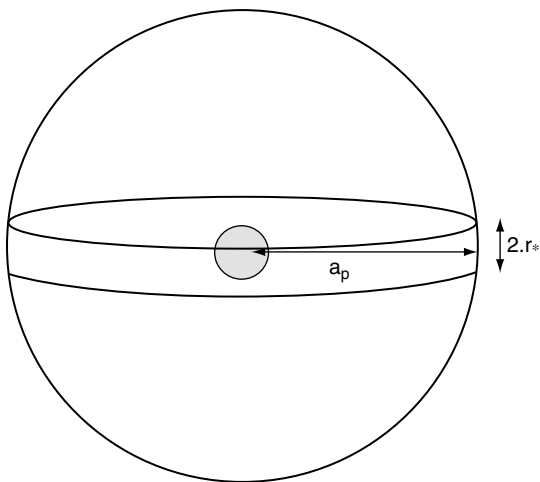


Fig. 2.6 The geometry for observing a transit

$$p_T = \frac{r_*}{p^{2/3}} \cdot \left(\frac{4\pi^2}{Gm_*} \right)^{1/3} \quad (2.21)$$

This probability is shown in Fig. 2.7, for a star that has a radius equal to the solar radius.

It will be seen that the probability decreases rapidly with the planet's orbital distance (and thus period). Detection of planets by observing their transits of a star will, therefore, independently of any considerations of detectability, be more favourable for planets at short distances from their parent star. The probability of observing a transit of a hot Jupiter is thus about 10 per cent (with a period of 3–4 days), whereas the probability of observing the transit of a planet such as the Earth (with a period of 365 days) is only about 0.5 per cent.

To calculate the duration of a transit, we will make the simplifying assumption that the orbit is circular, and consider the geometry shown in Fig. 2.8.

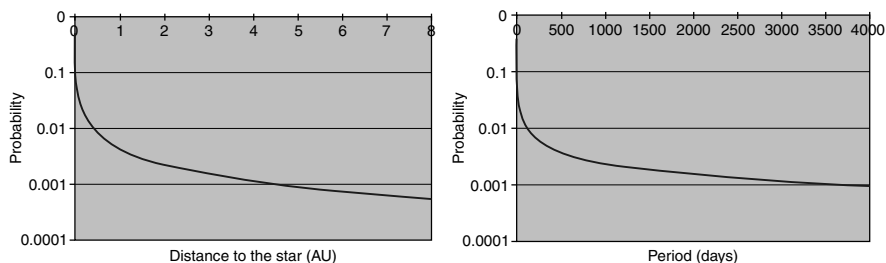


Fig. 2.7 Probability of a transit by a planet orbiting a solar-type star as a function of its distance from the star (*left*) and its orbital period (*right*)

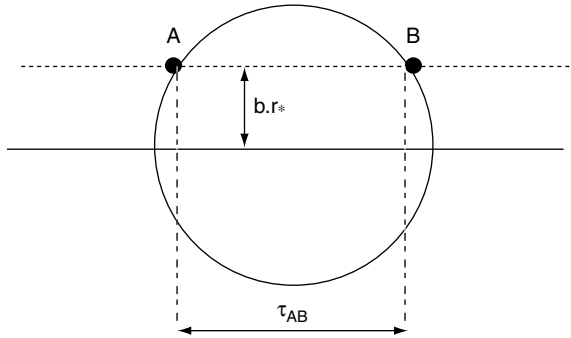


Fig. 2.8 Geometry of a transit

The duration of a transit corresponds to the time required for the planet to traverse the chord AB (Fig. 2.9). The impact parameter (denoted b and not p , to avoid confusion with notation used previously for the orbital period and the transit probability), is determined by the inclination of the orbit, i , relative to the line of sight and the relationship:

$$b.r_* = a_p \cdot \cos(i) \quad (2.22)$$

The length of the chord AB is:

$$l = 2.r_* \sqrt{1 - b^2} \quad (2.23)$$

The time taken to traverse it, for a circular orbit of period P , at velocity $v = 2.\pi.a_p/P$. From this we may derive the duration of the transit: $\tau_{AB} = l/v$, which may be written, replacing P by its expression as a function of a_p , using Kepler's Third Law, as:

$$\tau_{AB} = \frac{2.r_* \sqrt{1 - b^2}}{G.m_*^{1/2}} . a_p^{1/2} = \frac{(2\pi)^{2/3} . 2.r_* \sqrt{1 - b^2}}{(G.m_*)^{1/3}} . P^{1/3} \quad (2.24)$$

These equations may be written more simply, by taking a_p in astronomical units (the mean Earth–Sun distance), P in days, and m_* and R_* in solar units, to give:

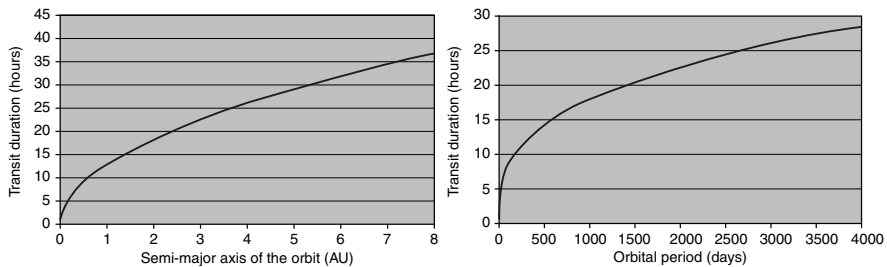


Fig. 2.9 Duration of a planetary transit (in hours) as a function of the orbit's semi-major axis (*left*) and of the orbital period (*right*)

$$\tau_{AB} = 13.0 \sqrt{1 - b^2} \cdot \frac{r_*}{m_*^{1/2}} \cdot a_p^{1/2} = 1.8 \sqrt{1 - b^2} \frac{r_*}{m_*^{1/3}} \cdot P^{1/3} \quad (2.25)$$

The relative amplitude of the extinction during a transit is, to a first approximation, equal to the ratio of the apparent surfaces of the planet and the star, and may be written:

$$\frac{\Delta F}{F} = \frac{r_p^2}{r_*^2} \quad (2.26)$$

So the transit of a giant planet like Jupiter causes a photometric extinction of 1 per cent, and that of a terrestrial planet like Earth an extinction of 0.01 per cent (10^{-4}). To detect a planet by the transit method, it is therefore necessary to be able to carry out photometry of the star to better than the relative extinction. So, to detect a giant planet, a photometric accuracy of roughly a fraction of one per cent is required, while for a terrestrial planet, the photometric accuracy needs to be around some 10^{-5} .

Although the detection of giant planets is possible from the ground, where one may achieve accuracies of around 10^{-3} , limited by atmospheric turbulence, the detection of terrestrial-type planets may only be contemplated from space. The difference in the accuracy of an observation made from the ground and one made from space is illustrated in Fig. 2.10, which shows the same transit, that of a giant planet passing in front of HD 209458, observed from the ground and from space.

The method's potential is not limited to just the identification of planetary candidates. In fact, when combined with the radial-velocity method (where the indeterminacy of the angle of inclination relative to the plane of the sky is removed by the observation of a transit), detection of a transit enables us to derive:

- the size of the object (the amplitude of the extinction)
- its mass (by measurement of the radial velocities)
- and, as a result, the density of the object, thus allowing us to distinguish between a gaseous planet and a terrestrial-type planet.

In some cases, when the observation may be made with a good signal-to-noise ratio, by using differential spectroscopy before and during the transit, it is possible to determine the composition (or at least the presence of certain elements) in the planet's atmosphere. This point will be discussed in Chap. 7. We should also mention that the transit method has been used in the infrared, exploiting the fact that the planet is eclipsed when it passes behind the star (known as a secondary transit). Accurate knowledge of the ephemerides of the system allows us, by subtraction of the flux during the secondary transit (star alone) from that before the secondary transit (planet + star), to derive some spectral information about the object. To date, this method has been applied to two objects: TrEs1 and HD 209459, using the Spitzer space telescope. It set limits on the flux emitted by the objects in four spectral bands (see Chap. 7). It should become more generally available when the James Webb Space Telescope (JWST) enters service.

At present there are more than twenty, ground-based projects for the systematic observation of transits, generally with small, automated, wide-field telescopes (see

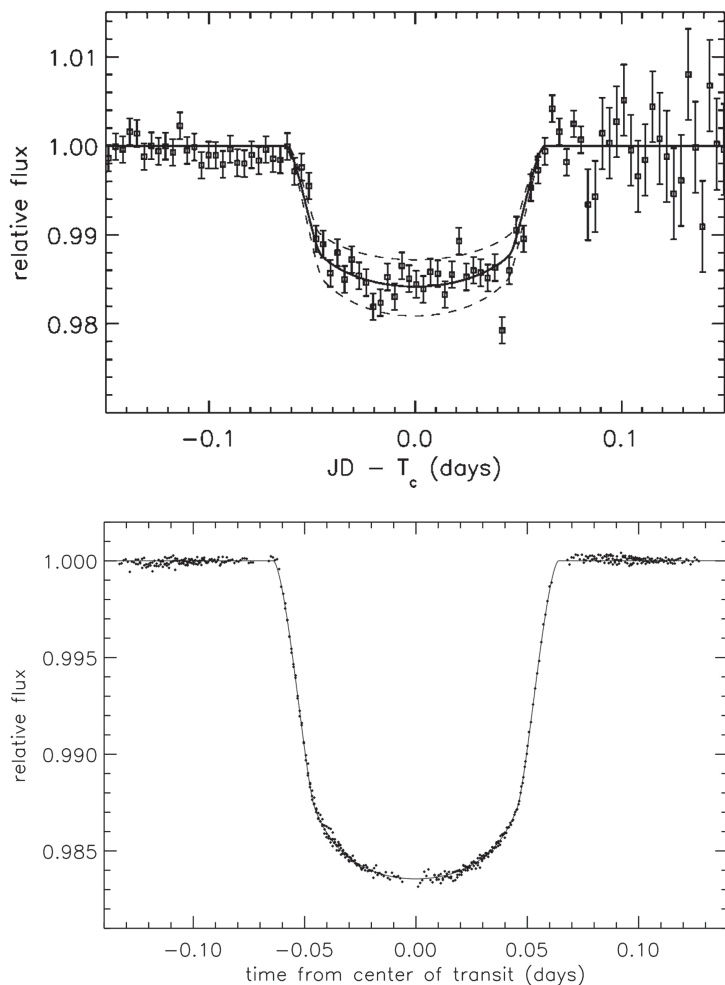


Fig. 2.10 A transit of HD 209458b observed (*left*) from the ground (Charbonneau et al., 2000) and (*right*) from space with the HST (Brown et al., 2001)

Chap. 8). In this case, the primary difficulties are the handling of a large quantity of data and the extraction of relative photometry of the objects, when the observations have been obtained under what are sometimes extremely varied conditions (in particular, the influence of local meteorology and the succession of day and night, which do not guarantee continuity of observation). The A-STEP project, with an observing programme from Dome-C in Antarctica is an extreme example of these ground-based observational programmes. The excellent atmospheric conditions associated with the long Antarctic night appear to be serious advantages for this programme. Nevertheless, the ultimate photometric accuracy of all these ground-based programmes appears to restrict the method to the detection of giant planets.

Two space missions designed to search for planets by the transit method are envisaged for the near future, and are discussed in detail in Chap. 8:

- the CoRoT mission, led by CNES, launched in late 2006, began operation in January 2007 and should be able to identify objects of a few terrestrial radii, orbiting close (with periods of less than 75 days) to their star;
- NASA's KEPLER mission should, from 2008, be able to detect terrestrial planets with periods up to about one Earth year.

2.2.2.2 Gravitational Microlensing

One of the most surprising aspects of Einstein's theory of relativity is the phenomenon of the gravitational lens (Einstein, 1936). In his theory, Einstein introduced energy/matter equivalence,³ with the consequence that the photon, the quantum of electromagnetic energy, is subject to gravitation, just like 'classical' baryonic matter, whose weight we all experience.

So a photon that passes at distance r from an object of mass M , which is assumed to be a point, undergoes a deflection α relative to its direction of propagation, and this is given by the following equation:

$$\alpha = \frac{4GM}{c^2 r} = \frac{2R_s}{r} \quad (2.27)$$

where R_s is known as the 'Schwarzschild radius' of the gravitational lens caused by the point mass.

When a massive object lies between the observer and the object being observed, the image of the latter is thus deformed by the mass of the deflecting object. The amplitude of the phenomenon depends on the mass and position of the deflector. When the deflecting object is sufficiently massive – in which case α is greater than the resolution of the instrument being used for the observation – the effect is known as a 'macrolens': the image of the object is multiplied into an odd number of secondary images. Historically, this was the first type of lens to be observed. The image of the double quasar Q0957+561 obtained by Huchra in 1985 revealed a symmetrical multiple structure related to the presence of a galaxy on the line of sight (Huchra, 1985). When the mass of the deflecting object is low and α is less than the angular size of the observing instrument's diffraction disk, the situation is described as a 'microlens' and multiplication of the image is not observed (although it still exists), but the mean lensing effect instead, which takes the form of an overall amplification of the intensity of the source being observed. The first observation of a gravitational microlensing event was that of the quasar Q2237+0305 (Racine, 1992).

³ This is the famous equation $E = mc^2$, where E is the rest energy of the particle, m its mass, and c the velocity of light. In the case of the photon, the rest energy is $h\nu$, where h is Planck's constant and ν the frequency of the wave associated with the radiation.

In a completely similar manner, if a massive object passes across the line of sight to a star, the latter's brightness is increased by the effect just described, simultaneously revealing the presence of the object that serves as a lens (Fig. 2.11). The gravitational amplification effect may, therefore, allow the transient observation of an object that cannot be detected prior to the microlensing event.

Gravitational amplification is given by the following relationship (Sackett, 1999):

$$A = \frac{u^2 + 2}{u\sqrt{u^2 + 4}} \quad (2.28)$$

where $u = \theta_S/\theta_E$, with θ_S the angular distance between the source and the deflector, and θ_E the Einstein radius, which is itself defined as:

$$\theta_E = \left(\frac{4GM}{c^2} \cdot \frac{d_{LS}}{d_L \cdot d_S} \right)^{1/2} \quad (2.29)$$

where G is the gravitational constant, M the mass of the lens, c the velocity of light, d_L the distance between the observer and the lens, d_S the distance between the observer and the source, and d_{SL} the distance between the source and the lens.

The variation in the brightness of the object during the passage of the lens is shown in Fig. 2.12.

In the case of a simple lens, the variation in luminosity is symmetrical, and centred on the position of the lensing star. The overall set of the apparent positions of

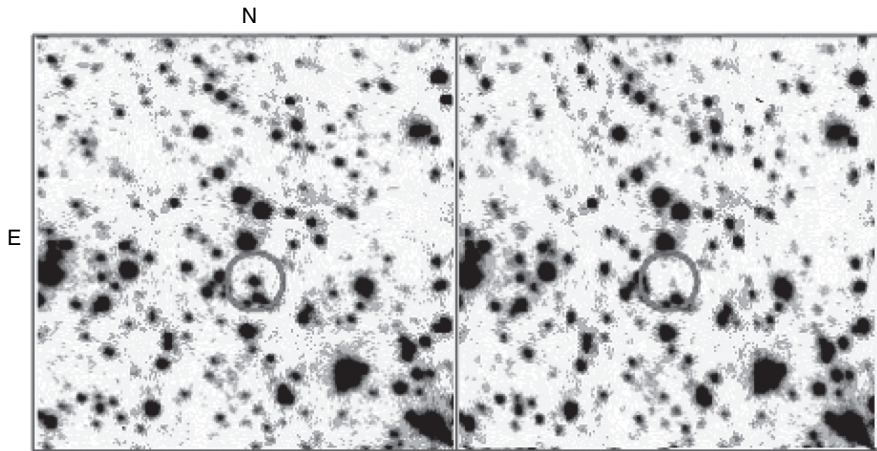


Fig. 2.11 The gravitationally amplified event 98-SMC-01 detected by the MACHO team, and which had a gravitational amplification factor of about 100. These images show the object during the amplification phase (*left*) and then following the amplification and the passage of the lensing body (*right*). The field size is about 70×70 arcsec. This particular event has been attributed to a low-mass binary deflector in the Small Magellanic Cloud

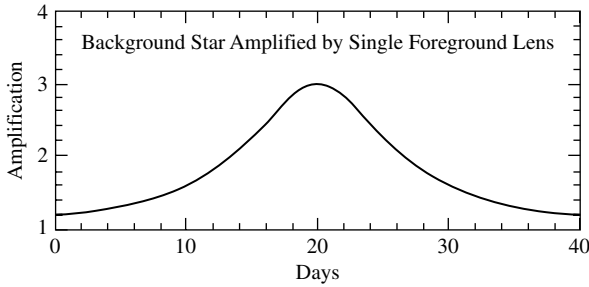


Fig. 2.12 The theoretical gravitational amplification curve as a function of time, for a simple lens

the lens where the flux of the star is amplified by a factor of 1.34, is a circle, whose radius is precisely the Einstein radius,⁴ as defined by Eq. (2.29).

This method has been used since 1995 by various groups (EROS, MACHOS, OGLE) to try to detect dark matter in the Galaxy. The theory to be verified is that dark matter consists of massive, compact, and non-luminous objects, whose overall mass could explain the difference between the mass deduced from the dynamics of the Galaxy, and that which is directly observable (stars and gas).⁵

For planet hunters, the interesting case is when the deflector is accompanied by one or more planets (a multiple lensing event). The structure of the amplification zone is then no longer symmetrical. Because of the presence of planets, a line appears between the star and the planet, known as a caustic, where the gravitational amplification is theoretically infinite for point sources such as stars. In the case of a simple lens (without a planet), the only point of infinite gravitational amplification is the position of the lens itself, i.e., when the source, lens, and observer are perfectly aligned. In the case of a multiple lens, the apparent intensity of the source undergoes significant variations when the apparent path of the source relative to the lens, approaches or crosses the caustic.⁶ It is the structure of the gravitational-amplification curve (Fig. 2.13) and, more especially, the study of its artefacts (number, duration, and intensity) that enables us to determine the characteristics of the planet, or planets (mainly the mass and projected angular distance), by use of a model that incorporates a significant number of parameters that may be adjusted to the data. The error bars associated with this method are rather large.

⁴ The Einstein radius is also the radius of the circle that is the image of the source produced by the deflecting object, when the source, the deflector, and the observer are perfectly aligned.

⁵ Calculations actually reveal an enormous mass-deficit (a problem which is known in cosmology as the ‘missing-mass problem’). A possible solution to this problem was to consider the existence of a vast number of brown dwarfs, which were not detectable by direct observation. This explanation, however, does not appear to be a convincing solution.

⁶ The most spectacular effects occur when the apparent path of the source star crosses the caustic, giving a gravitational-amplification curve that is extremely variable, depending on the exact case. The most frequent case is that when the path approaches the caustic, but does not cross it. In such a case, we observe anomalies in the characteristic Gaussian curve of a simple lens.

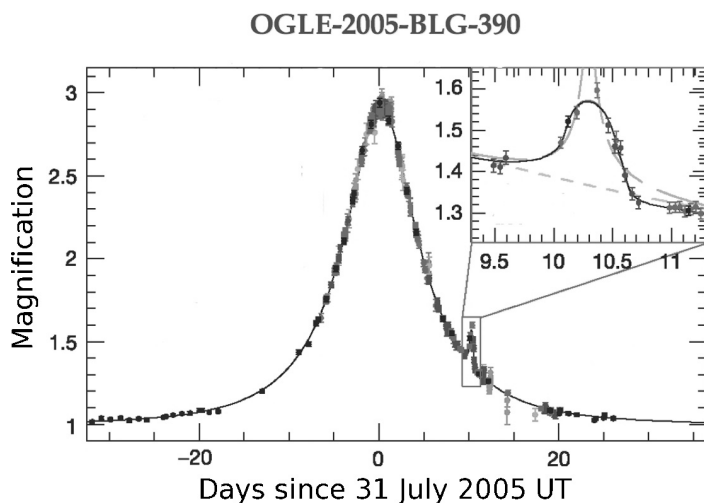


Fig. 2.13 The trace of the gravitational-amplification event OB-05-390, where the artefact indicates the presence around the deflecting star of an object of $5.8 M_{\oplus}$ (After Beaulieu et al., 2006)

This method is, in principle, extremely sensitive, and enables giant planets to be detected – for example, the candidate associated with the MACHO-97-BLG-41 event, with a mass estimated as about three Jupiter masses (Bennett et al., 1999). It also allows lower-mass planets (terrestrial planets) to be detected. It was using this method that, in 2005, Beaulieu et al. detected the object OB-05-390 b, which is currently the lightest planet ($5.8 M_{\oplus}$ to within a factor of 2) ever detected orbiting a star other than a pulsar (Beaulieu et al., 2005). The difficulty in this case occurs in monitoring the gravitational-amplification event. The teams that are carrying out the search for planets by this method are dependent on other teams who provide details of the gravitational-amplification targets as soon as an event begins. To search for planets, as continuous a curve as possible is required so that it may be examined for artefacts. To obtain adequate sampling (about one point every 30 min) and continuous coverage of the amplification events, a collaborative effort (PLANET) has been devised, using several telescopes around the world to obtain 24-h coverage of any event.

The gravitational microlensing method is most sensitive to objects at a moderate distance (a few AU) from their parent star. It is not suitable for close objects (but there are the transit and radial-velocity methods for that). It allows the preferential detection of objects orbiting cool stars, because the stellar population of lenses is dominated by low-mass (and low-mass) objects. The object OB-05-390 b is orbiting a cold star (of spectral type M) at a distance of 2.6 AU. Given the distance of the objects, the surface temperature of this planet cannot be higher than 50 K, which means that it compares with that of a Pluto-type object in our own Solar System.

At present only the OGLE and MOA teams are continuing to issue the gravitational-amplification alerts that are likely to lead to the discovery of other planets by suitable follow-up observations. The permanence of the method thus depends, in a crucial manner, on these programmes of systematic observations.

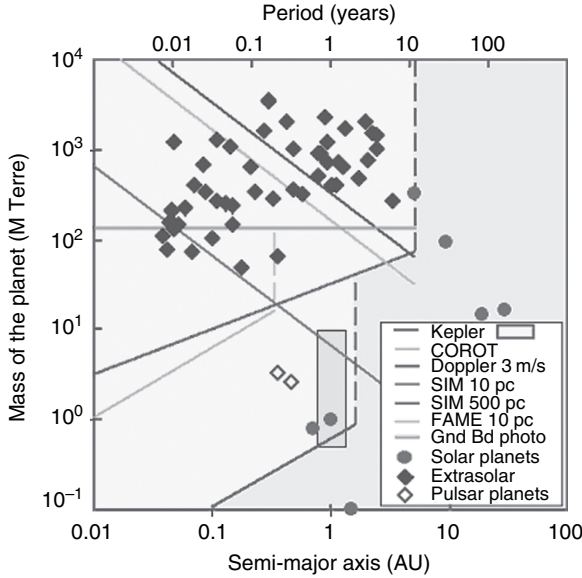


Fig. 2.14 A diagram summarizing the domains over which the various methods of indirect detection are most sensitive

2.2.3 Comparison of the Different Indirect Methods

In the previous sections we have considered the different methods for the indirect detection of extrasolar planets, their sensitivities, as well as their bias, either intrinsic, or linked to the choice of observational sample. Figure 2.14 summarizes the best observational domains for the different methods, with a diagram that plots the mass of the object against its distance from the central star (assuming the star to be of solar type).

2.3 Direct Detection of Exoplanets

In this section we intend to discuss direct detection of an exoplanet, in the sense of being able to detect directly photons arising from the planet, and not the effects of the latter on the central star. The objective of direct detection is to be able to separate

the photons from the star from those from the planet, to enable us to make a spectroscopic analysis and to deduce details of the planet's composition, or at least that of its atmosphere. Strictly speaking, however, the transit method, when carried out in the infrared, is a direct-detection method, in particular when the secondary transit is observed (the passage of the planet behind the star). By measuring the thermal flux before and during this secondary transit, we obtain, by subtraction, a measure of the flux from the planet. This method was used on HD 209458b and TrEs-1 with the help of the Spitzer space observatory, and enabled spectral information to be gathered in four channels between 5 and 25 μm (Charbonneau et al., 2005).

Here, however, we shall restrict discussion to the collection and analysis of photons from planets.

2.3.1 Choice of Spectral Region

The choice of a spectral region for direct observation involves a compromise between the contrast between the star and the planet (typically between 10^4 and 10^{10} , depending on the nature of the companion and the spectral region chosen), and the angular resolution of the instrument, which primarily depends on the size of the collector. The limiting angular resolution is set by diffraction. The image of a point at infinity through a telescope is not a point but a disc, the form and size of which depend on the shape and size of the telescope's aperture. For a circular aperture, the diffraction disc appears as shown in Fig. 2.15.

The radius r of the central diffraction disc is given by:

$$r = 1.22 \frac{\lambda}{D} \quad (2.30)$$

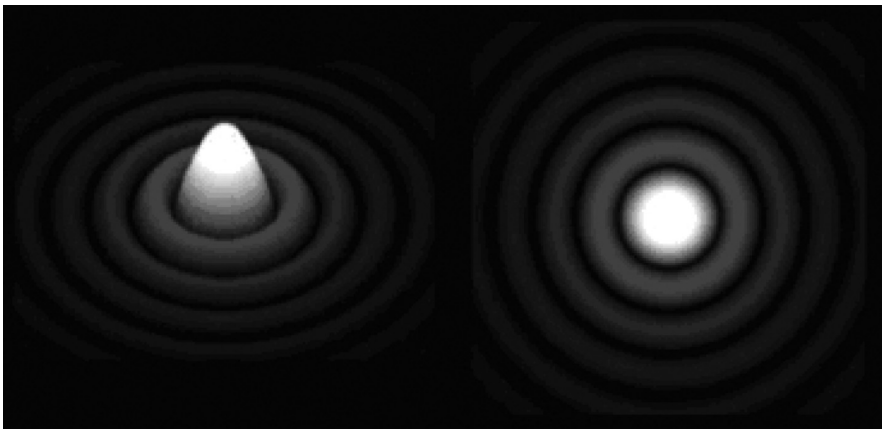


Fig. 2.15 Form of the diffraction disc given by a circular aperture with no central obstruction

where λ is the observational wavelength and D is the diameter of the aperture (the telescope's primary mirror).

To 'image' an exoplanetary system, taking account of its spatial size, the compromise finally comes down to a choice between two domains:

- The visible and near infrared. Here, the star-and-planet pair is spatially resolved by a telescope a few metres in diameter. The difficulty lies in overcoming the contrast, which is 10^9 – 10^{10} for a terrestrial-type planet. In this instance, coronagraphic methods are used, possibly together with adaptive optics.
- The thermal infrared (around $10\mu\text{m}$). Here, the contrast is optimal (10^7 for a terrestrial planet), but we have to resort to interferometry to obtain the angular resolution of the star-and-planet pair, because it would require a monolithic telescope several tens of metres in diameter to resolve the system.

The luminous flux of the objects also needs to be taken into account. This point is particularly important when we want to consider the question of imaging planetary surfaces. Table 2.5 gives an estimate of this flux.

Table 2.5 Flux received from an Earth-like planet at 10 parsecs from the Solar System

Spectral region	Integrated flux received on Earth (omitting atmospheric absorption)
0.3–2 μm	0.3 photon/s/m ²
6–20 μm	10 photons/s/m ²

Let us now discuss the different direct-detection techniques, together with the associated instrumentation.

2.3.2 Coronagraphic Methods and Adaptive Optics

2.3.2.1 A Historical Retrospect: The Lyot Coronagraph

The coronagraph was invented by Bernard Lyot at the beginning of the 1930s to observe the solar corona at times other than eclipses (Lyot, 1931; 1939). The principal approach – which was, nevertheless, empirical – that guided its inventor was to achieve the maximum reduction in the light scattered and diffracted by the optics. The principle of this instrument is shown in Fig. 2.16.

Lyot's trick was to re-image the primary image plane, and to place a diaphragm at the associated pupil plane to eliminate the effects of diffraction caused by the mask located as the primary image focal plane. This method, in conjunction with the use of extremely high-quality optics (with homogenous glass and exceptionally high quality polishing) enabled light scattered within the instrument to be reduced

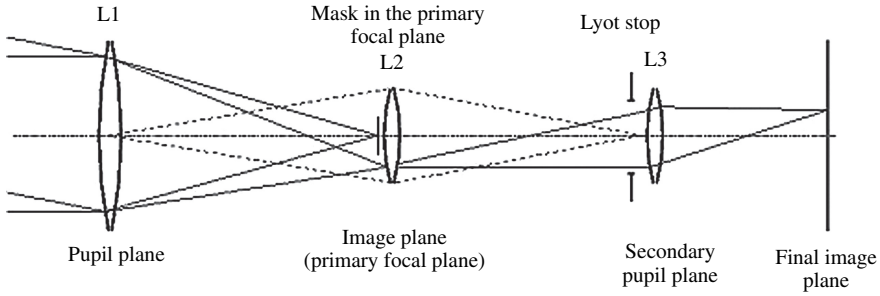


Fig. 2.16 The basic principles of the Lyot coronagraph

by a factor of between 10 and 100. Lyot's coronagraph thus allowed observation of the solar corona a few arcminutes from the solar limb.

An elegant description of how Lyot's coronagraph works may be made using Fourier optics and image formation theories. A good introduction to these topics may be found in Goodman, 1996.

From the point of view of the theory of the formation of images, the principle behind Lyot's coronagraph may be described in the way shown in Fig. 2.17:

- (a) at the entry pupil plane: the infinite (i.e., nominally plane) wavefront is sampled by the pupil

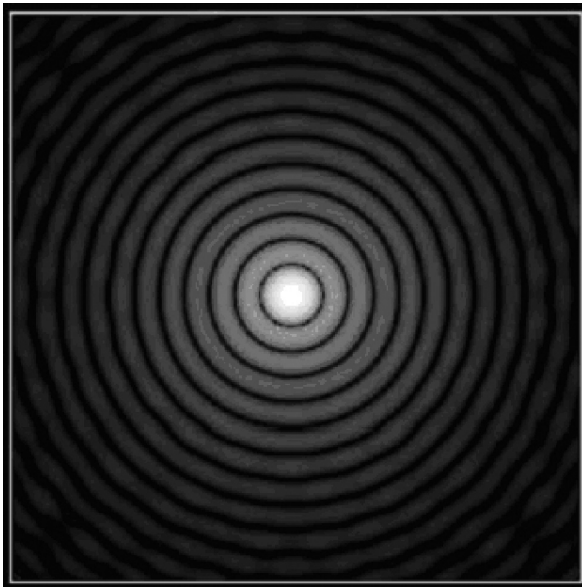


Fig. 2.17 The Airy function (Airy disk): the diffraction pattern produced by a circular diffracting aperture

- (b) at the primary image plane (the focus of the main mirror): The image of a single point at an infinite distance (which is known as the Point-Spread Function, PSF) is the square of the modulus of the pupil's Fourier transform. (The Fourier transform is a mathematical function that allows the transformation of a function into another, known as the frequency domain representation of the original function. For instance, a function of time may be transformed into one of frequency. Among the main characteristics of the Fourier transform is the fact that the operation known as a convolution, whereby two functions are manipulated to provide a third, may be simply represented in Fourier space by a single multiplication of the Fourier transform.) This point-spread function may thus be multiplied by the mask's transmission ($1-\Pi$, where Π is the aperture function, corresponding to a mask with an occulting centre. In this image plane we therefore have:

$$\text{image} = \text{PSF} - \text{PSF} \times \Pi \quad (2.31)$$

- (c) at the secondary pupil plane: the new pupil is once again the Fourier transform of the primary image, in other words: The Fourier transform of the Airy function (i.e., the initial pupil) less the convolution product of the Fourier transform of the Airy function (the initial pupil) with the Fourier transform of the aperture function (another Airy function with a size that is inversely proportional to the size of the mask). It will be recalled that the Airy function is the result of the diffraction of light by a circular aperture (Fig. 2.17). The final transmission of a coronagraph is thus equal to 0 on the optical axis, and there remains only an annulus (of scattered light), which is eliminated by the Lyot stop at the pupil (Fig. 2.18), which may be written as:

$$\text{pupil} = \Pi - \Pi \otimes \text{Airy} \quad (2.32)$$

where the symbol \otimes represents the second magnitude convolution product.

This method, with some variants, has been employed in all solar coronagraphs, including those installed on satellites such as SOHO (Koutchmy, 1988), as well as in certain stellar coronagraphs which, when combined with adaptive-optics systems,

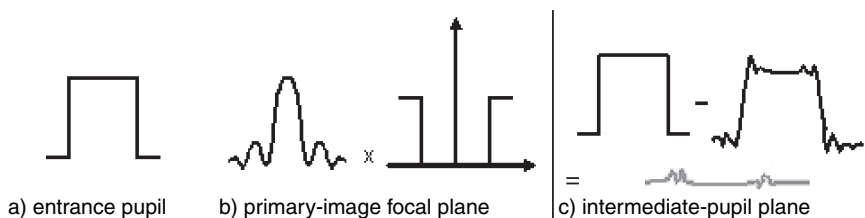


Fig. 2.18 Lyot coronagraph: (a) and (b) wavefront and image at the pupil planes and successive images. (c) in red, residual light in the plane of the secondary pupil, which is a ring of light eliminated by the Lyot stop. All photons from a source on the optical axis are therefore, in theory, completely eliminated

have attained remarkable performance (Malbet, 1996), such as the detection of a companion lying 10 Airy radii from a highly luminous star, which is 10^5 times as bright (Beuzit et al., 1997).

The stellar version of the Lyot coronagraph, even though it may be effective down to 2 arcsec of the star that is being occulted, when used with a telescope in the 4-m class, cannot observe beyond that limit. In fact, the closer to the star that one attempts to observe, the smaller the occulting disc needs to be, and the greater the diffraction that it creates, requiring a stop whose aperture rapidly becomes very small. In practice, the Lyot stop cannot cover less than the star's Airy disk. Yet it is precisely within this region that we want to observe. Short of using a large-diameter telescope which 'super-resolves' the system and eliminates the residual contrast, Lyot's original method cannot be used, as such, to search for terrestrial planets.

2.3.2.2 The Phase-Mask Coronagraph

An alternative to the Lyot coronagraph was suggested by François and Claude Roddier in 1997. It replaces the circular amplitude mask in the focal plane of the Lyot coronagraph with a circular phase mask. (The mask modifies the phase of the wavefront passing through it.) This mask is smaller, covering half (of the area) of the Airy disk, and applies an achromatic phase change (π) or, at least, one that is very weakly chromatic, given the low dispersion in the chosen spectral band. Light from the star is extinguished by destructive interference by virtue of the source's symmetry relative to the mask (which must be precisely located on the image spot), and rejected outside the geometric pupil, where it is blocked by the Lyot stop.

Several variants of a phase mask have been suggested. Among the most promising are the four-quadrant mask suggested by Daniel Rouan (Fig. 2.19). The theoretical performance of such a mask suggests that extinctions of 1010 may be attained, even inside the diffraction spot. The great advantage of this mask geometry is that it is 'spatially' achromatic (unlike Roddier mask, the size of which depends on the wavelength). Here, the efficiency does not depend on the size of the image spot but, uniquely, on the ability for the phase-change π , introduced by two of the four segments of the mask, to be achromatic.

Coronagraphs, whether of the amplitude mask or phase mask type, have requirements in common: the necessity for the stellar source to be perfectly centred on the centre of the mask, and for the stellar image to be as symmetrical as possible. Any deviation from symmetry results in a stellar residue, which may mask the image of the companion that is being sought. It is therefore essential:

- to have the best possible image of the star
- to permanently monitor the telescope's alignment relative to the centre of the star.

This is why all current coronagraphic systems operating from the ground are used in conjunction with adaptive optics.

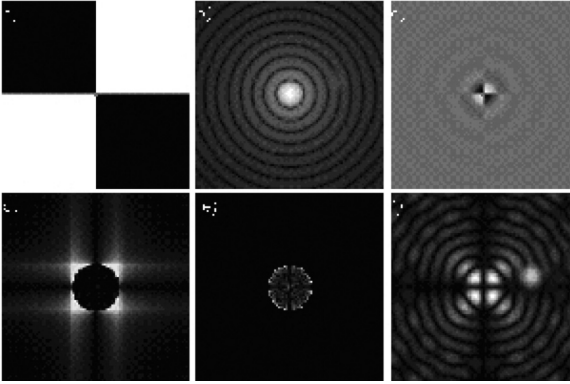


Fig. 2.19 The principle of the 4-quadrant coronagraph: (a) appearance of the phase mask (white: no phase-change, black: phase-change π); (b) image in the primary focal plane; (c) the complex amplitude after passing through the mask; (d) secondary pupil plane before the Lyot spot; (e) secondary image plane; (f) central portion of the secondary image focal plane (After Rouan et al., 2000)

2.3.2.3 The Phase-Induced Amplitude Apodization Coronagraph (PIAAC)

The PIAAC instrument has been proposed by Guyon et al., and is in fact the combination of two optical devices (Guyon et al., 2005):

- a high performance coronagraph
- a pupil ‘remapper’ to adapt the pupil shape of the telescope to the coronagraph.

The PIAAC concept comes from the observation that the diffraction pattern produced by a classical telescope leads to the presence of light (diffraction rings) far from the on-axis direction. Such diffraction rings are classically removed or reduced by ‘apodized’ apertures such as Gaussian apertures that can smoothen the aperture edge effects. The main drawback of such a technique is a drastic reduction of the source flux through modification of the pupil shape and thus its transmission, because of partial occultation of the pupil. Guyon proposes apodizing the pupil by a modification of the intensity distribution within the pupil. This is done by a global wavefront phase distortion, using aspheric mirrors (Guyon, 2003). This method allows apodizing the pupil without changing its overall transmission.

The PIAAC combines a first stage of apodization, which optimizes the shape of the pupil, and a coronagraph (which may be either a phase- or an amplitude-coronagraph). The principle of PIAAC is given in Fig. 2.20.

This concept has been demonstrated in the laboratory and is currently under test on the sky. It could also be adapted to a space-borne mission concept.

2.3.2.4 Adaptive Optics and Accurate Pointing: The Keys to Performance

Because of atmospheric turbulence, the image of a star in a large-diameter telescope is never the theoretical diffraction image as set by the limiting size of the telescope’s

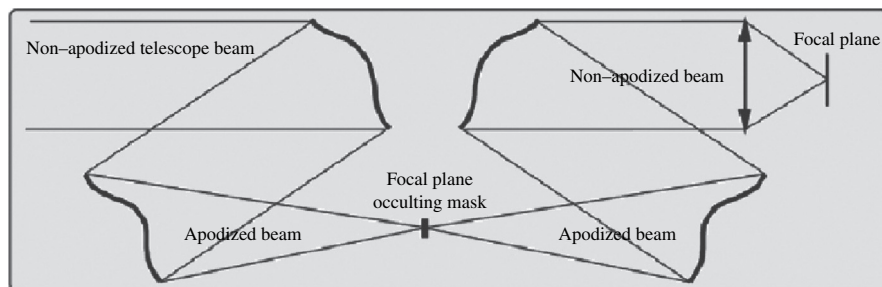


Fig. 2.20 Principle of a Phase-Induced Amplitude Apodization Coronagraph. The beam is apodized before reaching the first focal plane where the coronagraph mask is located. The beam is then apodized backwards to revert to its initial form and to retain the initial phase information contained in the entrance pupil. It is then imaged in the detection focal plane (After Guyon, 2003)

aperture. The star's image is, in fact, the sum of multiple 'speckles', the number and size of which depend on the level of turbulence (which is a characteristic of the site and the construction of the telescope), and of the telescope's size. During a long exposure the speckles give rise to an integrated image that is distinctly spread out, and the size of which is determined by the turbulence. In terms of resolution, the degradation of the image reduces the performance of these telescopes to those of smaller instruments, typically by a few centimetres to some tens of centimetres, depending on the quality of the site and the wavelength under observation. (A thorough introduction to atmospheric optics is given by Léna, 1996). The degradation of the image into speckles results from aberration of the wavefront emitted by the star during its passage through the atmosphere. This aberration affects the phase of the wave emitted by the source. Adaptive optics (Fig. 2.21) is a system that allows:

- the analysis of the aberration of the wavefront caused by its passage through the atmosphere: this is the role of the wavefront sensor. To do this the sensor monitors a bright star within the field of view. This star serves as a point-source reference.
- the partial compensation of the degradation by means of one or more deformable mirrors, which, acting locally, will add or subtract a phase element to the wavefront in such a fashion that it will restore the wavefront's plane form that it had before passing through the atmosphere. These mirrors are actuated by a control system which uses, in real time, the information from the wavefront-sensor to calculate the correction that needs to be applied by means of the deformable mirrors.

Current 'mono-conjugate' systems correct atmospheric turbulence by assuming that it is limited to a specific atmospheric layer (and have a sole adaptive mirror). This solution proves to be inadequate as soon as one wishes to image the field around the reference object. Studies are under way to realize 'multi-conjugate' systems, which take into account turbulence at several different altitudes in the atmosphere.

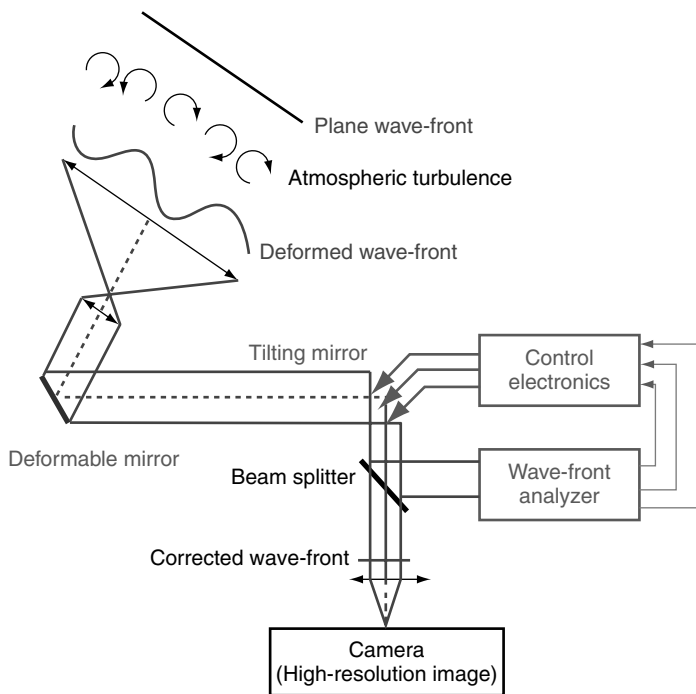


Fig. 2.21 Schematic diagram of an adaptive-optics system. The wavefront that has been deformed by the atmosphere is analyzed, and the effects of turbulence are corrected by use of a deformable mirror (by courtesy of ONERA)

Several parameters may be used to quantify the efficiency of an adaptive-optics system. Among these, we may mention the Strehl ratio, which is defined as the ratio between the peak intensity of the image of a star (i.e., the point-spread function) and the theoretical peak intensity for an image solely limited by diffraction. A Strehl ratio of 1 corresponds to an image that is fully coherent and perfect. An image that is not corrected by adaptive optics has a low Strehl ratio, which depends on the severity of atmospheric turbulence. Current, fine adaptive-optics systems are able to attain a Strehl ratio of 0.6. Future projects for the detection and investigation of extrasolar planets from the ground will require Strehl ratios of at least 0.9.

Use of a coronagraph requires a high-quality wavefront, and which it is therefore essential to correct by adaptive optics when observations are made from the ground. In the specific case of a phase-mask coronagraph, the guiding constraints (on positioning and on centring the mask) are extremely tight (the functioning of these systems is based on the symmetry of the stellar image relative to the phase mask). These constraints are met by a fast-response, tip-tilt mirror ahead of a high-quality adaptive system which stabilizes the position of the image at the image plane.

Most 8-metre-class telescopes are now fitted with adaptive-optics systems. Some also incorporate new-generation phase-mask coronagraphs. We may mention in

particular the NAOS-CONICA instrument on the VLTI. This instrument allowed Chauvin and his collaborators to obtain the first image of an exosystem (2M1207). In a few years' time, this instrument should be followed by the SPHERE instrument, which consists of a highly efficient adaptive-optics system coupled with a 4-quadrant coronagraph (see Chap. 8). These innovations, mainly developed in Europe, and more particularly in France, should, in due course, also be incorporated into American instruments.

2.3.3 Interferometry

As we have already seen in the section on astrometry, it is possible to measure the position and the proper motion of a star relative to a fixed reference by using interferometry. In that method, the position of the group of interference fringes and the length of the baseline are measured to derive the astrometric information. In this section we will show how analysis of the structure of the interference fringes allows us to extract information about the object itself and its environment. To understand this technique, one needs to understand some additional principles of interferometry.

2.3.3.1 Some Principles of Optical Interferometry

The Temporal-Spatial Coherence of Two Waves

When two sources of illumination are mutually coherent, superimposing the two waves produces a phenomenon known as 'interference'. Here, the intensity is not everywhere the sum of the intensities of the two beams taken separately, but is the result of an interference function, which we shall explain later.

Mathematically, the concept of temporal-spatial coherence is expressed by a quantity known as the 'complex degree of coherence', denoted $\gamma_{12}(\tau)$, and which is expressed as the equation:

$$\gamma_{12}(\tau) = \frac{\langle E(P_1, t + \tau) E^*(P_2, t) \rangle}{\left[\langle |E(P_1, t)|^2 \rangle \langle |E(P_2, t)|^2 \rangle \right]^{1/2}} \quad (2.33)$$

In this equation, $E(P, t)$ is the electric field (complex notation) at point P at time t ; $\langle E(P, t) \rangle$ is the temporal mean of this field, and τ represents the delay of wave 1 relative to wave 2. $\gamma_{12}(\tau)$ is a complex number where the modulus lies between 0 (no coherence) and 1 (complete coherence). In this expression, if $P_1 = P_2$ and $\tau \neq 0$, we are dealing with the wave's spatial coherence, and if $\tau = 0$, and $P_1 \neq P_2$, we are investigating the wave's spatial coherence.

Interferometry and Visibility

In observational astrophysics, an interferometer (Fig. 2.22) is, above all, an instrument that measures the degree of temporal-spatial coherence of the electromagnetic field emitted by a source and sampled by two telescopes.

Positions P_1 and P_2 are those of the two telescopes, the delay τ is obtained by modifying the observational configuration (the position of the object on the sky and the setting of the delay line).

If I_1 and I_2 are the intensities in the two arms of the interferometer and if $d = c \cdot \tau$ is the path difference, then the intensity at the exit from the beam combiner may be written as:

$$I(d) = I_1 + I_2 + 2\sqrt{I_1 I_2} \text{Re}(\gamma_{12}(d/c)) \quad (2.34)$$

Measurement of $I(d)$ allows us to measure the complex degree of coherence (or at least its real part) for each of the positions P_1 and P_2 . In the case of two monochromatic plane waves of wavelength λ , spatially and temporally coherent, the preceding equation becomes:

$$I(d) = I_1 + I_2 + 2\sqrt{I_1 I_2} \cdot \cos\left(\frac{2\pi d}{\lambda}\right) \quad (2.35)$$

This equation is interferometry's characteristic equation.

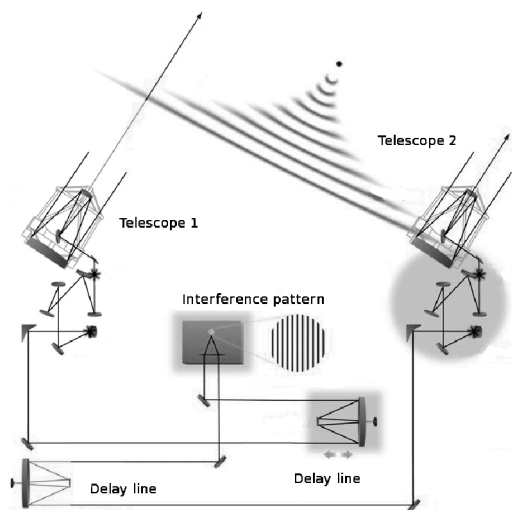


Fig. 2.22 A schematic diagram illustrating the principle of a twin-telescope interferometer of the VLTI type. The rays of light from the two telescopes are led to the beam combiner by a series of mirrors forming the optical train. The path difference is adjusted by using a delay line. (After ESO Internet site: www.eso.org.)

In practice, one records all or part of an ‘interferogram’ (Fig. 2.23), obtained by fixing positions P_1 and P_2 and varying τ (by means of a delay line).

From the graph of $I(d)$ we may determine a quantity known as the ‘visibility’, denoted V , and which is defined by:

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (2.36)$$

In the absence of instrumental perturbation, the visibility is a measure of the degree of mutual coherence. In practice, the experimental visibility (as measured) is the product of three terms:

- the intrinsic visibility of the object
- a term linked to the instrument function of the interferometer
- a term linked to atmospheric perturbations (for observation from the ground) or to the environment.

The last two terms are estimated by observing a ‘calibration’ object, which is a star whose intrinsic visibility is known, and generally a star that is not resolved with the interferometric baseline, and which is thus completely coherent (having an intrinsic visibility of 1).

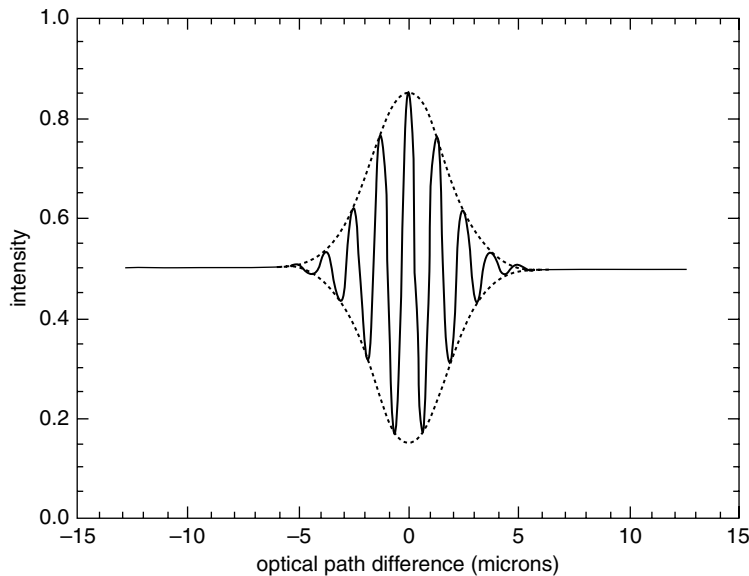


Fig. 2.23 An example of an interferogram obtained by varying the path difference by use of a delay line. For a path difference greater than $\pm 5 \mu\text{m}$, there is no temporal coherence. The spatial coherence is not perfect (the visibility is about 0.7)

Various studies may be found in the literature of the factors that degrade visibility, in particular, atmospheric turbulence. We shall not, therefore, discuss this aspect in detail here.

Visibility of the Source's Structure

The relationship between what may be observed (the experimental visibility) and the astrophysical source is obtained by employing the Zernicke-Van Cittert theorem, which may be expressed as follows:

If both the linear dimensions of the source of quasi-monochromatic radiation and the distance between the two points on the screen (here, the distance between the two telescopes, i.e., the distance between P_1 and P_2) are small relative to the distance between the source and the screen (here, between the source and the Earth), then the modulus of the complex degree of coherence (the experimental visibility) is equal to the modulus of the spatial Fourier transform of the intensity of the source, normalized to the total intensity of the source.

In other words, there is a simple Fourier relationship between the spatial structure of the source and the visibility function, measured at several different baselines (several spatial frequencies) as shown in Fig. 2.24.

In practice, the opposite problem has to be tackled. The visibility curve is obtained, and then, generally by adapting the model, one can deduce the structure and parameters of the source. For example, in the case of a stellar source, the simplest model is one consisting of an evenly illuminated disk. In this case, interpolation from the measured visibility points enables one to determine the diameter of the source (by determining the point at which the visibility function first becomes zero). Obviously, this requires the source to be resolved by the interferometer; that is that the object's angular resolution should be greater than the angular resolution determined by the interferometer's baseline (the value λ/B , where B is the distance between the telescopes and λ is the observational wavelength).

It should, however, be noted that the object space (the plane of the sky, where the coordinates are denoted by x and y), as well as the associated Fourier space (i.e., the one dealing with spatial frequencies and where the conjugate coordinates in x and y are denoted by u and v , described as the (u, v) plane), are two-dimensional

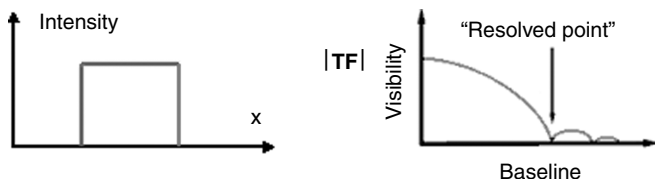


Fig. 2.24 The relationship between an object's spatial structure and the measured visibility function. An interferometer allows the visibility to be measured as a function of the length of the baseline. By adjusting one's model to fit the experimental points, one can then obtain the structure of the source by an inverse Fourier transform

spaces that are being sampled point by point. To reconstruct a correct representation of the astronomical object, it is essential, if the second dimension is to be properly covered, to have several observational baselines (several separations between the telescopes) and, equally, several different orientations of these baselines relative to the plane of the sky. In practice, it is generally the rotation of the Earth that alters the orientation of the fixed baselines, as, in particular, it is with the VLTI. In the case of a space interferometer, several different distances between the telescopes need to be employed, as well as several different orientations of those baselines relative to the sky. (It needs to be possible to turn the interferometer or, at the very least, to change its orientation relative to the line of sight to the target.)

One of the difficulties of interferometry by measuring visibility (apart, as we have seen, from measuring the actual visibilities accurately), is the necessity for having precise models that allow for the study of the visibility and its variations, in particular as a function of the observational wavelength (chromatic models). While it is not necessary for there to be a systematic knowledge of the visibility in absolute terms, it is often useful, or even indispensable, to study the relative variations of this value as a function of wavelength (for the observation of gaps in protoplanetary disks, for example – see Chap. 7).

2.3.3.2 Detection of Planets via Modulation of the Visibility

A star and a planet in orbit may be considered as a combination of two mutually incoherent sources, with a high level of contrast. In this case, the visibility of this combination via interferometry is equal to the sum of the visibilities, weighted to account for the relative intensity of the sources. Mathematically, this visibility may be expressed by the equation (Bordé, 2003):

$$V^2 = \frac{V_1^2 + V_2^2 + 2rV_1V_2\cos\psi}{(1+r)^2} \text{ where } r = \frac{I_2}{I_1} \quad (2.37)$$

V_1 and V_2 are the visibilities of the individual objects, I_1 and I_2 their relative intensities and ψ denotes the angular position of the star/planet system relative to the interferometer's baseline. This visibility function is shown graphically in Fig. 2.25.

Detection of objects with very low luminosity (high contrast) therefore requires an extremely accurate measurement of visibility. The maximum accuracy over several observations is in the region of 10^{-3} .

2.3.3.3 Cancellation Interferometry

Cancellation interferometry, also known as dark fringe interferometry or interference coronagraphy, differs uniquely from the method just described in its recombination technique. A diagrammatic representation of such an interferometer is shown in Fig. 2.26.

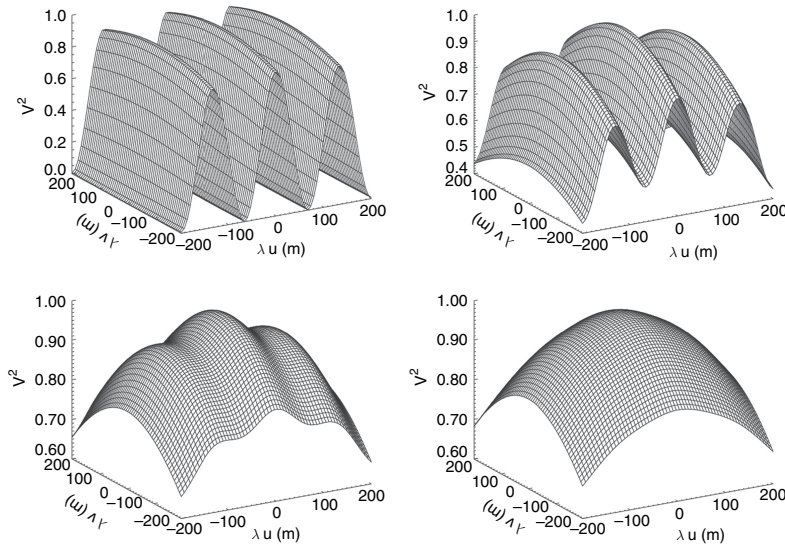


Fig. 2.25 Visibility functions for binary systems where the contrast is 1 (*top left*), 10 (*top right*), 100 (*bottom left*) and 1000 (*bottom right*) (After Bordé, 2003.)

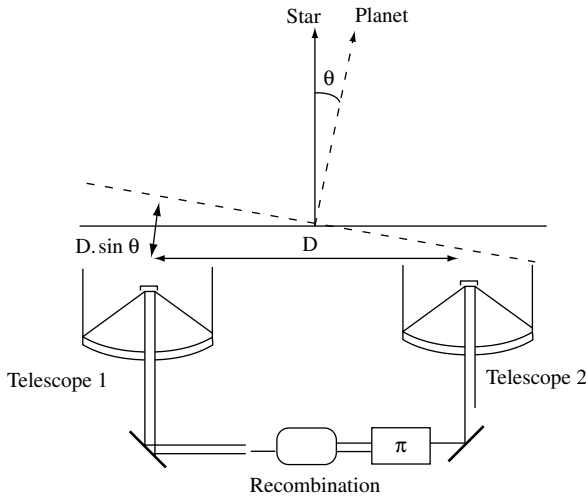


Fig. 2.26 Schematic diagram of a dark-fringe interferometer. Light arriving from the direction in which the instrument is aimed arrives simultaneously at the two telescopes. The achromatic dephaser, π , ensures destructive interference at the recombiner. Light from a point away from the target direction (from the planet) arrives at Telescope 1 with a delay relative to Telescope 2, because it has to cover an additional distance ($D \sin \theta$). By altering the distance between the telescopes (by varying D), it is possible to compensate for the π phase shift by the additional optical path, such that the interference is constructive in the direction of the planet. We thus obtain an instrument where the transmission is zero along the direction in which the instrument is pointed, and adjustable between 0 and 1 around that axis

Let us consider two telescopes T_1 and T_2 , which, because of diffraction are not able individually to resolve the star/planet pair. We point the two telescopes exactly in the direction of the star, and merge the beams of light from the two telescopes with an optical recombiner. In the direction of the star (i.e., where the two telescopes are pointing), the wavefront arrives simultaneously at T_1 and T_2 . If we recombine the two beams, they will be in phase and will produce constructive interference. If we then add an achromatic dephaser, π , in one arm of the interferometer (for example, the path from T_2), the light from the two telescopes will be recombined with opposite phases. In other words, the interference will be destructive and everything coming from the star's direction (in particular the star's flux) will be extinguished. In the direction of the planet (which typically lies at an angle $\theta \sim 0.1$ arcsec for a terrestrial analogue orbiting a star at a distance of 10 pc), we introduce a delay for T_1 relative to T_2 that is equal to $D \cdot \sin(\theta)$, where D is the distance between the two telescopes. If D is altered (by moving the telescopes), this may be done so that at an average wavelength (for example at the centre of a spectral band, or at two wavelengths so chosen as to maximize the spectral coverage) the difference in the path-lengths $D \cdot \sin(\theta)$ compensates for the dephasing that π introduces into the arm incorporating T_2 . This results in constructive interference in the direction of the planet. To summarize, such an instrument allows the observation of a faint object that lies outside the optical axis, where the theoretical transmission is zero. The response obtained with such an instrument is shown in Fig. 2.27. In its original version, the Bracewell interferometer was rotated such that the signal from the planet was modulated rel-

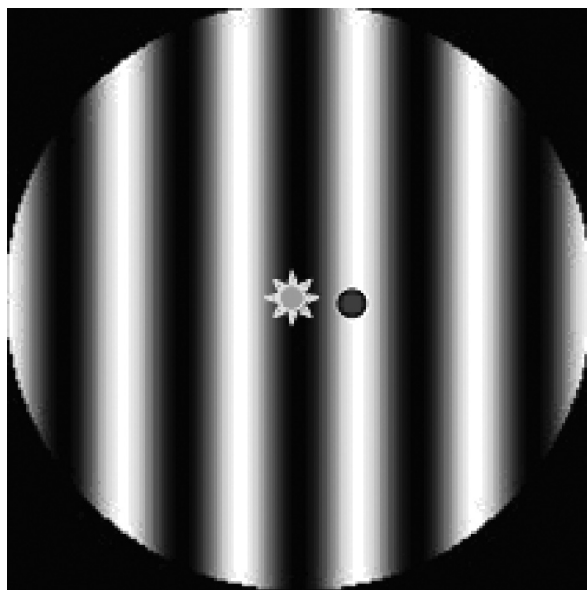


Fig. 2.27 Plot of the transmission of a dark-fringe interferometer with two telescopes (a Bracewell interferometer). The star is located on a dark fringe (it is extinguished), and the planet on a bright fringe

ative to the leakage of light from the star (because of the non-zero size of the star and of the finite size of the zone with zero transmission, the star is not perfectly extinguished).

In practice, a dark-fringe interferometer and its associated detection system (which may, in principle, be a single-pixel detector), behaves as a photometer with the spatial transmission shown in Fig. 2.27. In particular, measurements obtained with this instrument are measurements of the flux, which consists of several individual contributions:

- the off-axis object (for example, the planet),
- leaks from the star (because the central fringe is not perfectly dark and the star is not a perfect point), and the associated noise and fluctuations,
- other contributions related to the source, for example, a portion of the zodiacal disk in a planetary system,
- thermal emission from the instrument itself (in cases where such wavelengths are a problem).

The specific remedy that needs to be applied is to try to separate these different contributions to be able to isolate the component linked to the off-axis object that one is trying to detect and analyze. One classic technique consists of taking the geometry of the signals into account (the emission from a disk of zodiacal light is overall centrally symmetrical with respect to the plot of transmission, whereas that from the planet is not). To do this, the system is observed at several orientations of the interferometer, enabling discrimination of the different contributions. Another possibility is to use an internal modulation between several subsidiary interferometers.

This concept of a dark-fringe interferometer is the basis of the proposal for the DARWIN space mission and its American counterpart TPF (Terrestrial Planet Finder), described in Chap. 8.

It is interesting to note that from the instrumental point of view, the two interferometric concepts: the ‘measurement of visibility’ and dark-fringe interferometry, may, potentially, be combined in a single instrument. In particular, the achromatic dephaser does not prohibit the measurement of the experimental visibility. In such a case, it is necessary to limit exploration of the interference fringes to a small number of fringes, the characteristics of which (the amplitude and the phase) have been determined by a method that is precisely comparable with the classic method of measuring the visibility.

2.3.4 Interferometry and Imagery: Hypertelescopes

We have seen that the spatial resolution of a telescope is intrinsically limited by diffraction and, in practice, by the diameter of the telescope. The size of instruments is limited by current technology to:

- telescopes between 10 and 42 metres in diameter (ESO’s current Extremely Large Telescope projects). The largest current telescope is a multi-mirror system

- with an elliptical pupil (10×11 m): the Hobby Eberly Telescope, which is slightly larger than the two Keck telescopes (10 m in diameter). The largest monolithic telescopes are the two mirrors of the Large Binocular Telescope (each 8.4 m in diameter);
- space telescopes of a few metres in diameter. (The JWST is to have a diameter of about 6.5 m. This is the largest space instrument observing in the visible or infrared that has ever been constructed or is in the course of construction.)

However, if we are to contemplate the observation of planetary systems in the thermal infrared, or even direct imagery of exoplanetary surfaces in the visible, we will require diameters that are distinctly larger, and also with significantly greater collecting surfaces, because the flux from planets is weak. We soon encounter the limit for monolithic instruments – or at least with unified pupils (all points of the pupil are available without leaving the pupil).

Table 2.6 gives the minimal baseline (diffraction limited) as well as the collecting surface required to image the surface of a planet such as the Earth at a distance of 10 parsecs, with the corresponding spatial resolution (given as the number of pixels). The assumptions used for these calculations are:

- image of a planet like the Earth orbiting a solar analogue in the visible/near IR spectral region (0.1 photon/s/m² at the Earth, in the 0.3–1 μ m spectral region, in three colours, to obtain chromatic information i.e., to differentiate vegetation, continents, oceans, etc.);
- the integration time should be sufficiently short to ‘freeze’ the planet’s possible rotation and avoid a blurred image. In this calculation, the integration time is limited to one hour, but could be adapted according to the target if the rotation period were known;
- observation of the planet is limited by photon noise with a signal-to-noise ratio of 10 per pixel. As a result, an average flux of 100 planetary photons is required per bright pixel. (This relatively stringent assumption enables an estimate of the observation’s minimal characteristics.);
- at 10 parsecs, an Earth-like body would have an angular diameter of 8.5×10^{-6} arcsec;
- the overall transmission of the photometric chain (including the telescope and the detector’s efficiency) is assumed to be equal to 0.2.

Table 2.6 Parameters for imaging a terrestrial planet at a distance of 10 parsecs

Image size (pixels)	Collecting surface (m ²)	Minimum baseline (km)	Integration time
16 × 16	20 000	450	3 min
32 × 32	20 000	900	12 min
128 × 128	20 000	3600	50 min
256 × 256	320 000	7200	50 min
512 × 512	1 300 000	15 000	50 min
1024 × 1024	5 000 000	30 000	1 h

The minimum size of instruments required for planetary imagery clearly shows that we will have to abandon the idea of monolithic instruments and instead examine the possibility of imaging via interferometry.

The concept of imaging via interferometry is quite old, and several different configurations have been suggested. However, to obtain a field around the target, only the so-called Fizeau form is suitable. It is, nevertheless, very difficult to implement because it requires an optical system that guarantees homothetic mapping of the entry pupil to the exit pupil, and the accuracy of which increases as the spatial resolution increases (i.e., as the number of pixels in the image increases). If this homothetic mapping is not attained, the field is limited to the target direction (the centre of the image).

In addition, if we consider the baselines required to image planets (several 100 km), we are faced with the necessity of using small-sized telescopes (a few metres in diameter), separated by great distances (unless we use a large number of telescopes). As a result, if we want to preserve the homothetic mapping between the instrument's entry and exit pupils, the image will be very poor, and not capable of being exploited in terms of spatial frequencies. (Coverage of spatial frequencies is given by the pupil's autocorrelation function, which introduces many gaps in the frequency space if we use small telescopes separated by large distances) In other words, the image will be of very poor quality because it will contain few spatial frequencies.

An elegant solution for obtaining very-high-spatial-resolution imagery has been suggested by Labeyrie (1996), and this consists of combining the imaging capabilities of classic telescopes with the angular resolution given by interferometers. This concept, which has been called a 'hypertelescope' (Fig. 2.28), consists of several small telescopes that lie on the fictional mirror of a larger telescope (like Fizeau interferometers). An arrangement called a 'densified pupil' enables the frequency

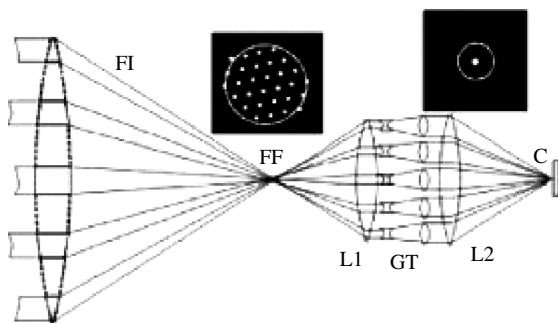


Fig. 2.28 A diagrammatic representation of the principle of a hypertelecope. Several small telescopes synthesize a larger aperture (shown schematically by a single lens). The intermediate image (FF) lacks spatial frequencies because the pupil is only weakly sampled. The densified-pupil combiner ($L_1 + GT + L_2$) enables a PSF to be attained that is comparable to that of a monolithic telescope with a size equal to the sum of those of the small telescopes. This gain is at the expense of the field that is covered. A compromise must be made between field size and frequency coverage (After Labeyrie, 1996)

coverage to be increased by assuming an equivalence between the sub-pupils of the hypertelescope's entry pupil. Admittedly, one then loses the homothetic mapping of the pupil (and thus reduces the field observed), but the whole interest in this arrangement is to find the best compromise between field and frequency coverage. The other special feature of hypertelescopes is that the (virtual) primary mirror that is covered with the small, individual telescopes is spherical. (The reason being that it simplifies the problem.) An optical corrector (a Mertz corrector) is then used to re-establish a stigmatic configuration at infinity, and to restore a large field.

The hypertelescope concept is currently being investigated. The Carlina project (see Chap. 8) should, in a few years' time, allow it to be validated and allow the first actual instruments to be defined, with initial baselines of a few hundred metres. The extension of this concept to baselines measured in kilometres or even more, and particularly within the framework of a space-borne implementation, will need to be accompanied by complementary efforts into positioning in precise formation. A hypertelescope in space, with long baselines, cannot be designed as a rigid structure. The sampling telescopes will therefore need to be autonomous in operation and positioning. The overall instrument will need to be maintained in a precise position, most especially when being aimed at a target, and during the observation. Several studies are under way, most notably within the framework of the DARWIN project (a dark-fringe interferometer), to validate ideas about formation positioning. Be that as it may, imaging the surface of a terrestrial-like exoplanet remains an extremely difficult task, which will require several generations of instruments before it becomes a reality. Intermediate stages will be necessary, such as imaging a whole planetary system. Luckily, it is not necessary to image the surface of a planet to obtain spectroscopic information about the planet itself, and thus be able to study it. This less sensitive method will therefore be favoured initially.

2.3.5 Detection by Radio

The problem of direct detection of exoplanets is generally described in terms of contrast and angular separation. This is, after all, the way in which we have initially attacked the subject. There is one spectral region where the contrast between the stellar flux and the planetary flux may be very low: this is a radio wavelengths, in particular at decametric wavelengths. In fact, planets with magnetic fields produce non-thermal auroral emissions (i.e., interactions between charged particles and the magnetic field in the polar regions as in the aurorae that occur on Earth). The intensity of these emissions is comparable, at these frequencies, with the emission from the star itself. The auroral emissions are very specific (Zarka et al., 1997):

- they have a typical duration of 30–300ms on the planets in our Solar System,

- their frequency spectrum is relatively uniform and broad (from less than 20 kHz to more than 40 MHz), with a typical spectral power that lies between 0.1 and 100 W Hz^{-1} , corresponding to a flux density that varies between 0.4 and 400 Jy.⁷

Detecting auroral emission from the Solar System's planets or from exoplanets is, however, rather tricky because of the presence of two main sources of parasitic signals:

- fluctuations in the sky background, of galactic origin, which have a very high brightness temperature (30 000–50 000 K at 25 MHz)
- parasitic signals of terrestrial (and human) origin. Such signals have the characteristic of being limited to relatively narrow frequency bands; and this obviously depends on the properties of the transmitter. Take, for example, CB⁸ signals, which completely saturate a frequency band around 27 MHz. On the other hand, the bands allotted to these emissions are numerous, spread out, and constantly being developed (Denis and Zarka, 1996).

In addition, at these wavelengths, the angular resolution is very low: the antenna's lobe is relatively broad (about 1° for an antenna 1 km in diameter, so the signals from both the star and the planet are detected simultaneously within the lobe). This breadth tends to increase the contribution of the galactic background in the signal that is detected. It is interesting to note that in this case it is not the noise in the detection chain that limits the sensitivity of the method, but the galactic contribution and any parasitic signals. Any detector, for example, can work at ambient temperatures.

Radio astronomers have suggested an observational strategy appropriate to these signals, and its specific features are as follows (Zarka et al., 1997):

- a receiving surface with a large area, so that the angular size of the antenna's lobe is reduced. In practice, the UTR-2 array at the Radio Astronomy Institute at Kharkov (Ukraine), which is currently the largest, and which is being used pending the commissioning of the future LOFAR (LOW Frequency ARray) and SKA (Square Kilometric Array) systems. The north-south arm of UTR-2 has an effective area of about $50\,000 \text{ m}^2$;
- the use of an acousto-optic spectrograph, which would enable time-frequency diagrams to be obtained with a temporal resolution (integration time) of about 250 ms, and spectral resolution of a few dozen kHz. The short duration of auroral emissions forbids the use of long integration times, because these would distinctly increase the contributions from parasitic signals and from the galactic background in every sample. It appears more sensible to integrate for short periods with durations that are comparable with those of the emission, and to eliminate empty samples so that only samples containing a signal are retained for integration;

⁷ $1 \text{ Jy} = 1 \text{ Jansky} = 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$.

⁸ Radio transmitting equipment, generally employed on vehicles. The development of mobile telephony (in a higher frequency region), however, is tending to marginalize their use, which appears to be restricted to professional drivers.

- handling the data with thorough observational and computational procedures, designed to eliminate parasitic signals (Zarka et al., 1997). In crude terms, this means detecting signals of terrestrial origin by observing two neighbouring areas of the sky, one on axis (and thus containing the astrophysical target), and the other off-axis. This technique may be used at Kharkov because the array's set of delay lines enables the array to be pointed in two different directions with an angular separation of 1° , and observations carried out in both lobes simultaneously.

Given the level of the background and parasitic signals, and using the method just described, it may be shown that, using UTR-2, it would be possible to detect Jupiter at about 0.2 parsec. This is, unfortunately, insufficient to even offer the hope of detecting extrasolar planets (remember that the star closest to the Sun, Proxima Centauri, lies at a distance of about 1.3 parsecs). On the other hand, Zarka has also shown that, in the case of giant planets close to their parent star (i.e., hot Jupiters), one could expect a mechanism that generated radio emission 10^3 – 10^5 times as strong as that produced by Jupiter. Under such circumstances, we could hope to detect emissions out to about 20–25 parsecs, which makes the method distinctly more attractive.

A series of observations has been made with UTR-2, but the complex processing to which the data are subject means that, to date, no candidates have yet been identified. When the SKA array is commissioned, it should increase the sensitivity of the method by reducing the size of the antenna's lobe and, as a result, the level of the galactic background.

Bibliography

- Beaulieu, J.-P., et al., 'Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing', *Nature*, **439**, 437–440 (2006)
- Bennett, D. et al., 'Discovery of a planet orbiting a binary star system from gravitational microlensing', *Nature*, **402**, 57–59 (1999)
- Beuzit, J.-L., Mouillet, D., Lagrange, A.-M. and Pauflique, J., 'A stellar coronagraph for the COMEON-PLUS adaptive optics system', *A&ASS*, **125**, 175–182 (1997)
- Bordé, P., *Détection et caractérisation de planètes extrasolaires par photométrie visible et interférométrie infrarouge à très haute précision*, PhD thesis, Université de Paris-VI (2003)
- Bracewell, R.N., 'Detecting nonsolar planets by spinning infrared interferometer', *Nature*, **274**, 780 (1978)
- Brown, T.M., Charbonneau, D., Gilliland, R.L., Noyes, R.W. and Burrows, A., 'Hubble space telescope time-series photometry of the transiting planet of HD 209458', *Astrophys. J.*, **552**, 699–709 (2001)
- Charbonneau, D., Brown, T.M., Latham, D.W. and Mayor, M., 'Detection of planetary transit across a sun-like star', *Astrophys. J.*, **529**, L45–L48 (2000)
- Charbonneau, D. et al., 'Detection of thermal emission from an extrasolar planet', *ApJ*, **626**, 523–529 (2005)
- Davis, M.M., Taylor, J.H., Weisberg, J.M., Backer, D.C., 'High-precision timing observations of the millisecond pulsar PSR 1937 +21', *Nature*, **315**, 547–550 (1985)
- Denis, L., Zarka, P., Interference problems at the Nançay Decameter Array and studies towards a better immunity, JCE Symposium "Interference problems in Radio Astronomy and Com-

- munications – or Cosmic Ecology”, U.R.S.I. XXVth General Assembly, Lille, France, p. 752 (1996)
- Einstein, A., *Science*, **84**, 506 (1936)
- Goodman, J.W., *Introduction to Fourier Optics*, 2nd edn, McGraw Hill, Columbus, Ohio (1996)
- Guyon, O., ‘Phase-induced amplitude apodization of telescopes pupils for extrasolar terrestrial planet imaging’, *Astron. Astrophys.*, **404**, 379–387 (2003)
- Guyon, O., Pluzhnik, E.A., Galicher, R. and Martinache, F., ‘Exoplanet imaging with a phase-induced amplitude apodization coronagraph. I. Principle’, *ApJ*, **622**, 744–758 (2005)
- Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E., Perley, R., ‘2237 + 0305 – A new and unusual gravitational lens’, *Astronom. J.*, **90**, 691–696 (1985)
- Koutchmy, S., ‘Space-borne coronagraphy’, *Space Sci. Rev.*, **47**, 95 (1988)
- Labeyrie, A., ‘Resolved imaging of extra-solar planets with future 10–100 km optical interferometric arrays’, *A & A*, **118**, 517–524 (1996)
- Léna, P., ‘Observational Astrophysics’, sec. Ed, Springer, Berlin (1998)
- Lyot, B., ‘Photographie de la couronne solaire en dehors des éclipses’, *C.R. Acad. Sci. Paris*, **193**, 1169 (1931)
- Lyot, B., ‘A study of the solar corona and prominences without eclipses’, *Mon. Not. R. Astron. Soc.*, **99**, 580 (1939)
- Malbet, F., ‘High angular resolution coronagraphy for adaptive optics’, *A&ASS*, **115**, 161–174 (1996)
- Racine, R., ‘Continuum and semiforbidden C III microlensing in Q2237 + 0305 and the quasar geometry’, *ApJ*, **395**, L65–L67 (1992)
- Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y. and Labeyrie, A., ‘The four-quadrant phase-mask coronagraph. I. Principle’, *PASP*, **777**, 1479–1486 (2000)
- Roddier, F. and Roddier, C., ‘Stellar coronagraph with phase mask’, *PASP*, **109**, 815–820 (1997)
- Sackett, P., ‘Searching for unseen planets via occultation and microlensing’ in *Planets Outside the Solar System: Theory and Observation*, Mariotti, J.-M and Alloin, D. Eds, NATO-ASI Series, Kluwer, 189–227 (1999)
- Wolszczan, A. and Frail, D., ‘A planetary system around the millisecond pulsar PSR1257+12’, *Nature*, **255**, 145–147 (1992)
- Zarka, P., Queinnec, J., Ryabov, V., Shevchenko, V., Arhipov, A., Rucker, H., Denis, L., Gerbault, A., Dierich, P., Rosolen, C., ‘Ground-based high sensitivity radio astronomy at decimeter wavelength’, in *Planetary Radio Emissions IV*, (eds) Ruckern, H.O. Bauer, S.J. and Lecacheux, A., Austrian Acad. Sci. Press, Vienna (1997)

Planetary Systems

Detection, Formation and Habitability of Extrasolar
Planets

Ollivier, M.; Encrenaz, T.; Roques, F.; Selsis, F.; Casoli, F.

2009, XIII, 344 p., Hardcover

ISBN: 978-3-540-75747-4