Astrophysics and Space Science Library 428

Valerio Bozza
Luigi Mancini
Alessandro Sozzetti
Editors

Methods of Detecting Exoplanets

1st Advanced School on Exoplanetary Science

ASSL

Springer

Methods of Detecting Exoplanets

Astrophysics and Space Science Library

Valerio Bozza • Luigi Mancini • Alessandro Sozzetti
Editors

Methods of Detecting Exoplanets

1st Advanced School on Exoplanetary Science

Springer

Editors
Valerio Bozza
Department of Physics
University of Salerno
Fisciano, Italy

Luigi Mancini
Planet and Star Formation
Max Planck Institute for Astronomy
Heidelberg, Germany

Alessandro Sozzetti
INAF – Osservatorio Astrofisico di Torino
Pino Torinese (TO), Italy

ISSN 0067-0057 ISSN 2214-7985 (electronic)
Astrophysics and Space Science Library
ISBN 978-3-319-27456-0 ISBN 978-3-319-27458-4 (eBook)
DOI 10.1007/978-3-319-27458-4

Library of Congress Control Number: 2016936368

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

To our princesses

Preface

Life as we know it on Earth intrinsically needs to explore and colonize new lands with suitable conditions in order to persist and propagate. In this sense, the "search for other world" has started long before the appearance of human beings. However, while the discovery and colonization of new lands was a relatively slow process for inferior organisms, the abilities of pre-historic men were enough to create villages on the whole habitable lands on our planet in the relatively short span of one million years. With so little left to explore on Earth, philosophers and astronomers in the last five thousand years started to wonder whether stars in the sky actually hosted other planets similar to ours. In more modern times, science fiction has been able to imagine many different worlds inhabited by more or less frightening or friendly creatures. At the end of the twentieth century, with such a strong cultural background, corroborated by the first successful explorations of our nearby celestial body (the Moon), the scientific proof of the existence of planets around other stars was highly expected and sought for. In more recent years, after several economic and political crisis, with the ghosts of global warming, pollution, and overpopulation, the existence of distant habitable worlds is no longer regarded as a simple satisfaction of human curiosity. The conscience of the fragility of our Earth forces us to study other planets to better understand their evolution and what the fate of our old Earth may finally be. Furthermore, it is foreseeable that in a (still) far future, with technologies that we cannot barely imagine yet, we shall move away from Earth looking for new home-worlds. With this very distant but inevitable perspective in mind, we can regard the present time as the beginning of a new era in which humanity first sights distant lands trying to scrutinize their habitability conditions in view of possible future colonization.

Indeed, with its philosophical and political implications, there is probably no other field in modern astrophysics for which the outreach to the general audience can be so easy as in the search for exoplanets. This fact has boosted this newly born field to top ranking in the attention by the media and by the funding agencies. In spite of this fact, the techniques used to detect and characterize extrasolar planets are far from being immediately understandable to a nonspecialist. Even within the same astrophysics community, there are not many people aware of

what the words "Strehl ratio," "caustic," "periodogram," and "Rossiter- McLaughlin effect" exactly mean. In order to fill this gap between an esoteric planet-finder community and the average physics and astrophysics student, we have gathered four top scientists, representative of the four most successful detection methods known today, in the enchanting cornice of Vietri sul Mare, in the Amalfi Coast. The lectures given by these renown scientists cover the direct imaging method, transits, radial velocities, and microlensing. Each of these methods has its own merits and lacks in investigating planetary systems.

Transits and radial velocities have produced the greater return in terms of number of exoplanets, also thanks to the first space mission (*Kepler*) fully devoted to exoplanets discovery. When combined together, these two techniques can yield the best characterization of the planets, including mass, radius, density, and orbital parameters. In some cases, it has been even possible to infer the chemical components of the atmospheres of transiting planets through fine spectroscopic techniques. However, these methods are best suited for planets very close to the parent star and only recently have started to graze the so-called habitable zone.

Microlensing can probe the frequency of planets orbiting at intermediate distances from the parent star, just beyond the so-called snow line, where giant planets are believed to form. It is also the only method to find planets that are very far from our Earth or even in other galaxies. It is finally the only way to find isolated planets, ejected from the system where they were born. Unfortunately, microlensing events are non-repeatable and do not allow further measurements to refine the planetary parameters.

Direct imaging is probably the most rewarding technique since it makes the planets shine out of the glare of their parent star. Very refined adaptive optics and coronographic techniques are needed to achieve such spectacular results. In some cases it is possible to follow the orbits of the planets and study their spectra. Of course, only planets very far from the star and still hot enough can be directly detected in this way.

These very short statements of the four methods are sufficient to understand how they complement each other as in a big puzzle where every piece is necessary for a full understanding of the global architecture of planetary systems. By probing the planetary frequency at different distances and in different conditions, these techniques are helping astrophysicists to reconstruct the scenarios of planetary formation and to give a robust scientific answer to the questions regarding the frequency of potentially habitable worlds. More difficult is to answer the question about the existence of forms of extraterrestrial life, because the conditions for habitability are always temporary and may not last long enough to allow the development of advanced creatures.

Nevertheless, a great effort is being lavished on the construction of new facilities, both on ground and in space, with the main aim of investigating these problems. It is no surprise that the search and characterization of exoplanets appears in the main goals of ALMA, E-ELT, JWST, and most of the spacecraft missions that are being designed by the main space agencies. With the increasing attention toward the search for exoplanets, it is then imperative to prepare the future generation of scientists to

take over from the present researchers, with the hope that they will be able to further expand our knowledge with innovative and enlightening ideas, stemming from the roots of our secular and instinctive spirit of exploration. In this respect, we hope that this book, by unveiling the tricks of the trade of planet detection to a wider community, will make a good service to science and humanity in general.

May 25, 2015

Valerio Bozza
Luigi Mancini
Alessandro Sozzetti

Acknowledgments

The organizing committee of the 1st Advanced School on Exoplanetary Science would like to thank the Max Planck Institute for Astronomy, the Department of Physics of the University of Salerno, and the Italian National Institute for Astrophysics (INAF) for their financial support. The organizing committee kindly thanks the director of the International Institute for Advanced Scientific Studies (IIASS) in Vietri sul Mare, *Prof. Emeritus Ferdinando Mancini*, for hosting the event. Finally, the organizing committee would also like to acknowledge the great help offered by the staff of the IIASS, *Mrs. Tina Nappi* and *Mr. Michele Donnarumma*, for the organization of the school.

Participants to the 1st Advanced School on Exoplanetary Science, Vietri sul Mare, May 25, 2015

Contents

List of Contributors

Andrew Collier Cameron SUPA, School of Physics and Astronomy, University of St Andrews, St Andrews, UK

Riccardo Claudi INAF – Osservatorio Astronomico di Padova, Padova, Italy

Andrew Gould Department of Astronomy, Ohio State University, Columbus, OH, USA

Artie P. Hatzes Thüringer Landessternwarte Tautenburg, Tautenburg, Germany

Part I
The Radial Velocity Method

Chapter 1
The Radial Velocity Method for the Detection of Exoplanets

Artie P. Hatzes

Abstract The radial velocity (RV) method has provided the foundation for the research field of exoplanets. It created the field by discovering the first exoplanets and then blazed a trail by detecting over 1000 exoplanets in orbit around other stars. The method also plays a vital role in transit searches by providing the planetary mass needed to calculate the bulk density of the exoplanet. The RV method requires a wide range of techniques: novel instrumentation for making precise RV measurements, clever techniques for extracting the periodic signals due to planets from the RV data, tools for assessing their statistical significance, and programs for calculating the Keplerian orbital parameters. Finally, RV measurements have become so precise that the measurement error is now dominated by the intrinsic stellar noise. New tools have to be developed to extract planetary signals from RV variability originating from the star. In these series of lectures I will cover (1) basic instrumentation for stellar radial velocity methods, (2) methods for achieving high radial velocity precision, (3) finding periodic signals in radial velocity data, (4) Keplerian orbits, (5) sources of errors for radial velocity measurements, and (6) dealing with the contribution of stellar noise to the radial velocity measurement.

1.1 Introduction

The radial velocity (RV) method has played a fundamental role in exoplanet science. Not only is it one of the most successful detection methods with over 1000 exoplanet discoveries to its credit, but also it is the method that "kicked off" the field. If it were not for RV measurements, we would probably not be studying exoplanets today. The first hints of exoplanet discoveries with this method date to the late 1980s and early 1990s (Campbell et al. 1988; Latham et al. 1989; Hatzes and Cochran 1993) and culminated with the discovery of 51 Peg b (Mayor and Queloz 1995). In the past 20 years exoplanets have become a vibrant field of exoplanet research. Although the transit method, largely due to NASA's Kepler mission

A.P. Hatzes (✉)
Thüringer Landessternwarte Tautenburg, Sternwarte 5, 07778 Tautenburg, Germany
e-mail: artie@tls-tautenburg.de

V. Bozza et al. (eds.), *Methods of Detecting Exoplanets*, Astrophysics and Space Science Library 428, DOI 10.1007/978-3-319-27458-4_1

(Borucki et al. 2009), has surpassed the RV method in terms of shear number of exoplanet discoveries, the RV method still plays an important role in transit discoveries by providing the mass of the planetary companion. Without this one cannot calculate the bulk density of the planet needed to determine its structure (gas, icy, or rocky planet).

The basic principle behind the RV method for the detection of exoplanets is quite simple. One measures the line-of-sight (radial) velocity component of a star as it moves around the center of mass of the star-planet system. This velocity is measured via the Doppler effect, the shift in the wavelength of spectral lines due to the motion of the star.

The Doppler effect has been known for a long time. Christian Doppler discovered the eponymous effect in 1842. The first stellar radial velocity measurements of stars using photography were taken almost 150 years ago (Vogel 1872). In fact, in arguably one of the most prescient papers ever to be written in astronomy, Struve (1952) over 60 years ago in his article "*Proposal for a Project of High-Precision Stellar Radial Velocity Work*" proposed building a powerful spectrograph in order to search for giant planets in short period orbits. As he rightfully argued "We know that stellar companions can exist at very small distances. It is not unreasonable that a planet might exist at a distance of 1/50 of an astronomical unit. Such short-period planets could be detected by precise radial velocity measurements." With such a long history one may ask "Why did it take so long to detect exoplanets?" There are two reasons for this.

First, the Doppler shift of a star due to the presence of planetary companions is small. We can get an estimate of this using Newton's form of Kepler's third law:

$$P^2 = \frac{4\pi^2 a^3}{G(M_s + M_p)} \tag{1.1}$$

where M_s is the mass of the star, M_p is the mass of the planet, P the orbital period, and a the semi-major axis.

For planets $M_s \gg M_p$. If we assume circular orbits (a good approximation, for the most part) and the fact that $M_p \times a_p = M_s \times a_s$, where a_s and a_p are the semi-major axes of the star and planet, respectively, it is trivial to derive

$$V[\mathrm{m\,s^{-1}}] = 28.4 \left(\frac{P}{1\mathrm{yr}}\right)^{-1/3} \left(\frac{M_p \sin i}{M_{\mathrm{Jup}}}\right) \left(\frac{M_s}{M_\odot}\right)^{-2/3} \tag{1.2}$$

where i is the inclination of the orbital axis to the line of sight.

Figure 1.1 shows the reflex motion of a one solar mass star due to various planets at different orbital radii calculated with Eq. (1.2). A Jupiter analog at an orbital distance of 5.2 AU will induce an $11.2\,\mathrm{m\,s^{-1}}$ reflex motion in the host star with an orbital period of 12 years. A Jupiter-mass planet closer to the star would induce an amplitude of several hundreds of $\mathrm{m\,s^{-1}}$. An Earth-like planet at 1 AU would cause a reflex motion of the host star of a mere $10\,\mathrm{cm\,s^{-1}}$. This only increases to $\approx 1\,\mathrm{m\,s^{-1}}$

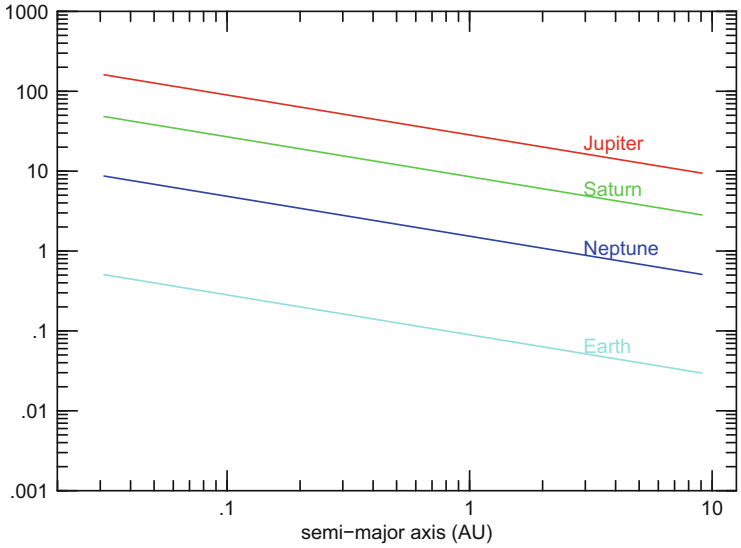

Fig. 1.1 The amplitude of the barycentric radial velocity variations for a one solar mass star orbited by Earth, Neptune, Saturn, or Jupiter at various orbital distances

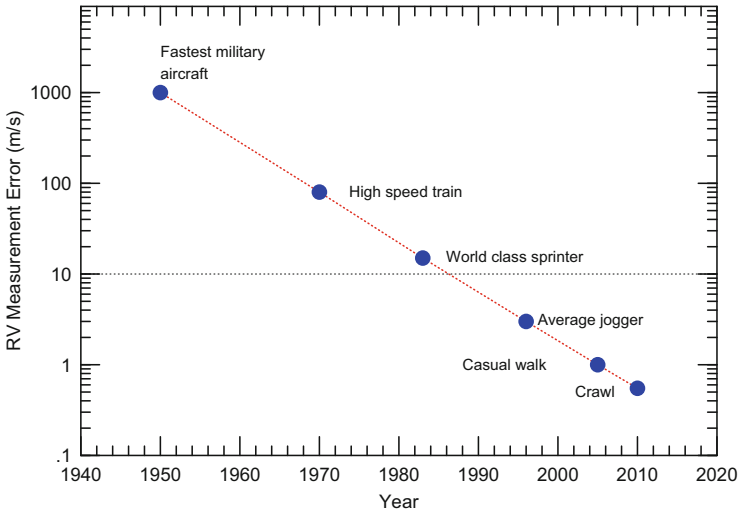

Fig. 1.2 The evolution of the radial velocity measurement error as a function of time. The *horizontal line* marks the reflex motion of a solar mass star with a Jupiter analog

by moving this planet to 0.05 AU. This figure shows that to detect planets with the RV method one needs exquisite precision coupled with long term stability.

Second, the measurement precision was woefully inadequate for detecting the reflex motion of a star due to a planet. Figure 1.2 shows the evolution of the RV

measurement error as a function of time. In the mid-1950s, using single-order spectrographs and photographic plates, one rarely achieved an RV measurement precision better than $1\,\mathrm{km\,s^{-1}}$, or about the speed of the fastest military aircraft (SR-71 Blackbird). By the end of the 1980s the measurement error had decreased to about $15\,\mathrm{m\,s^{-1}}$, or about the speed of a world class sprinter. At this time electronic detectors with high quantum efficiency had come into regular use. Simultaneous wavelength calibration methods were also first employed at this time. However, single-order spectrographs were still used and these had a restricted wavelength coverage which limited the RV precision.

Currently, modern techniques are able to achieve an RV precision of 0.5–1 $\mathrm{m\,s^{-1}}$, or about the speed of a fast crawl. The horizontal line shows our nominal RV precision of $10\,\mathrm{m\,s^{-1}}$ needed to detect the reflex motion of a solar-like star due to a Jupiter at 5 AU. It is no surprise that the first exoplanets were detected at about the time that this $10\,\mathrm{m\,s^{-1}}$ measurement error was achieved.

This spectacular increase in RV precision was reached though three major developments:

- High quantum efficiency electronic detectors
- Large wavelength coverage cross-dispersed echelle spectrographs
- Simultaneous wavelength calibration

In these lecture notes I will focus entirely on the Doppler method, the techniques to achieve the precision needed for the detection of exoplanets, how to analyze radial velocity data, and most importantly how to interpret results. I will *not* discuss results from Doppler surveys. These can be found in the now vast literature on the subject as well as on web-based databases (*exoplanet.eu and exoplanets.org*). The goal of these lectures is twofold. For those readers interested in taking RV measurements it should help to provide them with the starting background to take, analyze, and interpret their results. For readers merely wanting background knowledge on the subject, it well help them read papers on the RV detection of planets with a more critical eye. If you have some knowledge on how the method works, its strengths and pitfalls, then you can make your own judgments on whether an RV planet discovery is real, an artifact of instrumental error, or due to intrinsic stellar variations.

These notes are divided into six separate lectures (starting with Sect. 1.2) that deal with the following topics:

1. *Basic instrumentation*
   In this lecture a short description is given on the basic instrumentation needed for precise RV measurements.
2. *Precise stellar RV measurements*
   In this lecture I will cover the requirements on your spectrograph and data quality needed to achieve an RV precision for the detection of exoplanets. I will also describe how one eliminates instrumental shifts—the heart of any technique for the RV detection of exoplanets.

3. *Time Series Analysis: Finding planets in your RV data*
   Finding periodic signals in your RV data is arguably the most important step towards exoplanet detection. If you do not have a periodic signal in your data, then you obviously have not discovered a planet. In this lecture I will cover some of the basic tools that planet hunters use for extracting planet signals from RV data.
4. *Keplerian Orbits*
   Once you have found a periodic signal one must then calculate orbital elements. In this lecture I will cover basic orbital elements, the RV variations due to planets, and how to calculate orbits.
5. *Sources of Errors and Fake Planets*
   This lecture covers sources of errors, both instrumental and intrinsic to the star that may hinder the detection of exoplanets, or in the worse case mimic the RV signal of a planetary companion to the star.
6. *Dealing with the activity signal*
   The RV signal due to stellar activity is the single greatest obstacle preventing us from finding the lightest planets using the RV method. In this lecture I will cover some simple techniques for extracting planet signals in the presence of an activity signal.

1.2 Basic Instrumentation for Radial Velocity Measurements

The use of electronic detectors and echelle spectrographs provided the foundation for the improved RV precision needed to detect exoplanets. In this lecture I will briefly cover electronic detectors and the characteristics that may influence the RV precision, as well as the basics of spectrographs and how the RV precision depends on the characteristics of the spectral data.

1.2.1 Electronic Detectors

The use of photographic plates at the turn of the last century revolutionized astronomical observations. Astronomers could not only record their observations, but photographic plates enabled them to integrate on objects thus achieving a higher photon count. In spite of this development, photographic plates could rarely achieve a signal-to-noise ratio (S/N) more than 5–20. At this S/N level you might measure a Doppler shift of a few tens of km s^{-1}, but the Doppler shift due to the reflex motion of Jupiter would have been difficult.

The advent of electronic detectors played the first key role in improving stellar RV measurements. Figure 1.3 shows approximately the evolution in the quantum efficiency (QE) of electronic detectors as a function of time. Photomultipliers provided an increase in QE by an order of magnitude in the 1940s, followed by

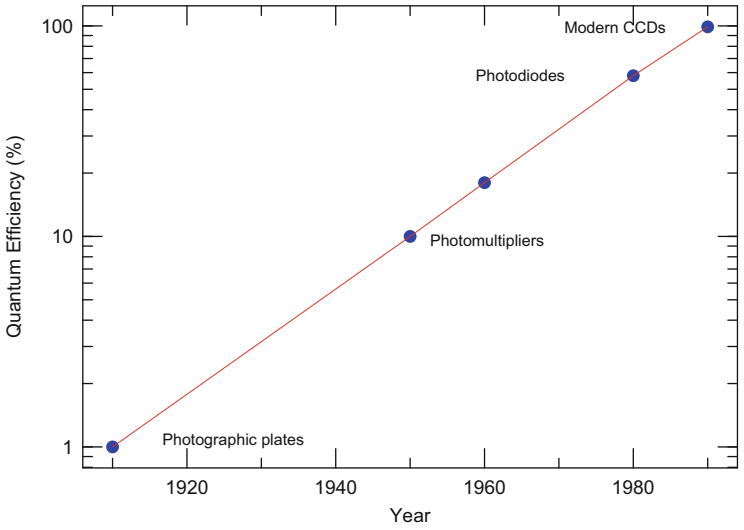

Fig. 1.3 The evolution of the quantum efficiency (QE) of astronomical detectors as a function of time

photodiode arrays. Charge coupled devices (CCDs) began to have routine use in the 1990s. Modern science grade CCD detectors can now reach quantum efficiencies close to 100 %. Note that it is no small coincidence that the discovery of exoplanets coincided with the use of 2-dimensional CCD detectors that incidentally were a glorious match to echelle spectrographs (see below).

Most RV measurements are made at optical wavelengths using CCD detectors. There are several issues when reducing data taken with CCD detectors that could be important for RV precision. First, there are slight variations in the QE from pixel to pixel that have to be removed via a so-called flat-fielding process. One takes a spectrum (or image) of a white light source (flat lamp) and divides each observation by this flat field. Flat fielding not only removes the intrinsic pixel-to-pixel variations of the CCD, but also any ghost images, reflections, or other artifacts coming from the optical system.

Figure 1.4 shows an example of the flat-fielding process as applied to imaging observations where you can better see the artifacts. The top left image is a raw frame taken with the Schmidt camera of the Tautenburg 2 m telescope. The top right image shows an observation of the flat lamp where one can see the structure of the CCD, as well as an image of the telescope pupil caused by reflections. The lower image shows the observed image after dividing by the flat lamp observation. Most of the artifacts and intensity variations have been removed by the division.

Fringing is another problem with CCDs that is caused by the small thickness of the CCD. It occurs because of the interference between the incident light and the light that is internally reflected at the interfaces of the CCD. Figure 1.5 shows a spectrum of a white light source taken with an echelle spectrograph (see below).

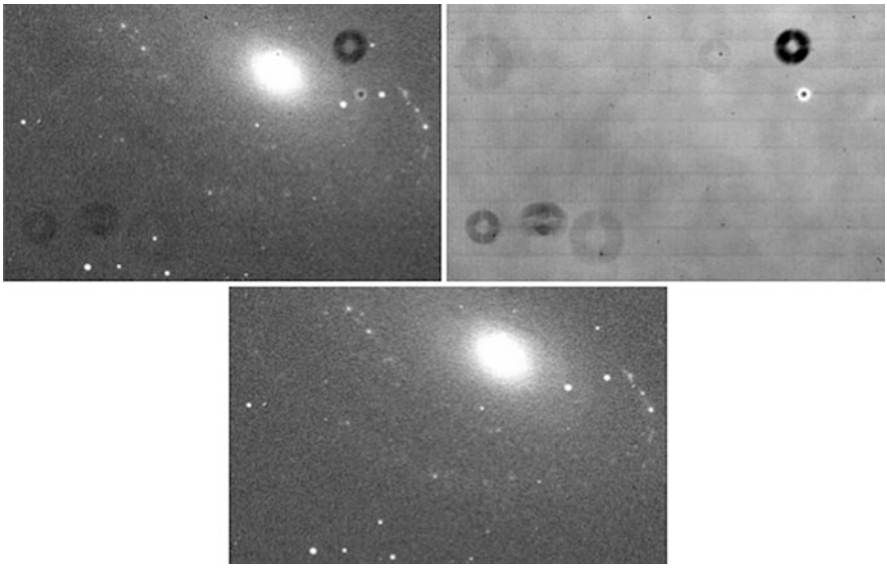

Fig. 1.4 The flat-fielding process for CCD reductions. (*Left top*) A raw image taken with a CCD detector. (*Right top*) An image taken of a white light source (flat field) that shows the CCD structure and optical artifacts. (*Bottom*) The original image after dividing by the flat field

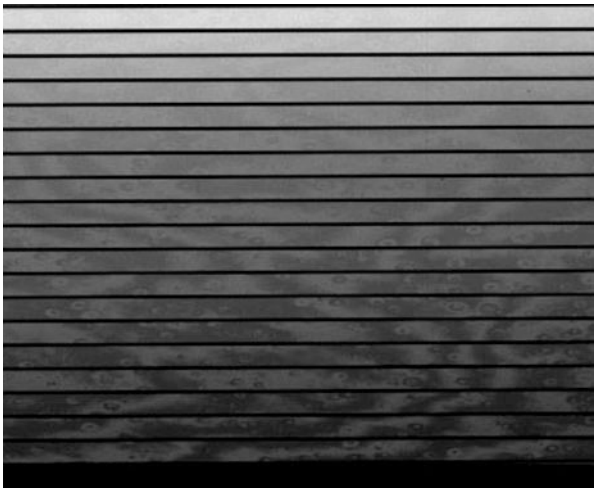

Fig. 1.5 Fringing in a CCD. Shown are spectral orders of an echelle spectrum (see below) of a white light source. Spectral orders with blue wavelengths ($\approx$5000 Å) are at the *top*, red orders ($\approx$7000 Å) are at the *bottom*. Fringing is more evident in the red orders

Red wavelengths are at the lower part of the figure where one can clearly see the fringe pattern. This pattern is not present in the orders at the top which are at blue wavelengths.

CCD fringing is mostly a problem at wavelengths longer than about 6500 Å. For RV measurements made with the iodine technique (see below) this is generally not a problem since these cover the wavelength range 5000–6000 Å. However, the simultaneous Th-Ar method (see below) can be extended to longer wavelengths where improper fringe removal may be a concern.

In principle, the pixel-to-pixel variations of the CCD and the fringe pattern should be removed by the flat field process, but again this may not be perfect and this can introduce RV errors. A major source of residual flat-fielding errors and improper removal of the fringe pattern is due to the fact that the light of the flat lamp does not go through the same optical path as the starlight. To replicate this one usually takes a so-called dome flat. For these one uses the telescope to observe an illuminated screen mounted on the inside of the dome. This light now follows as closely as possible the path taken by the starlight. For RV measurements the flat field characteristics may change from run to run and this may introduce systematic errors.

Finally, residual images can be a problem for poor quality CCDs or observations made at low light levels. After an observation regions of the CCD that have had a high count rate, or have been overexposed, may retain a memory of the previous image. This residual charge is not entirely removed after reading out the CCD (Fig. 1.6). This is generally a problem when a low signal-to-noise (S/N) observation is taken after one that had high counts. For precise stellar RVs the observations we generally are at high signal-to-noise ratios and there is little effect on the RV error even if residual images are present. However, this could introduce a significant RV error when pushing Doppler measurements to small shifts, i.e. $\sim$cm s^{-1}. If residual images are a problem, then one must read out the CCD several times before or after

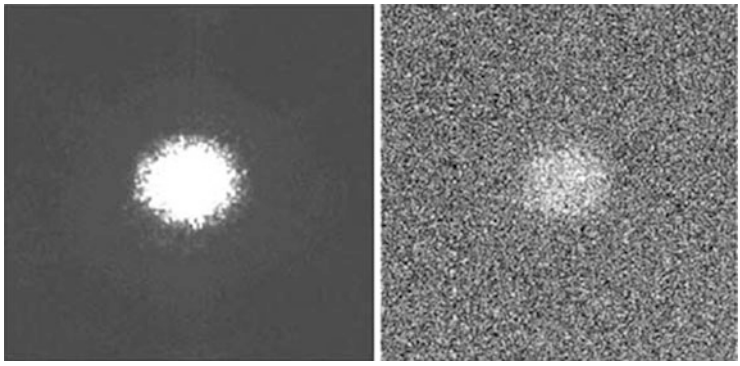

Fig. 1.6 Residual images in a CCD. (*Left*) An image of a star with a high count level. (*Right*) An image of the CCD after reading out the previous exposure. There is a low level (a few counts) image of the star remaining on the CCD

each new science observation. The exact number of readouts depends on the CCD. In general one should see how many readouts are required before the level of the residual image is at an acceptable level.

1.2.2 Echelle Spectrographs

The heart of any spectrograph is a dispersing element that breaks the light up into its component wavelengths. For high resolution astronomical spectrographs this is almost always a reflecting grating, a schematic that is shown Fig. 1.7. The grating is ruled with a groove spacing, σ. Each groove, or facet, has a tilt at the so-called blaze angle, ϕ, with respect to the grating normal. This blaze angle diffracts most of the light into higher orders, m, rather than the $m = 0$ order which is white light with no wavelength information.

Light hitting the grating at an angle α is diffracted at an angle β_b and satisfies the grating equation:

$$\frac{m\lambda}{\sigma} = \sin\alpha + \sin\beta_b \tag{1.3}$$

Note that at a given λ, the right-hand side of Eq. (1.1) is $\propto m/\sigma$. This means that the grating equation has the same solution for small m and small σ (finely grooved), or alternatively, for large m and large σ (coarsely grooved).

Over 30 years ago manufacturers were unable to rule gratings at high blaze angle. Blaze angles were typically $\phi \approx 20°$. To get sufficient dispersion one had to use finely ruled gratings with $1/\sigma$ = 800–1200 grooves/mm. The consequence of this was that astronomers generally worked with low spectral orders. There was some spatial overlap of orders so blocking filters were used to eliminate contaminating light from unwanted wavelengths.

Currently manufacturers can produce gratings that have high blaze angles ($\phi \approx 65°$) that are coarsely ruled ($1/\sigma \approx 30$ grooves/mm). Echelle gratings work at high

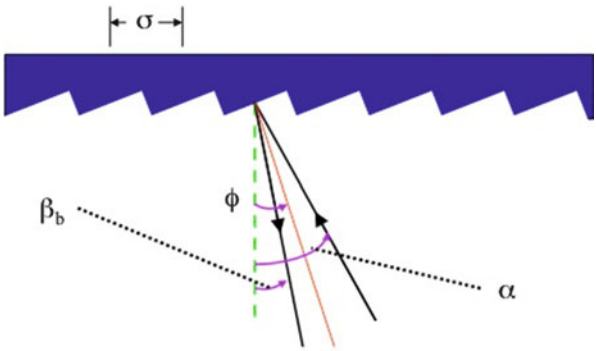

Fig. 1.7 Schematic of an echelle grating. Each groove facet has a width σ and is blazed at an angle ϕ with respect to the grating normal (*dashed line*). Light strikes the grating at an angle α, and diffracted at an angle β_b, both measured with respect to the grating normal

spectral orders ($m \approx 50$–100) so the large spacing σ, but large m means that the echelle grating has the same dispersing power as its finely ruled counterparts. They are also much more efficient.

A consequence of observing at high m values is that all spectral orders are now spatially stacked on top of each other. One needs to use interference filters to isolate the spectral order of choice. But why throw light away? Instead, almost all current echelle spectrographs use a cross-dispersing element, either a prism, grating or grism, that disperses the light in the direction perpendicular to the spectral dispersion. This nicely separates each of the spectral orders so that one can neatly stack them on your 2-dimensional CCD detector. It is no small coincidence that cross-dispersed echelle spectrographs become common at the time CCD detectors were readily available.

Figure 1.8 shows the schematic of a "classic" spectrograph consisting of a reflecting Schmidt camera. The converging beam of light from the telescope mirror comes to a focus at the entrance slit of the spectrograph. This diverges and hits the collimator which has the same focal ratio as the telescope beam ($F_{\rm col} = F_{\rm tel}$). This turns the diverging beam of starlight into a collimated beam of parallel light. The parallel beam then strikes the echelle grating which disperses the light into its wavelength components (only a red and blue beams are shown). These pass through a cross-dispersing element. The light finally goes through the Schmidt camera, complete with corrector plate, that brings the dispersed light to a focus. Although I have shown a reflecting camera, this component can also be constructed using only refractive (lens) elements, however these tend to be more expensive to manufacture.

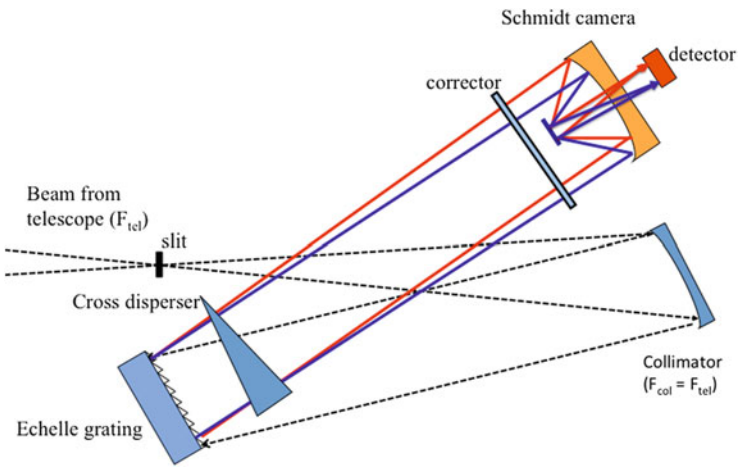

Fig. 1.8 Schematic of a classic spectrograph with a Schmidt camera. Light from the telescopes come to a focus at the slit. The diverging beam hits the collimator which has the same focal ratio as the telescope. This converts the diverging beam into a parallel beam that strikes the echelle grating which disperses the light. The dispersed light passes through the cross disperser, shown here as a prism, before the Schmidt camera brings the light to a focus at the detector

Before the advent of cross-dispersed elements, older spectrographs would have a similar layout except for the cross-dispersing element.

It is important to note that a spectrograph is merely a camera (actually, a telescope since it is also bringing a parallel wavefront, like starlight, to a focus). The only difference is the presence of the echelle grating to disperse the light and in this case, the cross-dispersing element. Remove these and what you would see at the detector is a white light image of your entrance slit. Re-insert the grating and the spectrograph now produces a *dispersed* image of your slit at the detector. This is also true for absorption and emission lines. These will also show an overall shape of the slit image, or in the case of a fiber-fed spectrograph a circular image. (The stellar absorption lines will of course also have the shape broadening mechanisms of the stellar atmosphere.)

Figure 1.9 shows an observation of the day sky (solar spectrum) taken with a modern echelle spectrograph. Because many spectral orders can be recorded simultaneously in a single observation a large wavelength coverage is obtained. Typical echelle spectrographs have a wavelength coverage of $\approx$4000–10,000 Å in one exposure. Often the full range is usually limited by the physical size of CCD detectors. As we will shortly see, having such a large wavelength coverage where we can use hundreds, if not thousands of spectral lines for calculating Doppler shifts is one of the keys to achieving an RV precision of $\approx 1\,\mathrm{m\,s^{-1}}$. Note that because this is a slit spectrograph the absorption lines appear to be images of the slit.

More details about astronomical spectrographs regarding their design, performance, use, etc. can be found in textbooks on the subject and is beyond the scope of these lectures. However there are several aspects that are important in our discussion of the radial velocity method.

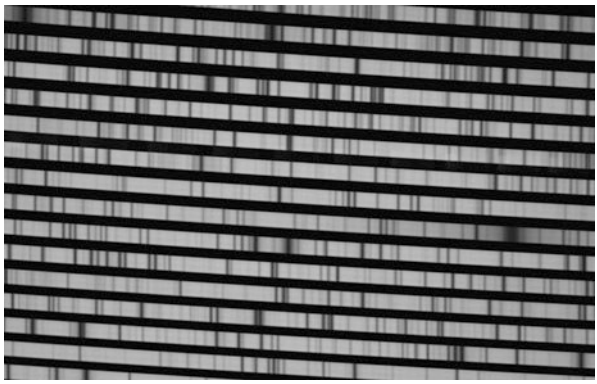

Fig. 1.9 A spectrum of the sun taken with a high resolution cross-dispersed echelle spectrograph

1.2.2.1 Spectral Resolving Power

Consider two monochromatic beams. These will be resolved by your spectrograph if they have a separation of $\delta\lambda$. The resolving power, R, is defined as

$$R = \frac{\lambda}{\delta\lambda} \tag{1.4}$$

Typically, to adhere to the Nyquist sampling (see below) $\delta\lambda$ covers two pixels on the detector. Note that R is a dimensionless number that gets larger for higher spectral resolution. The spectral resolution, $\delta\lambda$, on the other hand has units of a length, say Å, and is smaller for higher resolution. Many people often refer to Eq. (1.4) as the spectral resolution which is not strictly the case. High resolution echelle spectrographs for RV work typically have R = 50,000–100,000.

Let's take an R = 55,000 spectrograph. At 5500 Å this would correspond to a spectral resolution of 0.1 Å. To satisfy the Nyquist criterion the projected slit must fall on at least two detector pixels. This means the dispersion of the spectrograph is 0.05 Å per pixel. The non-relativistic Doppler shift is given by

$$v = \frac{(\lambda - \lambda_0)}{\lambda_0} c \tag{1.5}$$

where c is the speed of light, λ_0 the rest wavelength, and λ the observed wavelength. This means that a one pixel shift of a spectral line amounts to a velocity shift of $3\,\mathrm{km\,s^{-1}}$ which is our "velocity resolution."

1.2.2.2 The Blaze Function

The blaze function results from the interference pattern of the single grooves of the grating. Each facet is a slit, so the interference pattern is a sinc function. Recall from optical interference that the smaller the slit, the broader the sinc function.

For finely grooved gratings ($\approx$1000 grooves/mm) the groove spacing is small, so the blaze (sinc) function is rather broad. In fact in most cases one can hardly notice it in the reduced spectra from data taken with a spectrograph with a finely grooved grating. However, for echelle gratings that have wide facets the blaze function becomes much more narrow.

Figure 1.10 shows an extracted spectral order from an echelle spectrograph that has a rather strong blaze function. This can have a strong influence on your RV measurement, particularly if you use the cross-correlation method (see below). The blaze should be removed from all reduced spectra before calculating RVs. This is done either by dividing your science spectra with the spectrum (which also has the blaze function) of a rapidly rotating hot star, or fitting a polynomial to the continuum of the spectra.

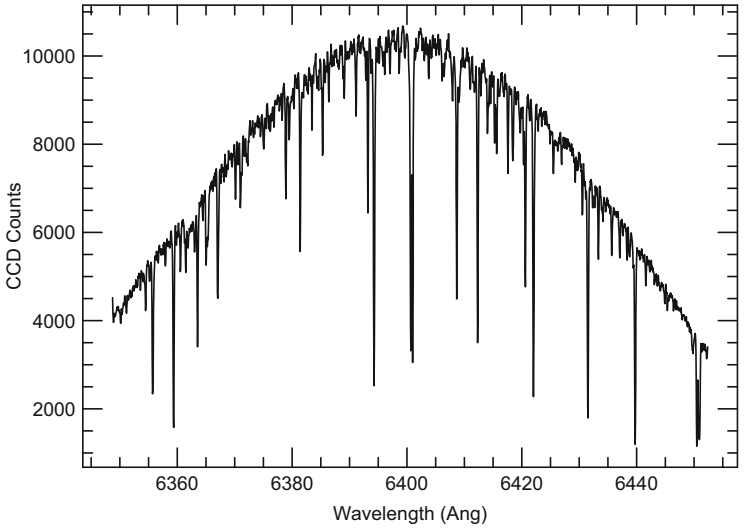

Fig. 1.10 A spectral order from an echelle spectrograph that has not been corrected for the blaze function

1.3 Achieving High Radial Velocity Measurement Precision

We now investigate how one can achieve a high RV measurement precision. This depends on the design of your spectrograph, the characteristics of your spectral data, and even on the properties of the star that you are observing. Important for achieving a high precision is the minimization of instrumental shifts.

1.3.1 Requirements for Precise Radial Velocity Measurements

Suppose that for one spectral line you can measure the Doppler shift with an error of σ. If one then uses $N_{\rm lines}$ for the Doppler measurement, the total error reduces to $\sigma_{\rm total} = \sigma/\sqrt{N_{\rm lines}}$. Thus for a given wavelength bandpass, B, the RV measurement error should be $\propto B^{-1/2}$. This is not strictly the case as some wavelength regions have more or less spectral lines, so the number of lines may not increase linearly with bandwidth, but this is a reasonable approximation.

How does the RV measurement precision depend on the noise level in your spectra? Figure 1.11 shows how the measurement error, σ, varies with the signal-to-noise ratio (S/N) of your spectral data based on numerical simulations. The simulated data had the same resolution and wavelength coverage, only the S/N is changing. These simulations show that $\sigma \propto ({\rm S/N})^{-1}$.

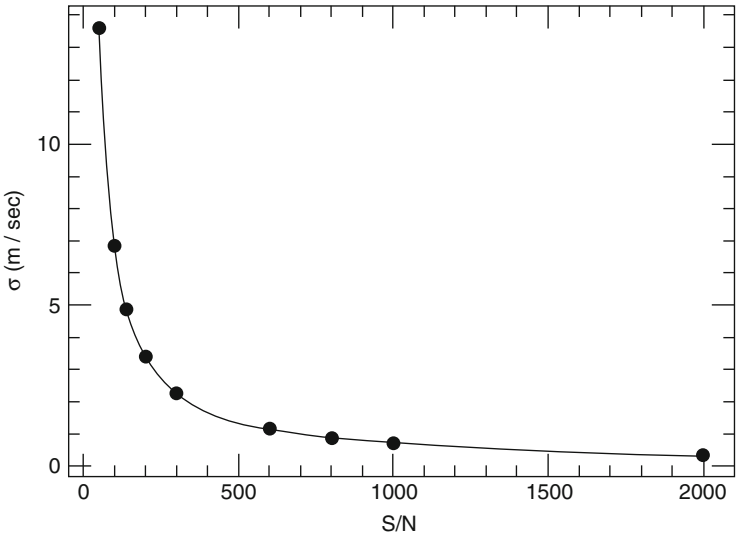

Fig. 1.11 The radial velocity measurement error as a function of signal-to-noise ratio (S/N) of the spectrum

It is worth mentioning that RV surveys on bright stars tend to aim for S/N = 100–200 as there is not a substantial decrease in measurement error by going to higher S/N and these have a high cost in terms of exposure time. Suppose that you had a measurement error of $\sigma = 3\,\mathrm{m\,s^{-1}}$ at S/N = 200 and you wanted to improve that to $\sigma = 1\,\mathrm{m\,s^{-1}}$. This would require S/N = 600. However, for photon statistics S/N = $\sqrt{N_{\mathrm{ph}}}$ where N_{Ph} is the number of detected photons. This means to achieve a S/N = 600 you need to detect nine times the number of photons which requires nine times the exposure time. If you achieve S/N = 200 in 15 min, you would have to observe the star for more than 2 h to get a factor of three reduction in measurement error. If this star did not have a planetary companion, you would have wasted a lot of precious telescope time. It is better to work at the lower S/N and include more stars in your program.

How does the velocity error depend on the resolving power, R? For higher resolving power each CCD pixel represents a smaller shift in radial velocity. For a given S/N if you measure a position of a line to a fraction of a pixel this corresponds to a smaller velocity shift. Thus we expect $\sigma_R \propto R^{-1}$, where the subscript "R" refers to the σ due purely to the increased resolving power. The RV error should be smaller for high resolution spectrographs. However, for a fixed-size CCD a higher resolving power (i.e., more dispersion) means that a smaller fraction of the spectral region will now fall on the detector. The band pass, B, and thus the number of spectral lines are proportional to R^{-1}, and $\sigma_B \propto B^{-1/2} = R^{1/2}$ where the subscript "B" denotes the σ due to the smaller bandpass of the higher resolution data. The final sigma for a fixed-sized detector, σ_D, is proportional to the product $\sigma_R \times \sigma_B \propto R^{-1/2}$. What

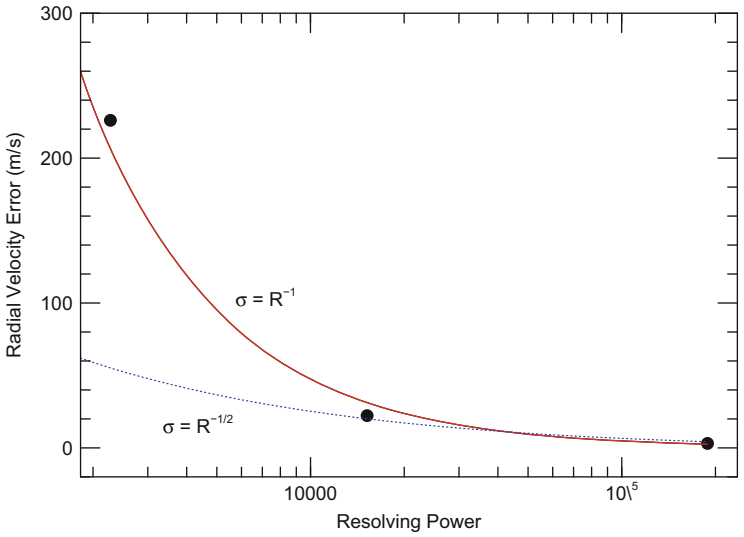

Fig. 1.12 (Points) The radial velocity error taken with a spectrograph at different resolving powers. This is actual data taken of the day sky all with the same S/N values. The *solid red line* shows a $\sigma \propto R^{-1}$ fit. The *dashed blue line* shows a $\sigma \propto R^{-1/2}$ fit. The detector size is fixed for all data, thus the wavelength coverage is increasing with decreasing resolving power

we gain by having higher resolving power is partially offset by less wavelength coverage.

Figure 1.12 shows actual RV measurements taken of the day sky (i.e. a solar spectrum) at several resolving powers: $R = 2300$, 15,000, and 200,000. The two curves show $\sigma \propto R^{-1/2}$ and $\sigma \propto R^{-1}$. The data more closely follow the $\sigma \propto R^{-1}$ curve. This indicates the RV error strictly from the resolving power should be $\sigma_R \propto R^{-3/2}$, i.e. the error of the data, $\sigma_D = \propto R^{-3/2}\,(\sigma_R) \times R^{1/2}\,(\sigma_B) = R^{-1}$.

Putting all this together we arrive at an expression which can be used to calculate the predicted RV measurement error for spectral data:

$$\sigma[\mathrm{m/s}] = C(\mathrm{S/N})^{-1} R^{-3/2} B^{-1/2} \tag{1.6}$$

where (S/N) is the signal-to-noise ratio of the data, R is the resolving power ($= \lambda/\delta\lambda$) of the spectrograph, B is the wavelength coverage in Å of the stellar spectrum used for the RV measurement, and C is a constant of proportionality.

The value of the constant can be estimated based on the performance of various spectrographs (Table 1.1). Plotting the quantity measurement error, σ, versus $R^{-3/2}B^{-1/2}$ (for data with the same S/N) shows a tight linear relationship with a slope $C = 2.3{\times}10^9$. With this expression one should be able to estimate the expected RV precision of a spectrograph to within a factor of a few.

Table 1.1 lists several spectrographs I have used for RV measurements. The table lists the wavelength range used for the RV measurements, the resolving power of

Table 1.1 The predicted RV measurement error ($\sigma_{\rm predicted}$) compared to the actual values ($\sigma_{\rm actual}$), measured over a wavelength range, $\Delta\lambda$, for various spectrographs of different resolving powers, R

Spectrograph	$\Delta\lambda$ (Å)	R $\delta\lambda/\lambda$	$\sigma_{\rm predicted}$ (m/s)	$\sigma_{\rm actual}$ m/s
McDonald CS11	9	200,000	8	10
McDonald tull	800	60,000	5	5
McDonald CS21	400	180,000	2	4
McDonald sandiford	800	50,000	7	10
TLS coude echelle	800	67,000	5	5
ESO CES	43	100,000	11	10
Keck HIRES	800	80,000	3	3
HARPS	2000	110,000	1.4	1

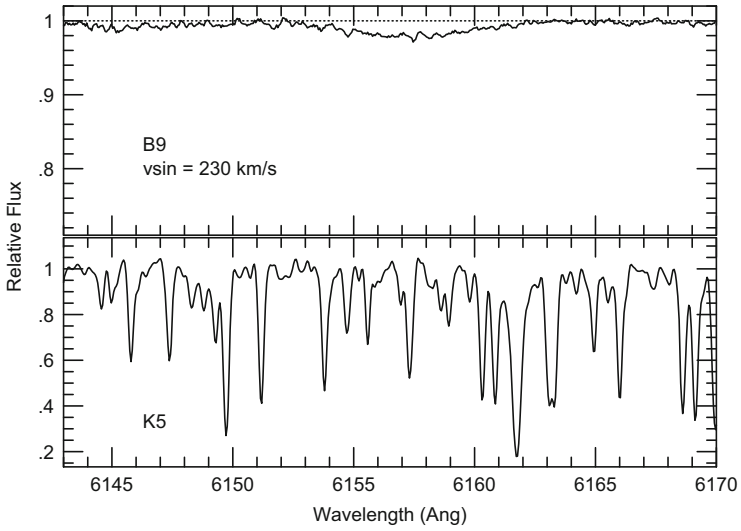

Fig. 1.13 (*Top*) A spectrum of a B9 star rotating at 230 km s^{-1}. (*Bottom*) The spectrum of a K5 star

the spectrograph, the measured RV precision, and the predicted value. All with the exception of the HARPS spectrograph used an iodine gas absorption cell (see below). The table shows that Eq. (1.6) can be used to estimate the RV measurement error of a spectrograph to within a factor of two.

1.3.2 Influence of the Star

The resolution and S/N of your data are not the only factors that influence the RV measurement precision. The properties of the star can have a much larger effect, as not all stars are conducive to precise RV measurements. Figure 1.13 shows the spectral region of two stars, the top of a B9 star and the lower panel of a K5 star. The hot star only shows one spectral line in this region and it is quite broad and

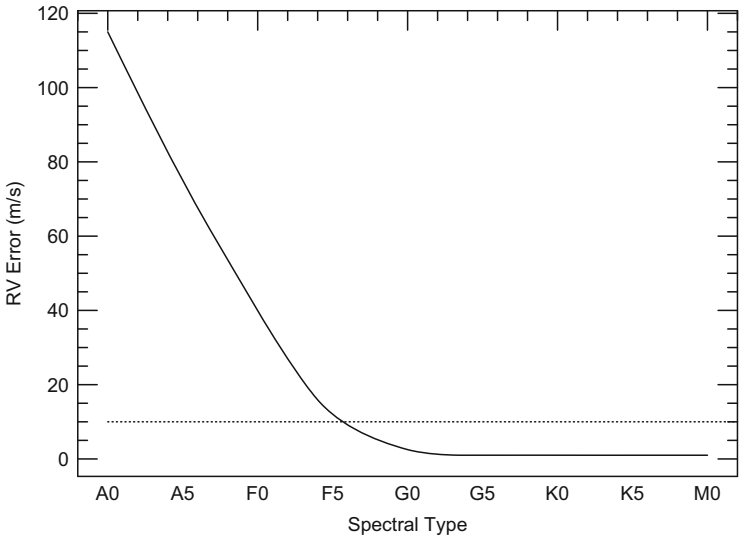

Fig. 1.14 The expected radial velocity error as a function of spectral type. This was created using the mean rotational velocity and approximate line density for a star in each spectral type. The *horizontal line* marks the nominal precision of $10\,\mathrm{m\,s^{-1}}$ needed to detect a Jovian-like exoplanet

shallow due to the high projected rotation rate of the star, $v \sin i_*$,[1] $\approx 230\,\mathrm{km\,s^{-1}}$. It is difficult to determine the centroid position and thus a Doppler shift of this spectral line. On the other hand, the cooler star has nearly ten times as many spectral lines in this wavelength region. More importantly, the lines are quite narrow due to the slow rotation of the star.

The approximate RV error as a function of stellar spectral type is shown in Fig. 1.14. The horizontal dashed line marks the nominal $10\,\mathrm{m\,s^{-1}}$ needed to detect a Jovian planet at 5 AU from a sun-like star. Later than about spectral type F6 the RV error is well below this nominal value. For earlier spectral types the RV error increases dramatically. This is due to two factors. First, the mean rotation rate increases for stars later than about F6. This spectral type marks the onset of the outer convection zone of star which is the engine for stellar magnetic activity, and this becomes deeper for cooler stars. It is this magnetic activity that rapidly brakes the star's rotation.

Second, the number of spectral lines in the spectrum decreases with increasing effective temperature and thus earlier spectral type. The mean rotation rate and the approximate spectral line density of stars as a function of stellar types were used to

[1] In this case i_* refers to the inclination of the stellar rotation axis. This is not to be confused with the i that we later use to refer to the inclination of the orbital axis. The two are not necessarily the same value.

produce this figure. This largely explains why most RV exoplanet discoveries are for host stars later than spectral type F6.

Figure 1.14 indicates that we need to modify Eq. (1.6) to take into account the properties of the star. Simulations show that the RV measurement error, $\sigma \propto v \sin i$ in km/s. This means that if you obtain an RV precision of $10\,\mathrm{m\,s^{-1}}$ on a star that is rotating at $4\,\mathrm{km\,s^{-1}}$, then you will get $100\,\mathrm{m\,s^{-1}}$ on a star of the same spectral type that is rotating at $40\,\mathrm{km\,s^{-1}}$, using data with the same spectral resolution and S/N.

One must also account for the change in the line density. A G-type star has roughly ten times more spectral lines than an A-type star and thus will have approximately one-third the measurement error for a given rotational velocity. An M-dwarf has approximately four times as many lines as a G-type star resulting in a factor of two improvement in precision. We can therefore define a function, $f(\mathrm{SpT})$, that takes into account the spectral type of the star. As a crude estimate, $f(\mathrm{SpT}) \approx 3$ for an A-type star, 1 for a G-type, and 0.5 for an M-type stars. Equation (1.6) can then be modified to include the stellar parameters:

$$\sigma[\mathrm{m/s}] = C(\mathrm{S/N})^{-1} R^{-3/2} B^{-1/2} (v \sin i/2) f(\mathrm{SpT}) \tag{1.7}$$

where v sin i is the rotational velocity of the star in $\mathrm{km\,s^{-1}}$ scaled to the approximate value for the star. Note this is for stars rotating faster than about $2\,\mathrm{km\,s^{-1}}$. For a star with a lower rotational velocity simply eliminate the $v \sin i$ term.

1.3.3 Eliminating Instrumental Shifts

Suppose you have designed a spectrograph for precise RV measurements with the goal of finding exoplanets. You have taken great care at optimizing the resolving power, wavelength range, and efficiency of the spectrograph. You start making RV measurements but you find that the actual error of your measurements is far worse than the predicted value.

The problem is that Eq. (1.6) does not take into account any instrumental shifts. A spectrograph detector does not record any wavelength information, it merely records the intensity of light as a function of pixel location on the CCD. To calculate a Doppler shift one needs to convert the pixel location of a spectral line into an actual wavelength. This can be done by observing a calibration lamp, typically a Th-Ar hollow cathode lamp. You then identify thorium emission lines whose wavelengths have been measured in the laboratory and mark their pixel location. A function, typically a high order polynomial, is fit to the pixel versus wavelength as determined from the identified thorium emission lines. This function is assumed to be valid for the spectrum of the star. The problem is that this calibration observation is taken at a different time (either before or after the stellar spectrum) and the light goes through a different optical path. The mechanical shifts of the detector between calibration exposures can be quite large making it impossible to achieve a precision sufficient to detect planets.

A Doppler shift of a spectral line will result in a physical shift, in pixels, at the detector. Let's calculate how large that would be for an $R = 100{,}000$ spectrograph. At this resolving power the resolution will be 0.05 Å at 5000 Å. For a well-designed spectrograph this should fall on two pixels of the detector which corresponds to a dispersion of 0.025 Å/pixel. By our Doppler formula that is a velocity resolution of $1500\,\mathrm{m\,s^{-1}}$ per CCD pixel. Thus a $10\,\mathrm{m\,s^{-1}}$ Doppler shift caused by a Jovian analog will create a shift of a spectral line of 0.0067 pixels. A typical CCD pixel has a size of about $15\,\mu\mathrm{m}$, so the shift of the spectral line will be 10^{-4} cm in the focal plane. A $1\,\mathrm{m\,s^{-1}}$ Doppler shift, a value easily obtained by modern methods, results in a physical shift of the spectral line at the detector of 10^{-5} cm or about one-fifth of the wavelength of the incoming light. It is unlikely that the position of the detector is stable to this level. It will not take much motion of the instrument or detector (e.g., rotation of the dome, vehicles driving past outside, small earthquakes, etc.) to cause an apparent positional shift of spectral lines at the detector.

Figure 1.15 shows the instrumental shifts of a spectrograph once used at McDonald Observatory. These show a peak-to-peak change of about $50\,\mathrm{m\,s^{-1}}$ in $\approx$2 h. The instrumental "velocity" can also change by this amount in only 2 min. The RMS scatter of the data is $27\,\mathrm{m\,s^{-1}}$. The predicted RV error according to Eq. (1.6) should be $\approx 10\,\mathrm{m\,s^{-1}}$. This means that the measurement errors are dominated by these instrumental errors. If you had not taken care to make your spectrograph stable, both mechanically and thermally, your instrument would have a difficult time detecting giant exoplanets.

The key to eliminating these instrumental shifts is to record *simultaneously* the calibration and the stellar spectra. Three methods have been used to do this.

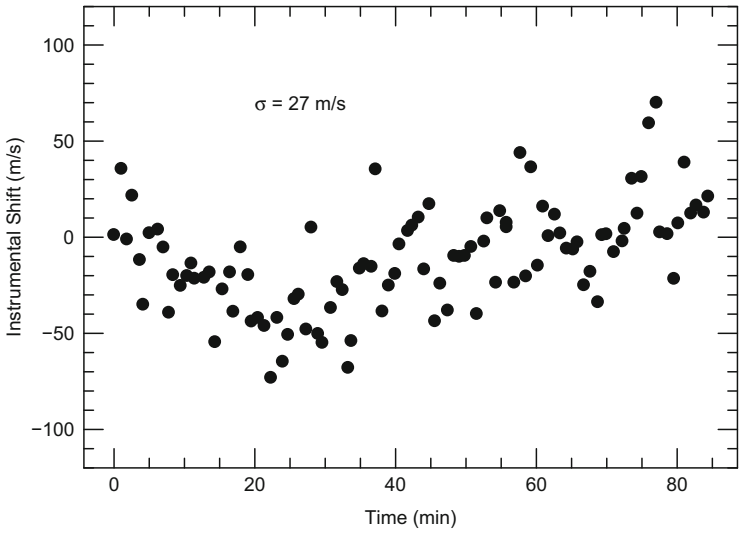

Fig. 1.15 The measured instrumental shifts of the coude spectrograph at McDonald Observatory as a function of time

1.3.3.1 The Telluric Method

The simplest (and cheapest) way to minimize instrumental shifts is simply to use telluric absorption features imposed on the stellar spectrum as the starlight passes through the Earth's atmosphere. Doppler shifts of the stellar lines are then measured with respect to the telluric features. Since instrumental shifts affect both features equally, a higher RV precision is achieved.

Griffin and Griffin (1973) first proposed the telluric method using the O_2 features at 6300 Å. They suggested that an RV precision of 15–20 m s^{-1} was possible with this method. Cochran et al. (1991) confirmed this in using telluric method to confirm the planetary companion to HD 114762. This method can also be extended to the near infrared spectral region using the telluric A (6860–6930 Å) and B (6860–6930 Å) bands (Guenther and Wuchterl 2003).

Although the method is simple and inexpensive to implement it has the big disadvantage in that it covers a rather limited wavelength range. More of a problem is that pressure and temperature changes, as well as winds in the Earth's atmosphere ultimately limit the measurement precision. You simply cannot control the Earth's atmosphere! It is probably difficult to achieve an RV precision better than about 20 m s^{-1} with this method.

1.3.3.2 The Gas Cell Method

An improvement to the telluric method could be achieved if somehow one could control the absorbing gas used to create the wavelength reference. This is the principle behind the gas absorption cell. A chemical gas that produces absorption lines not found in the stellar spectrum or the Earth's atmosphere is placed in a glass cell that is sealed and temperature stabilized. This cell is then placed in the optical path of the telescope, generally before the entrance slit to the spectrograph. The gas cell will then create a set of stable absorption lines that are superimposed on the stellar spectrum and these provide the wavelength reference.

In pioneering work, Campbell and Walker (1979) first used a gas cell for planet detection with precise RV measurements. They used the 3-0 band R branch of Hydrogen-Fluoride (HF) at 8670–8770 Å to provide the velocity metric. With this method Campbell and Walker achieved an RV precision of 13 m s^{-1} in 1979. Their RV survey first found evidence for the giant planet around γ Cep A (Campbell et al. 1988) that was ultimately confirmed by subsequent measurements using the iodine method (Hatzes et al. 2003).

Although the HF method was able to achieve good results and discover exoplanets, it suffered from several drawbacks: (1) the absorption features of HF only covered about 100 Å, so relatively few spectral lines could be used for the Doppler measurement. (2) HF is sensitive to pressure shifts. (3) To produce suitable HF absorption lines a large path length ($\approx$1 m) for the cell was required. This could present problems if your spectrograph had space restrictions. (4) The cell has to be filled for each observing run because HF is a highly corrosive and dangerous gas.

The process of filling the cell or breakage during its use would present a serious safety hazard to the astronomer and the telescope staff.

A safer alternative to the HF is to use molecular iodine (I_2) as the absorbing gas. There are several important advantages in the use of I_2:

1. It is a relatively benign gas that can be permanently sealed in a glass cell.
2. I_2 has useful absorption lines over the interval 5000–6000 Å or a factor of 10 larger than for HF.
3. A typical path length for an I_2 cell is about 10 cm so the cell can easily fit in front of the entrance slit of most spectrographs.
4. The I_2 cell can be stabilized at relatively modest temperatures (50–70 °C).
5. Molecular iodine is less sensitive to pressure shifts than HF.
6. The rich density of narrow I_2 absorption lines enables one to model the instrumental profile of the spectrograph (see below).
7. Molecular iodine presents no real health hazard to the user.

Figure 1.16 shows a spectrum of iodine, plus of a target star taken with and without the cell in the light path.

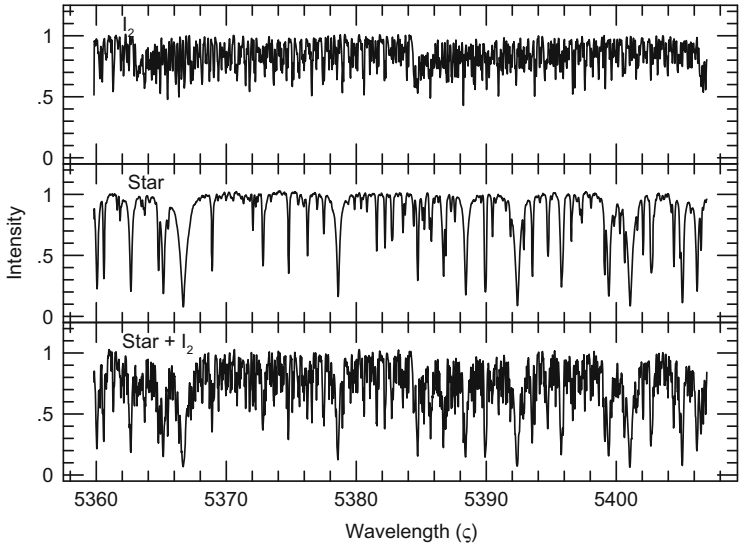

Fig. 1.16 (*Top*) A spectrum of molecular iodine in the 5360–5410 Å region. (*Center*) A spectrum of a sun-like star in the same spectral region. (*Bottom*) A spectrum of the star taken through an iodine absorption cell

1.3.3.3 Simultaneous Th-Ar

You may ask, why not simply eliminate these instrumental shifts by recording your standard hollow cathode calibration spectrum at the same time as your stellar observation? The advent of fiber optics allows you to do this. One fiber optic is used to feed light from the star into the spectrograph, and a second to feed light from a calibration lamp. Thus the calibration spectrum is recorded on the CCD detector adjacent to the stellar spectrum so any instrumental shifts will affect both equally. This technique is best exemplified by the high accuracy radial velocity planetary searcher (HARPS) spectrograph (Pepe et al. 2000).

Figure 1.17 shows a stellar spectrum recorded using the simultaneous Th-Ar technique. This spectrum was recorded using the HARPS spectrograph of ESO's 3.6 m telescope at La Silla. The continuous bands represent the stellar spectrum. In between these one can see the emission spectrum from the Th-Ar fiber. If you look carefully, you will see that the images of the thorium emission lines have a circular shape because they are an image of the fiber.

There are several advantages and disadvantages to using simultaneous Th-Ar calibration.

Advantages:

1. No starlight is lost via absorption by the gas in a cell.
2. There is no contamination of the spectral lines which makes analyses of these (e.g., line shapes) much easier.
3. It can be used over a much broader wavelength coverage ($\approx$2000 Å).
4. Computation of the RVs is more straightforward (see below).

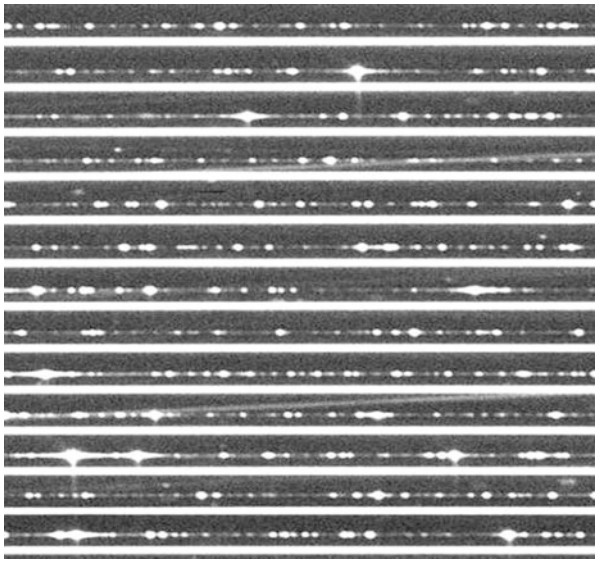

Fig. 1.17 A spectrum recorded with the HARPS spectrograph. The solid bands are from the star fiber. The emission line spectrum of Th-Ar above the stellar one comes from the calibration fiber. Note the contamination of the stellar spectrum with strong Th lines in the *lower left* and *upper center*

Disadvantages:

1. For Th-Ar lamps you have to apply a high voltage to the lamp in order to produce emission lines. Slight changes in the voltage may result in changes in the emission spectrum of the lamp.
2. Contamination by strong Th-Ar lines that spill light into adjacent orders. This can be seen in Fig. 1.17. This contamination is not easy to model out.
3. Th-Ar lamps age, change their emission structure, and eventually die. Using a new Th-Ar lamp will introduce instrumental offsets compared to the previous data taken with the old cell.
4. The wavelength calibration is not recorded in situ to the stellar spectrum, but rather adjacent to it. One has to have faith that the same wavelength calibration applies to different regions of the detector.
5. You cannot model any changes in the instrumental profile of the spectrograph (see below).

1.3.4 Details on Calculating the Doppler Shifts

Here I briefly describe how one calculates Doppler shifts with the various methods.

1.3.4.1 Simultaneous Th-Ar

Calculating Doppler shifts with the simultaneous Th-Ar method can be computationally simple. The standard tool for calculating these shifts is the cross-correlation function. If $s(x)$ is your stellar spectrum as a function of pixels, and $t(x)$ is a template spectrum, then the cross-correlation function is defined as

$$\mathrm{CCF}(\Delta x) = s(\Delta x) \otimes t(\Delta x) = \int_{-\infty}^{+\infty} s(x)t(x + \Delta x)dx \tag{1.8}$$

Since we are dealing with discretely sampled spectra whose CCFs are calculated numerically we use the discrete form of the CCF:

$$\mathrm{CCF}(\Delta x) = \sum_{x=1}^{N} s(x)t(x + \Delta x)dx \tag{1.9}$$

Δx is called the lag of the CCF. The CCF is most sensitive to Δx when the s and t are identical. The CCF will be a maximum for a Δx that matches both functions. For this reason the CCF is often called a matching or detection filter.

Unfortunately, the Doppler formula [Eq. (1.5)] is non-linear which means at different wavelengths the Doppler shift in pixels will be different. This can be remedied by re-binning the linear wavelength scale onto a logarithmic one

transforming the Doppler formula to

$$\Delta \ln \lambda = \ln \lambda - \ln \lambda_0 = \ln \left(1 + \frac{v}{c}\right) \tag{1.10}$$

The lag of the CCD is then $\Delta x = \Delta \ln \lambda = \ln(1 + v/c)$ which is a constant for a given velocity, v.

A variety of template spectra, $t(x)$, have been employed. The key to having a CCF that produces a good velocity measure is to have a template that matches as closely as possible your stellar spectra and that has a high S/N ratio. In the searching for exoplanets one is primarily interested in changes in the star's velocity, or relative RVs. In this case you can simply take one observation of your star as a template and cross-correlate the other spectra of the star with this one. An advantage of using an actual observation of your star as the template guarantees that you have an excellent match between star and template. However, this may introduce noise into the CCF. Alternatively you can co-add all of your observations to produce one master, high signal-to-noise ratio template. You can also use a synthetic spectrum. A common practice, as employed by the HARPS pipeline, is to use a digital stellar mask that is noise-free. This is a mask that has zero values except at the location of spectral lines. A different mask must be used for each spectral type that is observed. If you are interested in obtaining an absolute RV measure, you should use the spectrum of a standard star of known RV as your template. This standard star should have a spectral type near that of your target star.

The CCF can be quite sensitive to Doppler shifts even with low S/N data. Figure 1.18 shows a synthetic spectrum of a single spectral line generated using different values of S/N (= 100, 10, and 5). Each of these have been Doppler shifted by $+20\,\mathrm{km\,s^{-1}}$. These were cross-correlated with a noise-free synthetic spectral line. One can clearly see the CCF peak and appropriate Doppler shift even when the S/N is as low as five.

Since the CCF is a matching filter one can see why it is wise to remove the blaze function from your spectral data. If your template star also has the blaze function in it, then the CCF will always produce a peak at zero simply because you are matching the blaze functions of the two spectra and not the spectral features. If only the stellar spectrum has the blaze function, then this will distort the CCF and severely compromise the RV measurement precision.

Fortunately for astronomers you often do not have to write your own code to calculate CCFs for RV measurements. The HARPS data reduction pipeline produces the CCF and RV measurement as part of the observing process. The Terra-HARPS package is an alternative CCF pipeline to the standard HARPS reduction (Anglada-Escudé and Butler 2012). It uses a master, high signal-to-noise stellar template by co-adding all the stellar observations before computing the final CCF. The image reduction and analysis facility (IRAF) has packages for calculating Doppler shifts, most notably the package *fxcor*.

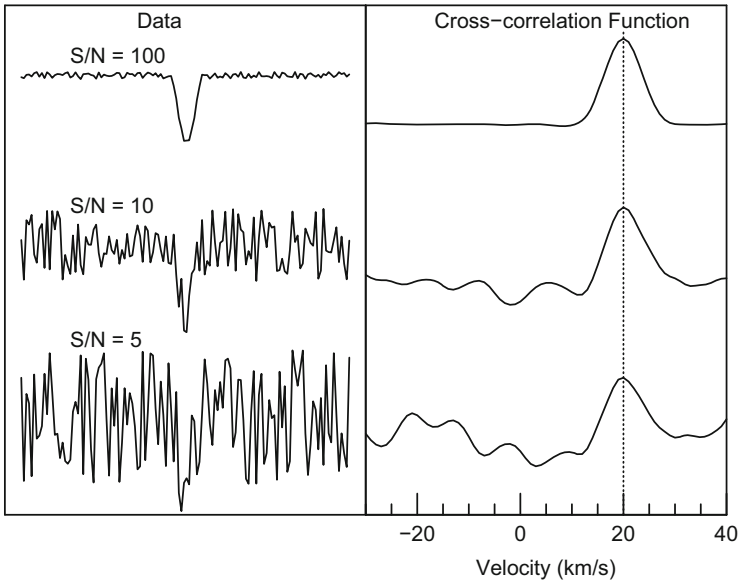

Fig. 1.18 Example of the use of the cross-correlation function (CCF) to detect a Doppler shift of a spectral line. (*Left*) Synthetic spectra generated at three levels of signal-to-noise (S/N = 100, 10, and 5) and with a Doppler shift of +20 km s^{-1}. (*Right*) The CCF of the noisy spectra cross-correlated with with a noise-free synthetic spectral line. Even with the noisy spectra the correct Doppler shift is recovered (*vertical dashed line*)

1.3.4.2 The Iodine Method

Calculation of Doppler shifts with the iodine method, when done correctly, is computationally more intensive since one does not use a simple cross-correlation function. With this method one actually models the observed spectrum taking into account changes in the so-called instrumental profile (IP).

The IP represents the instrumental response of your spectrograph. If you observed a monochromatic light source (i.e., ∂-function) with a perfect spectrograph, you would record a ∂-function of light at the detector. Real spectrographs will blur this ∂-function. For a well-designed spectrograph this blurring function is a symmetrical Gaussian whose full width at half maximum falls on the 2 pixel Nyquist sampling of the detector.

The problem for RV measurements is if the IP changes. The left side of Fig. 1.19 shows an IP that is an asymmetric Gaussian. The centroid of this Gaussian is +0.17 pixels, or a velocity shift of +250 m s^{-1} for an R = 100,000 spectrograph with respect to a symmetrical IP (dashed line). Since the stellar spectrum is convolved with this IP, this asymmetric shape will be imposed on all spectral lines.

For the RV detection of planets it does not matter if the IP in this case is asymmetric and that it introduces an RV shift of +250 m s^{-1} in all spectral lines. This is an absolute shift and we are only interested in relative shifts. So long as

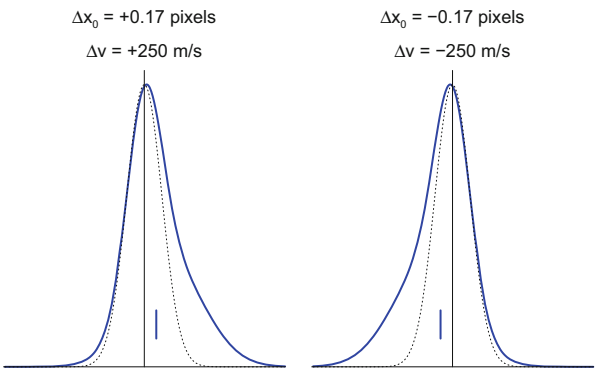

Fig. 1.19 (*Left*) The *solid line* is an asymmetric instrumental profile (IP). The *dashed line* is symmetrical IP profile. The *vertical line* represents the centroid of the asymmetric IP. It is shifted by +0.17 pixels or +250 m/s for an R = 100,000 spectrograph. (*Right*) An IP profile that is asymmetric towards the *blue side*. Such change in the IP would introduce a total instrumental shift of $\Delta v = 500\,\mathrm{m\,s^{-1}}$ in the RV measurement (Color figure online)

the IP does not change and the $+250\,\mathrm{m\,s^{-1}}$ is *always* the same this will not cause problems.

The problem arises if the IP were to change into the one on the right. In this case the Gaussian is asymmetric towards the blue and the centroid of each spectral line would shift by −0.17 pixels with respect to the centroid from a symmetric IP. One would thus measure a total change in relative shift in the velocity of the star by $500\,\mathrm{m\,s^{-1}}$ compared to measurements taken with the previous IP. This velocity change is not from the star, but from changes in the shape of the IP.

A tremendous advantage of the iodine method over the simultaneous Th-Ar method is that one can use information in the iodine lines to model the IP. This can be done because iodine lines are unresolved even at a resolving power of R = 100,000. Thus they carry information about the IP of the spectrograph. This is not the case for thorium emission lines from a hollow cathode lamp that have an intrinsic width comparable, if not greater than the width of the IP.

The "trick" to measuring the IP is to take a very high resolution (R = 500,000–1,000,000), high signal-to-noise spectrum of a white light source taken through the cell using a Fourier transform spectrometer (FTS). An FTS is used as these spectrographs provide some of the highest resolutions possible. One then rebins the FTS iodine spectrum to the same dispersion as your observation, typically taken with R = 60,000–100,000. This binned FTS spectrum is how an iodine cell observation would look like if your instrument IP was a pure ∂-function. One then finds a model of the IP that when convolved with the FTS iodine produces the iodine spectrum observed with your spectrograph. Figure 1.20 shows a section of the iodine spectrum taken with the coude spectrograph of the 2.7 m telescope at McDonald Observatory (dots). The dashed line is the FTS iodine spectrum binned

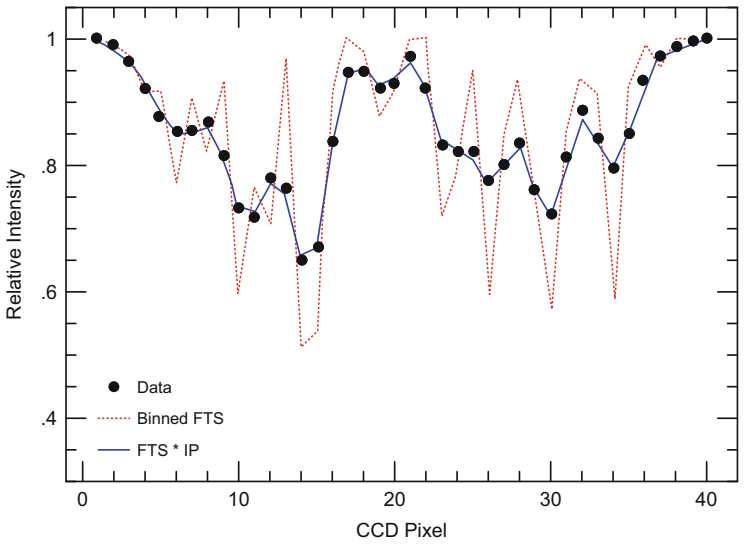

Fig. 1.20 (*Dots*) Measurements of the spectrum of I_2 over a short wavelength region. (*Dashed line*) Spectrum of I_2 taken at high resolution using an FTS and binned to the dispersion of the data. (*Solid line*) The FTS I_2 spectrum convolved with the model IP

to the dispersion of the observation. The solid line shows the binned FTS spectrum convolved with a model IP.

For a good RV precision one needs a good model of the IP. Most programs parameterize the IP as a sum of several Gaussian components following the procedure of Valenti et al. (1995). Gaussian profiles are chosen because the IP is to first order a Gaussian profile and the addition of satellite Gaussian components makes it easy to introduce asymmetries.

Figure 1.21 illustrates the IP parameterization process using one order from a spectrum taken with the coude echelle spectrograph on the Tautenburg 2 m Alfred Jensch Telescope. The IP in this case is modeled by a central Gaussian (red line) plus four satellite Gaussians. The black line (tall Gaussian-like profile) represents the sum of the Gaussians. When modeling the IP this is over-sampled by 5 pixels, so there are five IP pixels for every CCD pixel, often called "IP space" (Valenti et al. 1995).

One problem is that the IP can change across the spectral format, even across a single spectral order. Figure 1.21 shows the IP for the first 200 pixels (blueward), central 800–1000 pixels, and last 1800–2000 pixels (redward) of a spectral order modeled using the FTS iodine spectrum. One can see that the IP changes significantly as one moves redward along the spectral order becoming more "flat-topped." Recall that each spectral line at some level represents an image of the slit, which is a square function, so we should expect that the IP could have a slightly flat top. To model this feature with a sum of Gaussian functions requires two strong satellite components which results in the slight dip in the center of the IP. Note that this IP

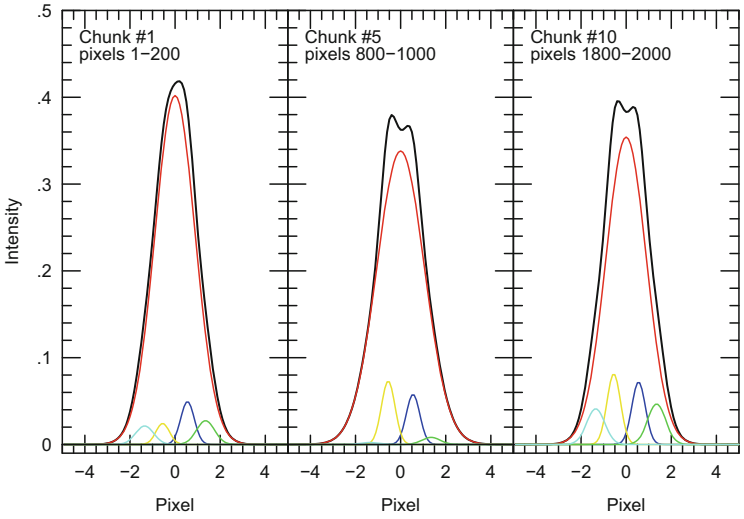

Fig. 1.21 A model of the IP using multi-Gaussian profiles. The smaller Gaussians represent satellite Gaussians that are combined with a larger central Gaussian shown in *red*. The *black line* represents the final IP (the tallest Gaussian) that is a sum of the central plus satellite Gaussians. The model IP is shown for the *blueside* (*left*), central (*center*), and *redside* (*right*) of the spectral order. The IP was calculated using 200 chunk pixels of an I_2 spectrum. Note that one pixel on the CCD ("detector" space) corresponds to five pixels in "IP space"

is five times over-sampled. When we rebin to the data sampling this dip disappears and we would have a flat-topped profile.

For calculating radial velocities using the iodine method requires three key ingredients:

1. A very high resolution spectrum of a white light source taken through your iodine cell using an FTS. This is your *fiducial*.
2. A high resolution spectrum of your star taken without the iodine cell. This is your *template*.
   This is a bit tricky because in the reduction process this stellar template spectrum will be convolved with our model IP. The problem is that the stellar spectrum was taken with your spectrograph so it already has been convolved with the IP. In the reduction process it will again be convolved with your model IP. You do not want to convolve your stellar spectrum twice!
   For the highest RV precision you should deconvolve the IP from the template spectrum before using it to calculate the RV. The IP used for the stellar deconvolution can be measured from an observation of a white light source taken through the cell with your spectrograph. This should be taken as close in time to the template spectrum so that there is little change in the IP. Alternatively, and the better way, is to take spectrum of a hot, rapidly rotating early type star through the cell prior to your stellar observation. As we have seen these stars have

few spectral features that are quite shallow due to the rotation of the star. What you observe is essentially the spectrum of iodine. This has the advantage in that the fiducial is taken through the same optical path as the template observation. This deconvolved spectrum must be numerically over-sampled by a factor of five using interpolation because the convolution will take place in the over-sampled "IP space."
3. A spectrum of your star taken through the cell and for which you want to calculate a Doppler shift.

You then divide each spectral order into 10–40 chunks. The exact number of chunks depends on how rapidly the IP changes along a spectral order. The radial velocity calculation is an iterative process where you solve the equation (Butler et al. 1996):

$$I_{\mathrm{m}} = k[T_{I_2}(\lambda) I_S(\lambda + \partial\lambda)] * \mathrm{IP} \tag{1.11}$$

where I_{S} is the intrinsic stellar spectrum, T_{I_2} is the transmission function of the I_2 cell, k is a normalization factor, $\partial\lambda$ is the wavelength (Doppler) shift, IP is the spectrograph instrumental profile, and $*$ represents the convolution.

In each of these spectral chunks you:

1. Remove any slope in the continuum. Given the small size of the chunks this can be done with a linear function. This requires two parameters.
2. Calculate the dispersion (Angstroms/pixel) in the chunk. Again, due to the relatively small chunks it is sufficient to use a second order polynomial that has three parameters.
3. Calculate the IP containing 5–10 (more can be used if needed) Gaussians. Each Gaussian (except for the central one) has variable positions, amplitudes, and widths. In principle allowing all the Gaussian parameters to vary may be too computationally intensive and the process may not converge. In practice the positions of the satellite Gaussians are fixed. For a five Gaussian fit to the IP a total of nine parameters are required, the width of central Gaussian and the widths and amplitudes of four satellite Gaussians. In each iteration these nine parameters are changed. The other Gaussian parameters remain fixed.
4. Apply a Doppler shift to your template spectrum. This is one parameter.
5. Combine the FTS iodine spectrum and the Doppler shifted template spectrum and convolve this with the IP produced in step 4. You compare this to your observed data by calculating the reduced χ^2. If you have not converged to the desired χ^2, go back to step 1 using the current values as your starting point and vary all the parameters.

So in the end we are determining a total of 15 parameters when all we care about is one, the Doppler shift!

Figure 1.22 shows radial velocity measurements of the RV constant star τ Cet and that demonstrate the need for doing the IP modeling (Endl et al. 2000). These data were taken with an iodine cell mounted at the CES spectrograph of the Coude

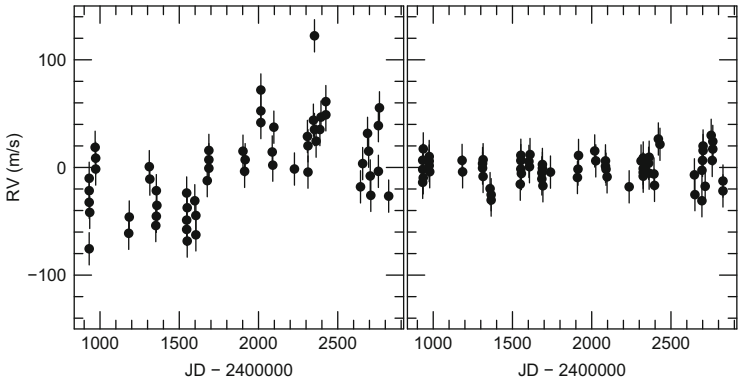

Fig. 1.22 (*Left*) Radial velocity measurements of τ Cet taken with the former CES at La Silla, Chile. No IP modeling was done in calculating the Doppler shifts. (*Right*) The calculated Doppler shifts using the same data but with IP modeling (Endl et al. 2000)

Auxiliary Telescope (both have since been decommissioned). The left panel shows the Doppler shifts calculated without IP modeling. There is a clear sine-like trend which mimics the Keplerian orbit of a planet. The scatter for these measurements is $27\,\mathrm{m\,s^{-1}}$. The right panel shows the Doppler shifts calculated with the inclusion of IP modeling. The sine-like trend has disappeared and the scatter has been reduced to $13\,\mathrm{m\,s^{-1}}$.

For more details about calculating RVs with the iodine method can be found in the literature (Valenti et al. 1995; Butler et al. 1996; Endl et al. 2000).

One disadvantage of the Th-Ar method is that one cannot use it to measure changes in the instrumental profile of the spectrograph (see below). This method implicitly relies on a very stable spectrograph and an IP that does not change with time. For these reasons the HARPS spectrograph was designed with thermal and mechanical stability in mind (Pepe et al. 2000).

HARPS is a state-of-the-art spectrograph specifically designed to achieve very high precision. It is housed in a vacuum chamber that keeps the pressure below 0.01 mbar and the temperature constant at 17 °C to within 0.01 °C. Care was also taken to minimize mechanical vibrations. The thermal and mechanical stability of the spectrograph insures that the IP remains constant. A key improvement to the HARPS spectrograph is the use of two sequential fiber optics in a double scrambler mode to ensure a stable illumination of the spectrograph that is insensitive to variations due to seeing and guiding errors. With such stability HARPS has been able to achieve a short term precision better than $1\,\mathrm{m\,s^{-1}}$ in the best cases.

1.4 Time Series Analysis: Finding Planets in the RV Data

You have acquired enough high precision RV measurements of a star, now comes the task of finding a possible periodic signal in the data. One needs to do this first before one can even fit a Keplerian orbit to the data.

Finding planets in your RV data consists of four major steps:

1. Finding a periodic signal in your data.
2. Determine if the signal is significant, i.e. whether noise is actually creating the signal.
3. Determine the nature of the variations, whether they are due to instrumental effects, intrinsic stellar variability, or a bona fide planet.
4. Fit a Keplerian orbit to the RV measurements.

Sometimes one can find a periodic signal in your data simply by looking at your time series. This only works if the amplitude of your signal is large and the sampling is good. However, if the amplitude of the signal is comparable to the measurement error, the signal is not so easy to detect by eye. The top panel of Fig. 1.23 shows a segment of a sampled sine wave that has noise added with a root mean square (RMS) scatter comparable to the signal amplitude. Even with the input sine wave to guide you, it is impossible to detect the signal by eye. The lower panel shows a periodogram of the data and one can clearly see a strong peak at the frequency of

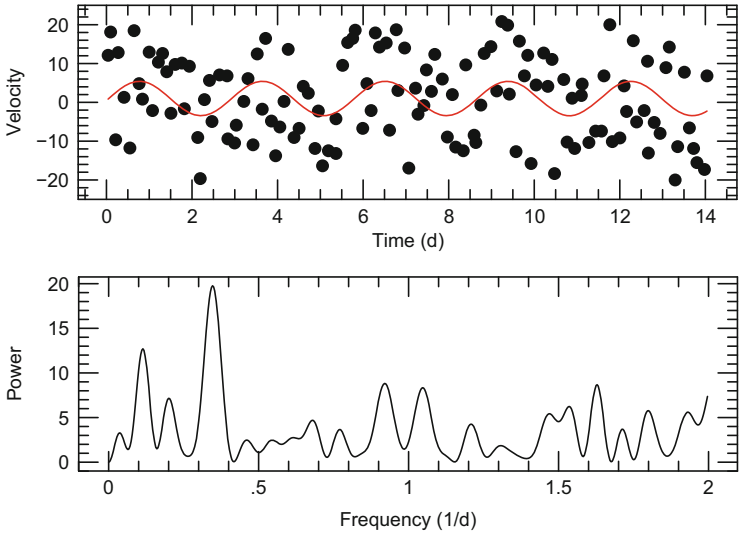

Fig. 1.23 (*Top*) Time series (*dots*) of a sampled sine function (*curve*) with a period of 2.85 days whose amplitude is the same level as the noise. The sine variations are not visible in the time series. (*Bottom*) The DFT power spectrum of the above time series. A dominant peak at the correct frequency is clearly seen

the input sine wave (0.35 day^{-1}). We need special tools to extract signals from noisy data.

Finding periodic[2] signals in time series data is a problem found in many aspects of science and engineering, not just exoplanet research. The most used tools are largely based on the discrete Fourier transform (DFT). If you have a time series of measurements $x(t_n)$ where t_n is the time of your nth measurements, the DFT is

$$\mathrm{DFT}_X(\omega) = \sum_{i=1}^{N} X(t_j) e^{-i\omega t_j} \tag{1.12}$$

where $e^{i\omega t}$ is the complex trigonometric function $\cos(\omega t) + i\sin(\omega t)$, N is the number of data points sampled at times t_j, and ω the frequency.[3]

The foundation for the DFT is the fact that sines and cosines are orthogonal functions that form a basis set. This means that any function can be represented as a linear combination of sines or cosines.

The power is defined by

$$P_X(\omega) = \frac{1}{N}|\mathrm{FT}_x(\omega)|^2 = \frac{1}{N}\left[\left(\sum_{j=1}^{N} X_j \cos\omega t_j\right)^2 + \left(\sum_{j=1}^{N} X_j \sin\omega t_j\right)^2\right] \tag{1.13}$$

and this is often called the classic periodogram.

The DFT of a real time series can have a real and imaginary part, but in astronomy we are interested in real quantities. The DFT is often represented as a power spectrum $P(\omega)$ where P is the complex conjugate—a real value—or sometimes as the amplitude spectrum $A(\omega)$ where $A = \sqrt{P_X}$.

The utility of the DFT is that if a periodic signal is in your data it appears (nearly) as a ∂-function in Fourier space at the frequency and with the amplitude of your sine wave. In the presence of noise the periodic signal can be more readily seen in the frequency domain (Fig. 1.23).

Ideally, one would like to have data that are taken in equally spaced time intervals, but this is rarely possible with astronomical observations. There are DFT algorithms available for unequally spaced data.

A useful tool for DFT analyses is the program *Period04* (Lenz and Breger 2005). *Period04* enables you to perform a DFT on unequally spaced time series and plot the amplitude spectrum. Peaks in the amplitude spectrum can then be selected and

[2]It has become a common practice in RV exoplanet discoveries to plot period along the abscissa. I prefer to use frequency as this does not distort the periodogram. I will frequently interchange the use of period with frequency in the discussion. You can easily go from one to the other using the frequency-to-period converter on your hand calculator, i.e. the "$1/x$" key.

[3]Frequency is often measured as angular frequency which is related to the period by $\omega = 2\pi/P$. Throughout this paper when I refer to a frequency it is merely the inverse of the period, or day^{-1}.

a sine function fit to the data made. Later we will employ *Period04* for some time series analysis.

Many planet hunters use an alternative form of the DFT, namely the Lomb–Scargle periodogram (Lomb 1976; Scargle 1982). This is defined by the more complicated expression:

$$P_X(\omega) = \frac{1}{2}\left\{\frac{\left[\sum_j X_j \cos\omega(t_j-\tau)\right]^2}{\sum_j \cos^2\omega(t_j-\tau)} + \frac{\left[\sum_j X_j \sin\omega(t_j-\tau)\right]^2}{\sum_j \sin^2\omega(t_j-\tau)}\right\} \tag{1.14}$$

where τ is defined by

$$\tan(2\omega\tau) = \sum_j \sin(2\omega t_j) / \sum_j \cos(2\omega t_j)$$

The periodogram defined in this way has useful statistical properties that enable one to determine the statistical significance of a periodic signal in the data. One of the main problems of time series analysis is finding a periodic signal that is real and not due to noise. LSP gives you an estimate of the significance of such a signal.

Note that the Lomb–Scargle periodogram [Eq. (1.14)] does not take into account weights on the data values and assumes that the time series has zero average. The generalized Lomb–Scargle periodogram includes an offset to the data as well as weights (Zechmeister and Kürster 2009).

The DFT and the LSP are intimately related. In fact, Scargle (1982) showed that the LSP is the equivalent of sine fitting of data, essentially a DFT. It is worth mentioning, however, the differences between the DFT and the LSP. In the DFT the power at a frequency is just the amplitude squared of the periodic signal. The amplitude (or power) of the signal can be read directly from the DFT amplitude (power) spectrum. On the other hand the power in the LSP is related in a non-linear way to the statistical significance of the signal.

Figure 1.24 highlights both the similarities and differences between the two. It shows the DFT (right) and LSP (left) of an input sampled sine wave with no noise. The two look nearly identical except in the power. The DFT has a power of 100, the square of the input amplitude of 10. The LS has a large power that tells you nothing about the amplitude of the periodic signal. However, its large value of $\approx$2500 immediately tells you that the signal is significant and that it is virtually impossible that it is due to noise. For noise-free data there are no real advantages for using the LSP over the DFT. The LSP shows its utility when the noise becomes comparable to the signal amplitude.

As the statistical significance increases, the power in the LSP increases. The right panel of Fig. 1.25 shows the LSP of noisy data consisting of a sine function whose amplitude is comparable to the measurement error. After using 50 measurements points for calculating the LSP the peak is now reasonably significant. After using

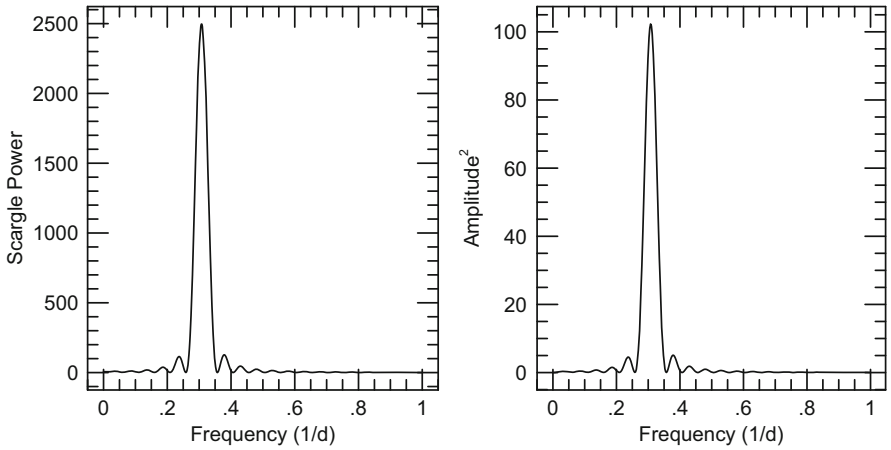

Fig. 1.24 (*Left*) The Lomb–Scargle periodogram of a time series with a periodic signal. (*Right*) Power spectrum from the discrete Fourier transform of the same time series. Note that both periodograms look identical, but have different power levels. The LSP power is related to the statistical significance while the DFT power is the amplitude^2 of the signal

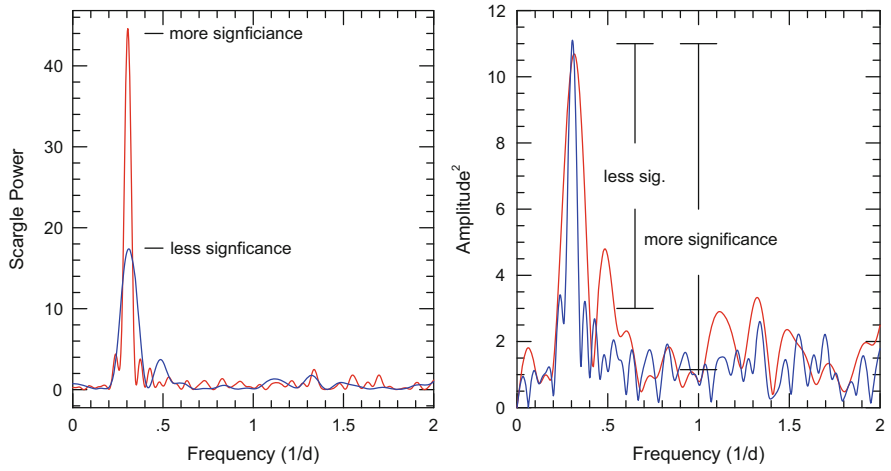

Fig. 1.25 (*Left*) Lomb–Scargle periodograms of periodic time series with noise. The *blue curve* (smaller amplitude) is for a short time length. The *red curve* (higher amplitude) is for a longer time length of the data. In this case the signal becomes more significant thus the LS power increases. (*Right*) The same but for the power spectrum from the DFT. Since the power is related to the amplitude of the signal it does not increase with more data. However, the noise floor decreases. In this case the significance of the signal is measured by how high above the surrounding noise level the peak is

100 measurements the detected signal becomes more significant and the LSP power greatly increases. The significance of the signal is measured by the power level.

The left panel of Fig. 1.25 shows the DFT of the same time series, again for 50 measurements (red line) and 100 measurements (blue line). In this case the power of the DFT remains more or less constant in spite of more measurements and thus higher statistical significance. There are slight differences in the amplitude, but this is because as you get more measurements the amplitude is better defined. The power remains constant in the DFT because it is related to the amplitude of the signal and that is not changing. Note, however, that the level of the noise drops—the power level of the surrounding noise peaks drops with respect to the signal peak. In the case of the DFT, the significance of a signal is inferred by how high a peak is above the surrounding noise floor.

1.4.1 Assessing the Statistical Significance

If one has noisy data that is poorly sampled, it is often easy to find a periodic signal that fits the data. Figure 1.26 shows 10 simulated RV measurements. A periodogram found a signal at $P = 0.59$ day. The fit to the data looks quite good. The only problem is that the data were generated using pure random noise—there is no periodic signal in the data. If you have noisy data of limited time span and sparse sampling, you almost *always* will be able to fit a sine curve to your data.

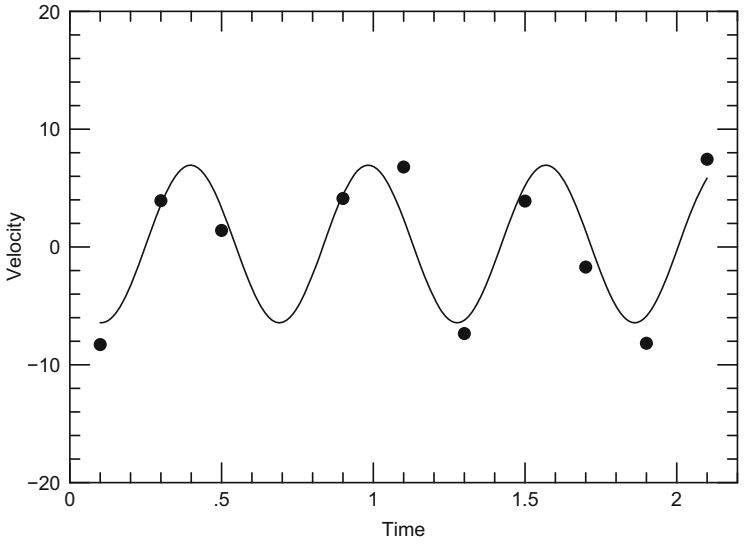

Fig. 1.26 A sine fit (*curve*) to ten simulated RV measurements consisting only of random noise

1.4.1.1 Lomb–Scargle Periodogram

In assessing the statistical significance of a peak in the Lomb–Scargle periodogram there are two cases to consider. The first is when you want to know if noise can produce a peak with power higher than the one you found in your data, the so-called false alarm probability (FAP) over a wide frequency range. The second case is if there is a known periodic signal in your data. In this case you want to ask what the FAP is *exactly* at that frequency.

If you want to estimate the FAP over a frequency range ν_1 to ν_2, then for a peak with a certain power the FAP is (Scargle 1982)

$$\mathrm{FAP} = 1 - (1 - e^{-z})^N \tag{1.15}$$

where z is the power of the peak in the LSP and N is the number of independent frequencies. It is often difficult to estimate the number of independent frequencies, but as a rough approximation one can just take the number of data measurements. Horne and Baliunas (1986) gave an empirical expression relating the number of independent frequencies to the number of measurements. For large N the above expression reduces to $\mathrm{FAP} \approx \mathrm{Ne}^{-z}$.

If there is a known period (frequency) in your data, say ν_0, then you must ask what is the probability that noise will produce a peak higher than what is observed *exactly* at ν_0. This is certainly the case for a transiting planet where you know the period of the orbiting body and you are interested only in the power at that orbital frequency. In this case $N = 1$ as there is only one independent frequency. The above expression reduces to

$$\mathrm{FAP} = e^{-z} \tag{1.16}$$

For example, if you have 30 RV measurements of a star, then you need a power of $z \approx 8$ for an unknown signal to have a FAP of 0.01. If there is a known frequency in your data, then you need only $z \approx 4.6$ at that frequency. As a personal rule of thumb for unknown signals: $z = 6$–$10 \rightarrow$ most likely noise, $z = 8$–$12 \rightarrow$ most likely noise, but might be a signal, $z = 12$–$15 \rightarrow$ interesting, get more data, $z = 15$–$20 \rightarrow$ most likely a signal, but nature still might fool you; $z > 20 \rightarrow$ definitely a signal, publish your results.

The FAP estimated using the above expression should just be taken as a rough estimate. A more accurate value from the FAP comes from a bootstrap analysis using two approaches. The first method is to create random noise with a standard deviation having the same value as the RMS scatter in your data. Calculate the LSP and find the highest peak in the periodogram. Do this a large number (10,000–100,000) of times for different random numbers. The fraction of random data sets having LSP higher than that in your data is the FAP.

This of course assumes that your noise is Gaussian and that you have a good handle on your errors. What if the noise is non-Gaussian, or you are not sure of your true errors? The most common form of the bootstrap is to take the actual data

and randomly shuffle the data keeping the time values fixed (Murdoch et al. 1993). Calculate the LSP, find the highest peak, then re-shuffle the data. The fraction of the shuffled data periodograms having power larger than the original data gives you the FAP. This method more or less preserves the statistical characteristics of the noise in your data. Of course if you still have a periodic signal in your data, then this will create a larger RMS scatter than what you would expect due to measurement uncertainties. The bootstrap will produce a higher FAP in this case and would thus be a conservative estimate of the FAP.

1.4.1.2 Fourier Amplitude Spectrum

It is possible to get an estimate of the FAP from the Fourier amplitude spectrum. Through Monte Carlo simulations Kuschnig et al. (1997) established that if a peak in the amplitude spectrum has a height approximately 3.6 times higher than the mean peaks of the noise that surrounds it, then this corresponds to a FAP ≈ 0.01. Figure 1.27 shows the FAP as a function of the peak height above the noise. This was generated using a fit to Fig. 4 in Kuschnig et al. (1997). Again, this should only be taken as a crude estimate. Refined values can be obtained using the bootstrap method.

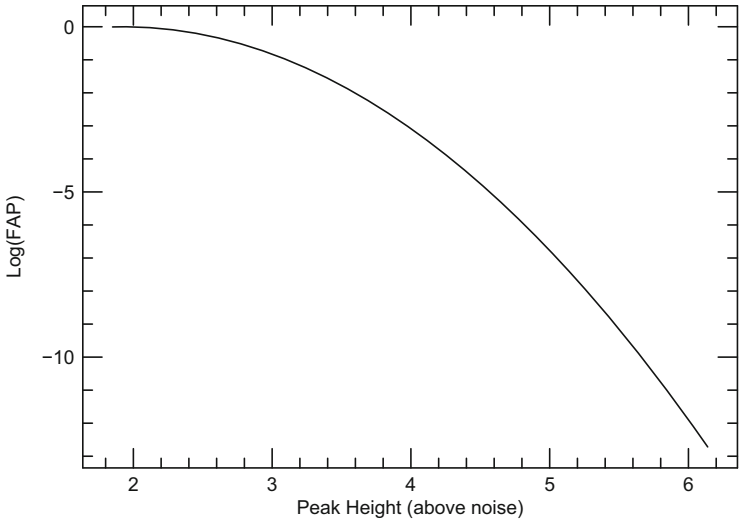

Fig. 1.27 The FAP as a function of the peak height above the noise for a peak in the DFT amplitude spectrum

1.4.2 Aliases and Spectral Windows

When analyzing a time series for periodic signals one would prefer to have equal time sampling and a very long uninterrupted time string (as well as no noise!). The real world is never ideal. One can rarely observe in equally spaced time intervals and the time series will always be interrupted. In the case of astronomical observations these interruptions are due to sunrise, bad weather, or unsympathetic time allocation committees who would not grant you enough telescope time!

In understanding how the sampling window affects the structure seen in DFTs and LSPs there are two useful rules to keep in mind:

Rule 1: A function that is narrow in the time domain will be broad in the Fourier domain. It is well known that the Fourier transform of a Gaussian is another Gaussian. However, if the Gaussian is broad in the time domain, it will be narrow in the Fourier domain. Likewise, optical interferometry produces a Fourier transform of your intensity pattern. The interference pattern of a 1-D slit is the sinc function. The narrower the slit, the broader the sinc function in the Fourier domain.

Rule 2: The convolution of two functions in the time domain translates into a multiplication of the Fourier transforms of the individual functions in the Fourier domain.

The convolution of two functions $f(t)$ and $g(t)$ is defined as

$$f * g = \int_{-\infty}^{+\infty} f(\tau)g(t-\tau)d\tau \tag{1.17}$$

Note that the convolution is intimately related to the cross-correlation function [Eq. (1.8)]

In the Fourier domain the convolution of these two functions is just the product of their respective Fourier transforms, $G(\omega)$ and $F(\omega)$:

$$f(t) * g(t) \equiv F(\omega) \cdot G(\omega) \tag{1.18}$$

This mathematical procedure of convolution involves reversing a function (g) and sliding it with respect to f, all the while integrating under both functions. It is often called a smoothing function since if you convolve a time series, or spectral data, by say a box of width N, you are producing a running average of points within that box. This complicated mathematical procedure is reduced to a simple product, $F \cdot G$, in Fourier space.

Going back to our smoothing example. Suppose you had a noisy spectrum and you wished to smooth it by four pixels. The "brute force" way is to convolve your spectrum with a box of width four pixels, i.e. you slide the box along your spectrum taking the running average at each point. The "elegant" way would be to perform a Fourier transform of your spectrum, a Fourier transform of your box smoothing

function (a sinc!), multiply the two, and compute the inverse Fourier transform. You will arrive at the same smoothed spectrum.

The convolution theorem works both ways, that is to say if you multiply two functions in the time domain, that is the same as convolving the individual Fourier transforms in the Fourier domain. As we will see, this fact can produce some rather complicated periodograms when looking for true periodic signals in RV data.

For a pure sine wave an LSP or DFT should produce a ∂-function, but that is only the case for an infinite time series. Real observations are always over a finite time interval and with gaps. In the time domain this is a sine wave that is multiplied by a box function whose width is the time span of your observations. In the Fourier domain (DFT and LSP) this multiplication turns into a convolution of the Fourier transforms of the individual functions. In Fourier space a sine wave is a delta function, and a box is a sinc function. The convolution causes the ∂-function to appear as a sinc function centered on the frequency of your signal and with the same amplitude.

This process is shown in Fig. 1.28 where we have used the more realistic case of several box functions simulating observations over several nights. Again, the window function, this time a bit more complicated, appears at the frequency of the input sine wave.

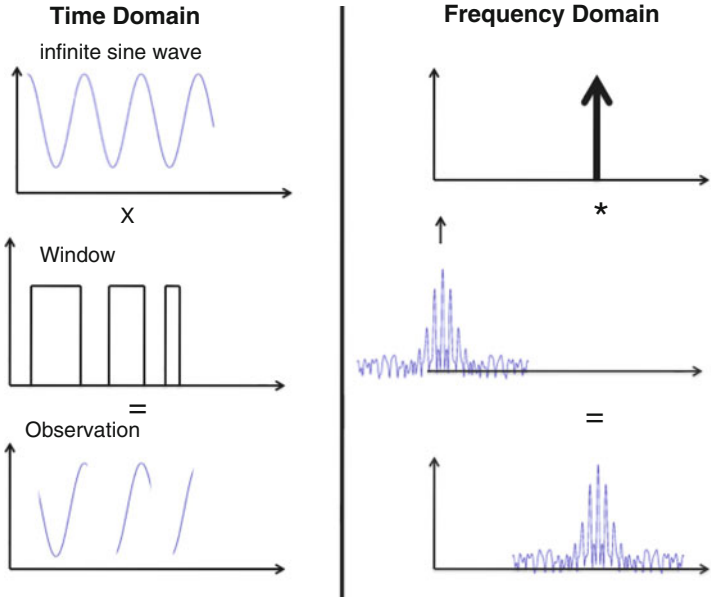

Fig. 1.28 (*Left panels*) In the time domain the observations consist of an infinite sine wave multiplied by the sampling window that produces a sine with gaps. (*Right*) In the Fourier domain the sine function appears as a ∂-function that is convolved (symbolized by the *star*) with the Fourier transform of the window function

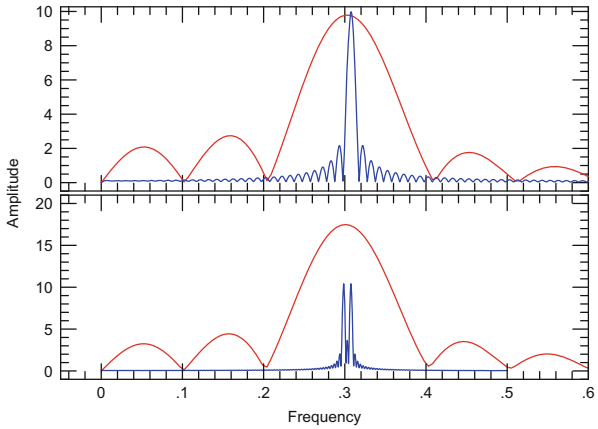

Fig. 1.29 (*Top*) DFT amplitude spectrum of a periodic signal at P = 3.25 days for a 10-day time window (*broad red curve*) and a 100-day time window (*narrow blue curve*). (*Bottom*) DFT amplitude spectrum from a series consisting of two closely spaced sine functions with P = 3.24 days and P = 3.35 days and the same amplitude. The *red* (*broad*) *curve* is for a 10-day time series and the *blue* (*narrow*) *curve* is for a 300-day time series

The longer the observing window, i.e. longer box, the narrower the sinc function. Recall that the narrower the function in the time domain, the wider it is in the Fourier domain. The top panel of Fig. 1.29 shows the DFT amplitude spectrum of an input sine wave with a period of 3.25 days and with different lengths of the observing window. The red (broad) curve is for a 10-days observing time span (uninterrupted sampling). The blue (narrow) curve is for a 100 days observing window.

When trying to resolve two very closely spaced frequencies it is important to have a wide observing window. Such a case may arise if, say, you are trying to extract a planet signal whose period is only slightly different from the rotational period of the star. The length of your observing window, ∂T, corresponds to a frequency of $1/\partial T$ and this must be much less than the frequency separation of the two signals you are trying to detect.

The lower panel of Fig. 1.29 shows the DFT of two input sine functions of equal amplitude with periods of 3.25 and 3.35 days, or a frequency separation of 0.009 day^{-1}. The DFT in red is for a time string of only 10 days. The observing window thus has a frequency width of 0.1 day^{-1}, insufficient to resolve the two signals. Instead the DFT produces a single peak at $\approx$0.3 day^{-1} but with twice the amplitude. Over this short time the time series looks like a single sine curve. The DFT in blue is for a time string of 400 days that has a window frequency width of 0.0033 day^{-1}. In this case both signals are easily resolved.

Multi-periodic signals result in multiple ∂-functions in the Fourier domain. This means that the window function will be superimposed at the frequency location of every signal peak (Fig. 1.30). This can result in a complicated DFT or LSP. The DFT in Fig. 1.30 looks as if it has a myriad of signals, but it in fact only contains

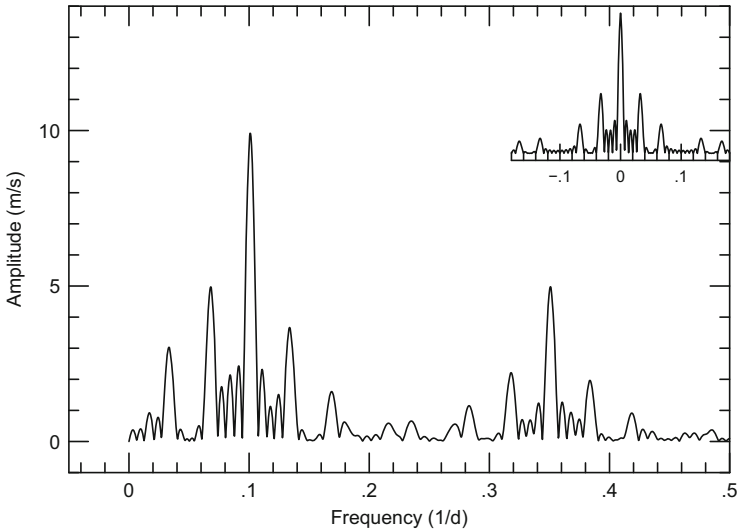

Fig. 1.30 The sampled time series consisting of two sine functions with periods of 10 days (ν = 0.1 day^{-1}) and 2.85 days (ν = 0.35 day^{-1}) . The DFT of the window function is shown in the *upper panel*. Because the observed DFT is a convolution of the Fourier transform of the window function (spectral window) with the data transform (two ∂-functions) the spectral window appears at each signal frequency

two periodic signals. All the others are due to the window function (panel in figure). This window function is superimposed at every real frequency. There are only two signals in this case, but the DFT can get quite messy if one has many periodic signals in the data (e.g., multi-planet system). One can easily check whether a peak is due to a real signal or the window function by fitting a sine wave to the data with the appropriate period and subtracting this from the data. This will remove not only the signal, but all of its peaks due to the window function will also disappear.

Aliasing is also a problem that may cause one to find a wrong period in time series data. An alias period is due to under-sampling and will cause an under-sampled short period signal to appear as a much longer period. This is demonstrated in Fig. 1.31 which shows two sine functions of different periods. If the short period sine wave was actually in the data and you sampled it at the rate shown by the dots, you would not be able to distinguish which signal was in your data. Both would fit your measurements.

To avoid aliases one must satisfy the Nyquist sampling criterion. If your sampling rate is ∂T, this corresponds to a sampling frequency of $f_s = 1/\partial T$. If you want to detect higher frequency signals than f_c, then the Nyquist criterion states

$$f_s \geq 2f_c \tag{1.19}$$

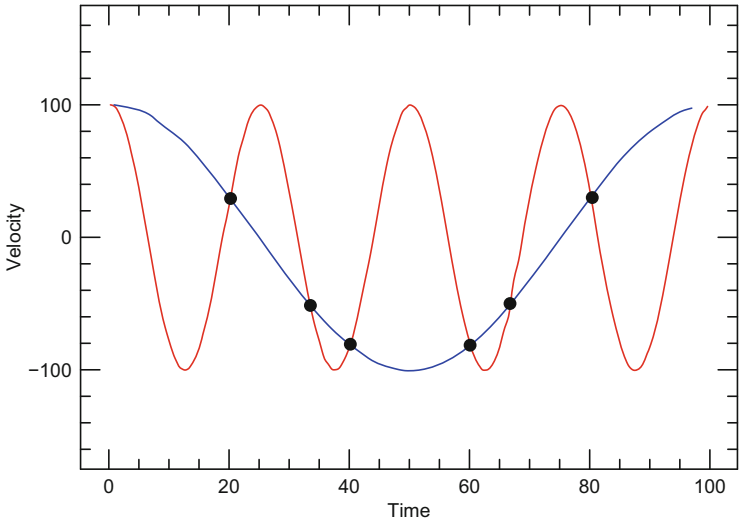

Fig. 1.31 Six RV measurements (*dots*) that can be fit by a short or a long period sine function. Because of the poor sampling the long period sine wave appears as an alias of the shorter period sine

For example, if you observe a star once a night, your sampling frequency is 1 day^{-1}. The Nyquist frequency (f_c) in this case is 0.5 day^{-1}. This means that you will not be able to detect a periodic signal with a frequency longer than $2f_s$, or equivalently a period shorter than 2 days. If shorter periods were in the data, these will appear at a longer period in your Fourier spectrum. So, if you have a planet with an orbital period of 0.8 day, the wrong strategy would be to take only one observation per night. To detect such a period requires several observations per night, scattered throughout the night, and over several consecutive days.

Ground-based observations are always interrupted by the diurnal cycle of the sun caused by the Earth's rotation. This means that the 1-day aliases are ubiquitous in ground-based measurements. For ground-based observations if you have a frequency f_d in your data, then in the periodogram alias frequencies will appear at

$$f_{\rm alias} = f_d \pm 1 \tag{1.20}$$

There is a well-known example where the alias problem resulted in astronomers reaching a wrong conclusion and with important consequences. The planet 55 Cnc-e was one of the first hot Neptunes discovered by the RV method. The planet had a mass of 17.7 $M_\odot$ and an orbital period reported to be 2.85 days (McArthur et al. 2004). This period, however, was the alias of the true period at 0.74 day.

Figure 1.32 shows the DFT power spectrum of the RV measurements for 55 Cnc plotted beyond the Nyquist frequency of $\approx$0.5 day^{-1}. One sees a peak at the "discovery" orbital frequency at 0.35 day^{-1} (P = 2.85 days). However, there is a

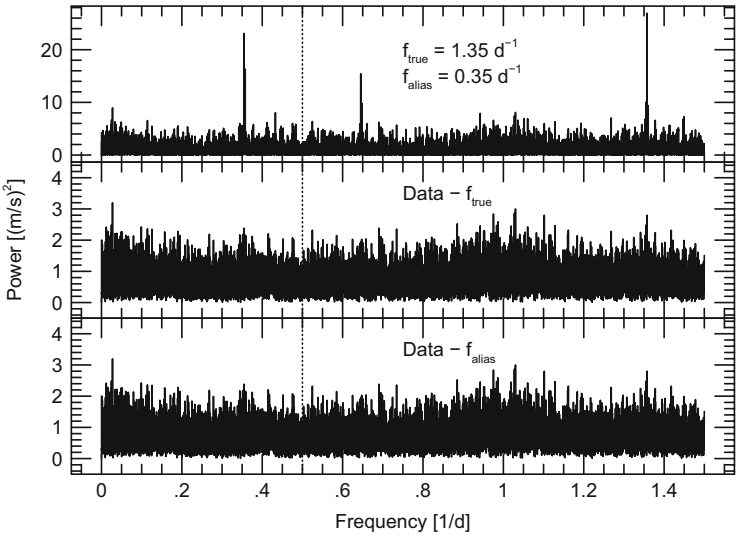

Fig. 1.32 (*Top*) DFT power spectrum of the RV measurements for 55 Cnc. The *vertical dashed line* represents the Nyquist frequency. (*Center*) DFT power spectrum of the RVs after removing the true orbital frequency of the planet at 1.35 days^{-1}. (*Bottom*) DFT power spectrum of the RVs after removing the signal of the alias frequency at 0.35 day^{-1}

slightly higher peak at the alias frequency of 1.35 days^{-1}. There is also another, most likely alias at 0.65 day^{-1}, or about one-half the higher frequency. One can check that these are aliases by fitting sine waves to the data using the 1.35 days^{-1} frequency, and then the other. Note that in both cases all peaks are suppressed, indicating that only one frequency is in the data.

So, how does one determine the true frequency? One can get a hint via simulations. Take a sine wave with one of the periods, sample it like the real data, and add the appropriate level of noise. Compute the periodogram. Then create another data set with the other period, sampled like the data and with noise. The periodogram that best matches the real periodogram gives a hint as to which period is most likely in your data.

Of course the best way to exclude alias periods is by adhering to the Nyquist criterion. In the case of 55 Cnc this requires a sampling rate with a frequency $f_s \geq 2.7$ days^{-1}, or a time interval shorter than 0.37 day. At best, observe the star throughout the night over several nights.

In the case of 55 Cnc, the discoverers missed out on a golden opportunity. Because the true period was so short, that meant that the transit probability was higher. Indeed, space based photometric measurements were able to detect a transit at the shorter period (Winn et al. 2011; Demory et al. 2011). 55 Cnc would have not only been the first hot Neptune, but the first transiting hot Neptune! Furthermore, because the true orbital period was shorter the mass of the planet was reduced to 11.3 $M_{\odot}$ (true mass!). The discoverers probably did not bother to look beyond the

Nyquist frequency because at the time having a planet with an orbital period less than 1 day would have been unprecedented.

So the lesson learned: look beyond the Nyquist frequency when doing time series analysis, and do not bias yourself as to what periods to be looking for.

1.4.3 Finding Planetary Systems with Pre-whitening

Exoplanets most likely are in multiple systems. It is safe to say, that where there is one planet there are surely others. Indeed, the Kepler space mission has shown that multiple planet systems are common (see Latham et al. 2011).

A fast and efficient means of finding multiple planets in your RV data is through the process of pre-whitening. This is a technique commonly used in the study of stellar oscillations where one often has to extract multi-periodic signals from your time series. It is easier to find additional signals in the periodogram after removing dominant signals.

In pre-whitening, one performs a DFT on the time series in order to find the dominant peak in the data. A sine fit is made to the data using that frequency and this is subtracted from the data. Note that this procedure also removes all the aliases due to this signal. One then performs a DFT on the residual data to find the next dominant peak which you then fit and subtract. The process stops when the final residual peak in the DFT is at the level of the noise. A good level to stop is when the final peak is less than four times the surrounding noise level.

The procedure is called pre-whitening because a "white" Fourier spectrum has more or less equal power at all frequencies. By removing large power peaks one is making the spectrum "whiter." Alternatively, you can view this as Fourier component analysis, i.e. you are sequentially finding the dominant Fourier components in your data.

The program *Period04* provides a nice environment along with a graphical interface for pre-whitening data. In fact, the program was specifically developed for finding multi-periodic signals in time series data of oscillating stars. With literally a few clicks of a mouse button one can produce a DFT, fit a sine function to a selected peak, and then view the residuals which can be searched for additional periods.

I will demonstrate the pre-whitening procedure on the multi-planet system GJ 876. Jenkins et al. (2014) published orbital solutions for six planets. Figure 1.33 shows the DFT amplitude spectrum for the RV data used by Jenkins et al. (2014). It shows a strong peak at $K = 200\,\mathrm{m\,s^{-1}}$ at $f_b = 0.0163\ \mathrm{day^{-1}}$ ($P = 60$ days) the orbital frequency of GJ 876b (Marcy et al. 1998). The panel below shows the DFT after removal of f_b. The residual peak corresponds to the orbital frequency of the second planet, $f_c = 0.033\ \mathrm{day^{-1}}$ ($P = 30.28$ days). The pre-whitening procedure then continues top, bottom, and again on the right column panels.

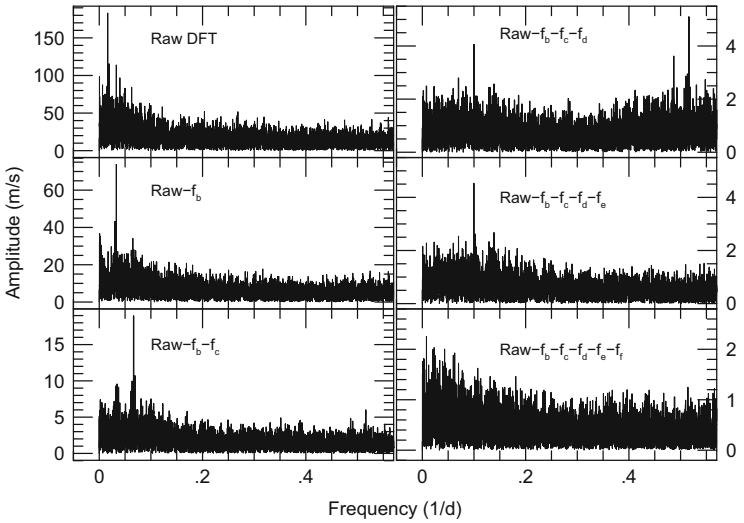

Fig. 1.33 Pre-whitening procedure applied to the RV data of the six planets system GJ 876

Table 1.2 Planets found in the GL 876 system by the pre-whitening process compared to published values (Jenkins et al. 2014)

	ν (day^{-1})	P (days)	K (m s^{-1})	$P_{\text{published}}$ (days)	$K_{\text{published}}$ (m s^{-1})
f_b	0.0163	61.03 ± 0.001	211.7 ± 0.4	61.03 ± 0.03	211.6 ± 32.9
f_c	0.0330	30.28 ± 0.004	89.0 ± 0.3	30.23 ± 0.03	88.7 ± 13.2
f_d	0.0664	15.04 ± 0.004	20.76 ± 0.28	15.04 ± 0.004	20.7 ± 3.2
f_f	0.0998	10.01 ± 0.03	5.70 ± 0.27	10.01 ± 0.02	5.00 ± 0.80
f_e	0.5160	1.94 ± 0.001	6.21 ± 0.23	1.94 ± 0.001	5.91 ± 0.98
f_g	0.0080	124.88 ± 0.02	3.19 ± 0.35	124.88 ± 90	3.37 ± 0.53

Table 1.2 shows the resulting signals found in the RV data for GJ 876 using pre-whitening compared with the published values from Jenkins et al. (2014). All six planets are recovered at the correct periods (frequencies) and amplitudes. I should note that the entire process of finding these planets with *Period04* took about 1 min.

A DFT analysis uses sine functions which correspond to circular orbits. Planets, however, can have eccentric orbits. The pre-whitening procedure can be refined to account for eccentric orbits. One first finds all the "planets" in the RV data via pre-whitening. You then isolate the RV variations of one planet by removing the contribution of all the others. These data can then be fit with a Keplerian orbit that includes a non-zero eccentricity. This orbital solution is then removed from the data and you proceed to the next planet.

1.4.4 Determining the Nature of the Periodic Signal

Figure 1.34 shows the LSP of the RV data shown in the left panel of Fig. 1.22. It shows a strong peak at $\nu = 0.000497\ \mathrm{day}^{-1}$ corresponding to a period of 2012 days and a K-amplitude of $37.9\ \mathrm{m\,s}^{-1}$. This is the signal expected for a reflex motion of a star due to a 2.4 M_{Jupiter} companion. A bootstrap analysis reveals that the FAP $< 5 \times 10^{-6}$. This is a real signal in the data and not due to noise. However, as we have seen these variations are instrumental shifts caused by variations in the IP of the spectrograph.

You have found a periodic signal in your RV data and have determined that it is real and not due to noise. Before you can fit an orbital solution and announce to the world your exoplanet discovery, you must first discern the nature of the RV variations. The problem is that a wide range of phenomena can mimic a planet signal and often discerning the true nature of a signal is the most difficult part in finding exoplanets with RV measurements.

Besides planets, periodic signals can arise from three broad areas:

1. Systematic measurement errors often caused by instability of the instrument that is periodic. Figure 1.22 is an example of this.
2. Improper data analysis, most notably in the barycentric correction of the Earth's motion.
3. Intrinsic variability of the star.

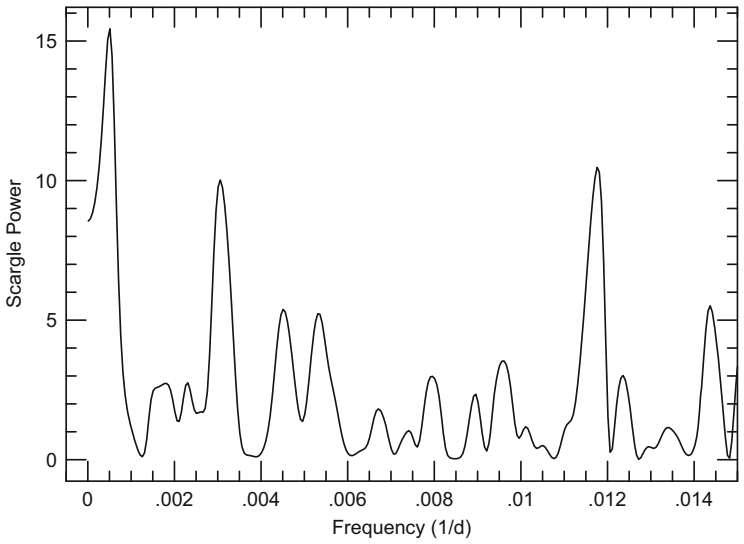

Fig. 1.34 LSP of the RV data shown in the *left panel* of Fig. 1.22. The statistical significance of the signal is FAP $< 5 \times 10^{-6}$

In Sect. 1.6 I will cover in more detail sources of errors, particularly from stellar variability. Here I briefly summarize a few checks you can make.

The first hint of the nature of the RV variations comes from the period. If you have a 1-year period in your data, look hard into the barycentric correction. Exoplanets certainly can have a period near 365 days, but more than likely it is due to an improper barycentric correction. A 1-year period should always make you uncomfortable. In my work I had two cases where I found a planet signal in the data near 365 days. One turned out to be a real planet, the other due to a subtle bug in the RV program. Bottom line, one should always be cautious when one finds periods in your RV data that coincide with well-known astronomical periods, i.e. 1 day, 1 month, or integral fractions thereof.

Seasonal temperature variations in the spectrograph can also cause a 1 year, $\approx$6 months, or $\approx$3 months period to appear, especially if your spectrograph is not temperature stabilized. A 1-month period may point to contamination in your spectra by moonlight.

Intrinsic stellar variability is one of the more difficult problems in discerning the nature of RV variations and one that will be dealt with in more detail in Sects. 1.6 and 1.7. Is the RV period what you expect from stellar rotation or stellar oscillations? One of the strongest arguments that the RV signal of 51 Peg b was actually due to a planet was that the 4-day orbital period was much shorter than the rotation period of the star, yet much longer than expected oscillations in a sun-like star (Mayor and Queloz 1995).

Ancillary measurements are also important to establish that the RV period is not the rotation period of the star. The stellar rotation can be determined through a variety of measurements: photometric, Ca II H & K lines, or spectral line shapes. This will be covered in more detail in Sect. 1.6. Bottom line, a planet signal should only be present in the RV data, if you see the same period in other quantities, then you probably do not have a planet.

There is one simple test one can use to exclude other phenomena as a cause of your RV variations. A planetary companion will cause a periodic signal in your RV data that is always present. After all, it is unlikely that your planet disappears! This means that as you acquire more data the statistical significance of your detection should increase. However, signals from other phenomena such as stellar activity should not have such coherent and long-lived signals. Spot features come and go, they migrate, and there might be times when they simply disappear. This causes the strength of a signal due to, say, stellar activity, or even pulsations, to change. This means that as you collect more data the statistical significance of the signal will decrease. However, for signals that are long-lived and coherent, as you acquire more data the power in the Lomb–Scargle periodogram should always increase. Figure 1.35 shows the Scargle power of the RV signal that was attributed to the planet GJ 581g (Vogt et al. 2010, 2012) as a function of the number of data points used for the periodogram. It shows an approximately linear rise reaching a power consistent with a FAP $\approx 10^{-6}$—this is a highly significant signal. However, with more data ($N > 220$) the power dramatically drops. The solid squares shows the expected increase in power for a simulated planet with the same orbital period and

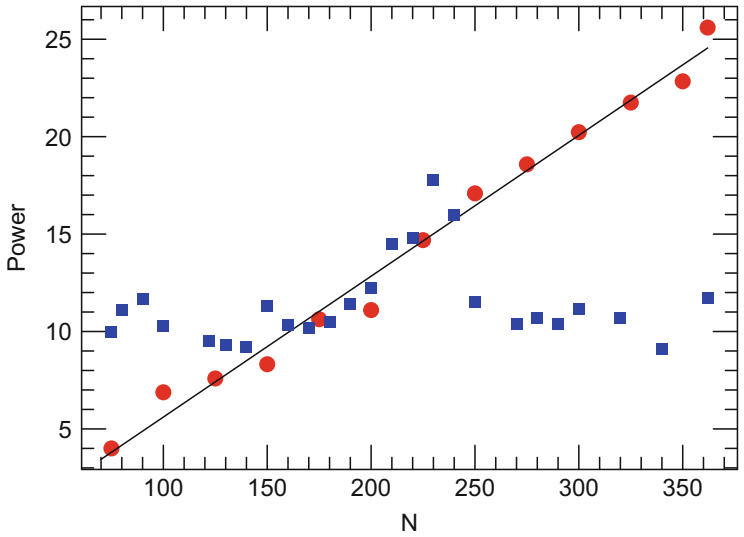

Fig. 1.35 (*Blue squares*) The LS power of the RV signal due to GJ 581g as a function of the number of data points (*N*) used to compute the LSP. (*Red dots*) The predicted increase in LS power as a function of *N* for a simulated signal from GJ 581g and the appropriate noise level. Note that these follow a linear trend (*line*)

K-amplitude as GJ 581g that was sampled the same way as the data. As expected the power, and thus significance, increases. The fact that the power of the signal does not increase with more data is an indication that it may be an artifact of activity (Hatzes 2013b).

1.5 Keplerian Orbits

In this lecture I will briefly cover Keplerian orbits. Michael Perryman's *The Exoplanet Handbook* (Perryman 2014) has an excellent description of orbits that I largely will follow here.

1.5.1 Specifying the Orbit

Figure 1.36 shows the basic elements of a Keplerian orbit. These elements include:

Reference plane: the plane tangent to the celestial sphere.
Line of nodes: the line segment defined by the intersection of the orbital plane with the reference plane.

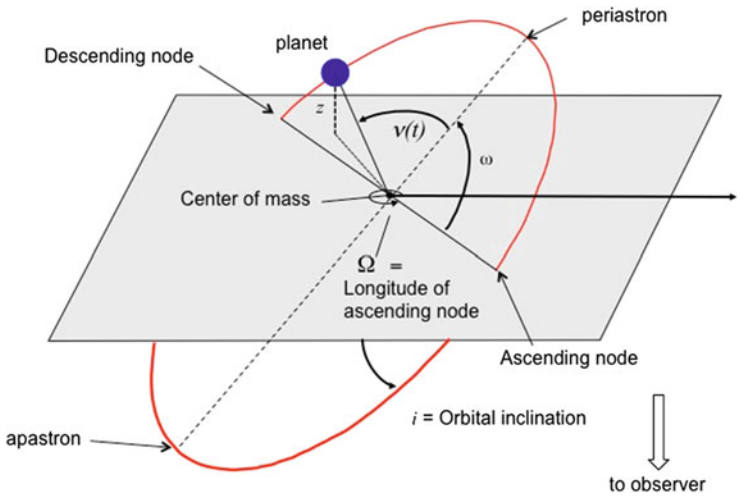

Fig. 1.36 The elements of a Keplerian orbit

Ascending node: point where the planet crosses the reference plane and moving away from the observer.
Descending node: point where the planet crosses the reference plane and moving towards the observer.
Ω = longitude of ascending node. The angle between the vernal equinox and the ascending node. It is the orientation of the orbit in the sky.

Fully parameterizing a Keplerian orbit requires seven parameters: a, e, P, t_p, i, Ω, *and* ω:

a : semi-major axis that defines the long axis of the elliptical orbit.
e : eccentricity describes the amount of ellipticity in the orbit.
P : orbital period.
T_0 : the time of periastron passage.
Ω : longitude of ascending node.
ω (argument of periastron): angle of the periastron measured from the line of nodes.
i (orbital inclination): angle of the orbital plane with respect to the line of sight.

RV measurements can determine all of these parameters except for two, Ω and the orbital inclination. The angle Ω is irrelevant for the mass determination. The orbital inclination, however, is important. Because we measure only one component of the star's motion we can only get a lower limit to the planet mass, i.e. the mass times the sine of the orbital inclination.

For RV orbits there is one more important parameter that one derives and that is of course the velocity K-amplitude.[4] This is the component along the line of sight of the star's orbital velocity amplitude. We do not measure the mass of the planet directly, but through the amplitude of the reflex motion of the host star.

1.5.2 Describing the Orbital Motion

For Keplerian orbits the star and planet move about the center of mass, or barycenter, in elliptical orbits with the center of mass at one focus of the ellipse. The ellipse is described in polar coordinates by (see Fig. 1.37):

$$r = \frac{a(1-e^2)}{1+e\cos\nu} \tag{1.21}$$

The eccentricity is related the semi-major (a) and semi-minor (b) axes by

$$b^2 = a^2(1-e^2)$$

The periastron distance, q, and apastron distance, Q, are given by

$$q = a(1-e)$$
$$Q = a(1+e)$$

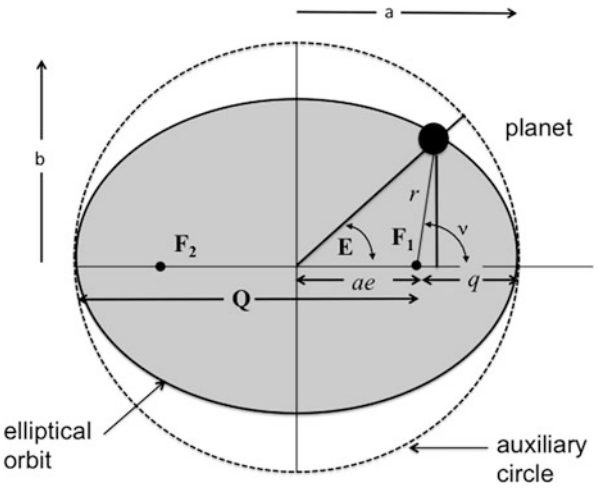

Fig. 1.37 Elements of an elliptical orbit. The auxiliary circle has a radius equal to that of the semi-major axis, a. The semi-minor axis is b. The true anomaly, ν, describes points on the orbit. Alternatively, one can use the eccentric anomaly, E. The focus, F_1, is the system barycenter

[4]Historically the radial velocity amplitude is denoted by the variable K and is often referred to as the K-amplitude.

In calculating orbital motion there are several important angles that, for historical reasons, are called *anomalies*:

$\nu(t)$ = true anomaly: The angle between the direction of the periastron and the current position of the planet as measured from the center of mass.

$E(t)$ = eccentric anomaly: This is the corresponding angle referred to the auxiliary circle (Fig. 1.37) having the radius of the semi-major axis.

The true and eccentric anomalies are related geometrically by

$$\cos\nu(t) = \frac{\cos E(t) - e}{1 - e\cos E(t)} \tag{1.22}$$

$M(t)$ = mean anomaly: This is an angle relating the fictitious mean motion of the planet that can be used to calculate the true anomaly.

For eccentric orbits the planet does not move at a constant rate over the orbit (recall Kepler's second law, i.e. equal area in equal time). However, this motion can be specified in terms of an average rate, or mean motion, by

$$n \equiv \frac{2\pi}{P}$$

The mean anomaly at time $t - T_0$ after periastron passage is defined as

$$M(t) = \frac{2\pi}{P}(t - T_0) \equiv n(t - T_0) \tag{1.23}$$

The relation between the mean anomaly, $M(t)$, and the eccentric anomaly, $E(t)$, is given by Kepler's equation

$$M(t) = E(t) - e\sin E(t) \tag{1.24}$$

To compute an orbit you get the position of planet from Eq. (1.23), solving for E in the transcendental equation (1.24) and then using Eq. (1.22) to obtain the true anomaly, ν.

1.5.3 The Radial Velocity Curve

We now look into how the orbital parameters influence the observed RV curve. Referring to Fig. 1.37, a planet moving by a small angle $d\nu$ will sweep out an area $\frac{1}{2}r^2 d\nu$ in a time dt. By Kepler's second law $r^2 d\nu\, dt =$ constant. The total area of an

ellipse is just $\pi a^2(1-e^2)^{1/2}$ which is covered by the orbiting body in a period, P. Therefore

$$r^2\frac{d\nu}{dt} = \frac{2\pi a^2(1-e^2)^{1/2}}{P} \tag{1.25}$$

The component of r along the line of sight is $r\sin(\nu+\omega)\sin i$. The orbital velocity is just the rate of change in r:

$$V_0 = \sin i\left[r_1\cos(\nu_1+\omega)\frac{d\nu_1}{dt} + \sin(\nu_1+\omega)\frac{dr_1}{dt}\right] + \gamma$$

The subscripts "1" and "2" refer to the star and planet, respectively. The above equation is thus for the star whose reflex motion we are measuring. The term γ comes from the overall radial velocity of the barycenter and is often called the γ-velocity. Since we are only interested in relative velocity measurements we do not care about the γ-velocity which is irrelevant for the mass determination.

We can use Eqs. (1.21) and (1.25) to eliminate the time derivatives and arrive at

$$V_0 = K_1\left[\cos(\nu+\omega) + e\cos\omega\right]$$

$$K_1 = \frac{2\pi a_1\sin i}{P(1-e^2)^{1/2}} \tag{1.26}$$

For eccentric orbits there will be a maximum positive RV value of the orbit, and a maximum negative value. If we let A_1 and B_1 represent the absolute values of these quantities, then $A_1 \neq B_1$ and

$$A_1 = K_1(1+e\cos\omega)$$
$$B_1 = K_1(1-e\cos\omega)$$

and

$$K_1 = \frac{1}{2}(A_1+B_1)$$

The RV curves from Keplerian orbits can have a bewildering zoo of shapes depending primarily on the eccentricity and the view angle from the Earth. Figure 1.38 shows four examples of Keplerian orbits. Circular orbits (top left panel) show the familiar sine RV curve. As the eccentricity increases this turns into a step-like function. It is not just the eccentricity that affects the shape, but also ω. The shape of the RV curve can look markedly different for the same eccentricity, but different values of ω (right panels of Fig. 1.38).

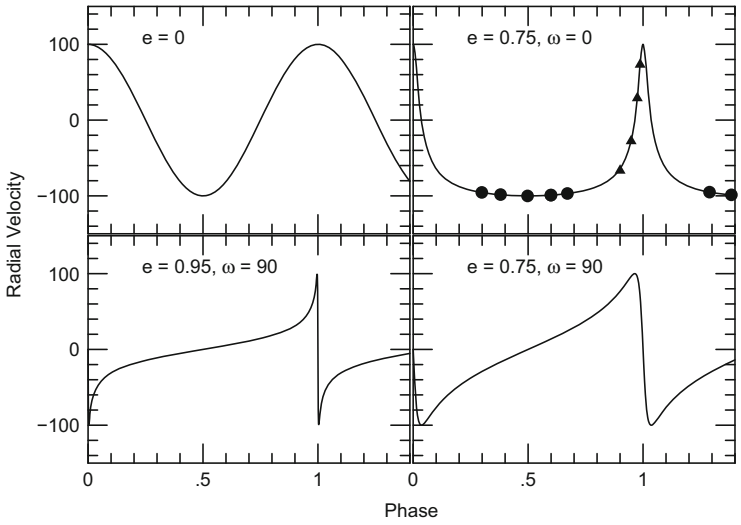

Fig. 1.38 Sample RV curves from Keplerian orbits. (*Top left*) A circular orbit. (*Bottom left*) An eccentric orbit with $e = 0.95$ and $\omega = 90°$. (*Top right*) An eccentric orbit with $e = 0.75$ and $\omega = 0°$. (*Bottom right*) Orbit with the same eccentricity but with $\omega = 90°$. Sparse RV measurements of a star with a planet in an eccentric orbit can have variations that look constant (*dots*) or mimic a binary companion (*triangles*)

Although several exoplanets have been discovered in highly eccentric orbits, the number may be much higher simply because eccentric orbits are difficult to detect. For highly eccentric orbits all the "action" takes place over a very narrow time interval. For instance, if you made RV measurements at the phases of the panel shown as circles in the top right panel of Fig. 1.38, then you would conclude this was a constant RV star without a planet and you would eliminate it from your target list.

Likewise, maybe you have taken your measurements on the steep part of the RV curve (triangles in Fig. 1.38). The star will show a very rapid increase in the RV, similar to what you would expect from a stellar companion. You would then conclude that this was just a boring binary star and you would again eliminate this star from your target list. A good observing strategy would be to continue to observe your target stars even if they appear to be RV constant, or seem to be a stellar binary system.

It is also difficult to get good orbital parameters for exoplanets in highly eccentric orbits. For these you often need to get RV measurements at the extrema of the orbit and this can occur over just a few days. If the planet has a long period orbit and you miss the peak of the RV due to bad weather, you may have to wait several years for the next opportunity.

Note that for circular orbits ω is not defined. A planet in a circular orbit has no periastron passage since it keeps a constant distance from the star. The angle of

periastron passage it thus undefined. In this case the phase of the orbit is simply defined by the epoch or mean anomaly.

There is one instance worth noting where you do use a non-zero ω for a circular orbit. This is the case when you have a transiting planet whose epoch is usually defined by the mid-point of the transit and not by another epoch. In this case the RV curve has to have the correct phase, i.e. one where the velocity of the host star goes negative shortly after the mid-transit. Since the phase of the orbit can no longer be absorbed in the epoch which is now fixed, one must change ω in order to have the correct phase. In this case $\omega = 90°$ for circular orbits.

1.5.3.1 Periodograms of Eccentric Orbits

Periodograms (DFT and LSP) find the amplitudes of sine functions in your RV time series. Naively, one would expect that these cannot find eccentric orbits in RV data. This is not the case. Recall that periodograms are essentially Fourier transforms, and for eccentric orbits you not only see the peak of the dominant orbital frequency, but also at its harmonics.

Figure 1.39 shows the LSP of an RV curve having an eccentricity of 0.75 and a period of 74.5 days ($\nu = 0.013$ day^{-1}). The LSP easily finds the correct frequency of the orbit, but also its higher harmonics. All the visible peaks are integer multiples of the orbital frequency ($\nu = n \times 0.013$ day^{-1} where n is an integer). If one sees a peak in the LSP and at least one of its harmonics, then one should fit a Keplerian

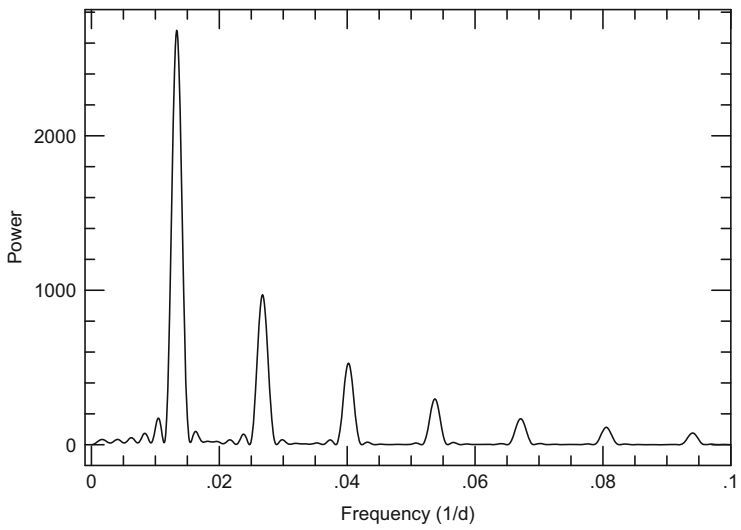

Fig. 1.39 Lomb–Scargle periodogram of RV data from an eccentric orbit ($e = 0.75$). The dominant peak is at the orbital frequency $\nu_{\rm orb} = 0.0133$ day^{-1}. Other peaks occur at harmonics $\nu = n\nu_{\rm orb}$ where n is an integer

orbit with non-zero eccentricity to see if the harmonic arises from an eccentric orbit. Otherwise, you may reach the wrong conclusion that you have a multi-planet system with all the planets in resonant orbits.

1.5.4 The Mass Function

Once we have calculated the orbital elements of a planet we need to derive its mass and this comes from the mass function. We can write Kepler's law

$$\frac{G}{4\pi^2}(M_1 + M_2)P^2 = (a_1 + a_2)^3$$

where now we have included the components of both the star (1) and planet (2).

$$= a_1^3 \left(1 + \frac{a_2}{a_1}\right)^3$$

$$= a_1^3 \left(1 + \frac{M_1}{M_2}\right)^3$$

where we have used $M_1 a_1 = M_2 a_2$

$$\frac{G}{4\pi^2}(M_1 + M_2)P^2 \sin^3 i = a_1^3 \sin^3 i \left(\frac{M_1 + M_2}{M_2}\right)^3$$

From Eq. (1.26) we can solve for $a_1 \sin i$ in terms of K, P, and e. After substituting and re-arranging we arrive at

$$f(m) = \frac{M_2^3 \sin^3 i}{(M_1 + M_2)^2} = \frac{K_1^3 P (1 - e^2)^{3/2}}{2\pi G} \approx \frac{M_2^3 \sin^3 i}{{M_1}^2} \tag{1.27}$$

where for the later expression we have used the fact that $M_1 \gg M_2$ for planetary companions.

Equation (1.27) is known as the mass function, $f(m)$, that can be calculated from the orbital parameters K, P, and e. There are two important things to note about the mass function. First, it depends on the stellar mass, M_1. This means if you want to get a good measurement of the mass of your planet, you have to know the mass of the star. In most cases this is an educated guess based on the spectral type of the star. Only in cases where you have asteroseismic measurements (e.g., Hatzes et al. 2012) or the host star is a component of an astrometric binary do you know the stellar mass.

Second, you do not derive the true mass of the planet, M_2, only $M_2^2 \sin^3 i$ which you have to take the cube root to get $M_2 \sin i$. The Doppler effect only gives you one component of the velocity of the star. The orbital inclination can only be measured using astrometric measurements, or for transiting planets. In the latter case the orbital inclination must be near 90° even to see a transit.

It is important to note that the mass function is the only quantity related to the mass of the planet that is derived from orbital solutions. It is also a quantity that is constant for a given system and as such it should *always* be given in publications when listing orbital parameters derived from RV measurements. The stellar mass may change with refined measurements, but the mass function stays the same. With a published mass function value it is easy for the reader to calculate a new planet mass given a different mass for the host star.

1.5.5 Mean Orbital Inclination

RV measurements, on their own, will never give you the true mass of the companion, only the minimum mass, or the product of the mass times the sine of the orbital inclination. Only in cases where you have a transiting planet will RVs alone give you the true mass. A large number of RV measurements coupled with a few astrometric measurements can also give you the companion mass (Benedict et al. 2002, 2006).

What if you just so happen to be viewing an orbit nearly perpendicular to the planet. In this case this might just be a stellar binary companion. Since you can only derive the minimum planet mass with the RV method, is it really all that useful? It is therefore important to ask "What is the probability that the companion mass is much higher than our measured value?" Also, for a random distribution of orbits: "What is the mean orbital inclination?" The probability that an orbit has a given orbital inclination is the fraction of celestial sphere that orbit can point to while still maintaining the same orbital orientation, i. This gives a probability function of $p(i) = 2\pi \sin i\, di$. The mean inclination is given by

$$\langle \sin i \rangle = \frac{\int_0^\pi p(i) \sin i\, di}{\int_0^\pi p(i)\, di} = \frac{\pi}{4} = 0.79 \tag{1.28}$$

This has a value of 52° and you thus measure on average 80 % of the true mass.

We have seen that for orbits it is the mass function [Eq. (1.27)], $f(m)$, that is important and $f(m) \propto \sin^3 i$. So for orbits the mean value of $\sin^3 i$ is what matters:

$$\langle \sin i \rangle = \frac{\int_0^\pi p(i) \sin^3 i\, di}{\int_0^\pi p(i)\, di} = 0.5 \int_0^\pi \sin^4 i\, di = \frac{3\pi}{16} = 0.59 \tag{1.29}$$

How likely is it that we are viewing an orbit perpendicular to the orbital plane ($i \sim 0°$)? The probability that the orbit has an angle i less than a value θ is

$$p(i < \theta) = \frac{2\int_0^\theta p(i)di}{\int_0^\pi p(i)di} = (1 - \cos\theta) \tag{1.30}$$

The probability that an orbit has an inclination less than 10° is about 1.5 %. So fortunately for RV measurements viewing orbits perpendicular to the orbital plane is not very likely. This does not mean that this never happens. RV measurements found evidence for a planet candidate around the star HD 33636 with a minimum mass of 10.2 $M_{\rm Jup}$ (Vogt et al. 2002). Astrometric measurements made with the Fine Guidance Sensor of the Hubble Space Telescope measured an orbital inclination of 4° which resulted in a true companion mass of 0.142 $M_\odot$ (Bean et al. 2007). The companion is not a planet but a low mass star! The probability of this occurring is small, about 0.2 %, but with a thousand or so planet candidates discovered by the RV method it sometimes will happen.

The GAIA astrometric space mission should be able to derive true masses for most of the giant planets discovered by the RV methods. It could be that a few, but not all, of our exoplanets will disappear.

1.5.6 Calculating Keplerian Orbits with SYSTEMIC

SYSTEMIC is an excellent program if you want to get into the business of calculating Keplerian orbits from RV data (Meschiari et al. 2010). You can download it from the site *www.stafonm.org/systemic*. It has a graphical interface where you can first perform a periodogram to find periodic signals in your RV data, fit these with a Keplerian orbit, subtract this, and look for additional planets in the data. Orbital parameters can then be optimized to find the best solution. SYSTEMIC has a database so that you can compute orbits using RV data from already known exoplanets, but you can also work on your own RV data. A nice feature of the program is that you can combine different data sets with different zero point velocity offsets.

In calculating Keplerian orbits it is important to have good estimates of the parameters. The first step is to get a good estimate of the period which comes from the periodogram analysis. You then fit the velocity K-amplitude and find an appropriate phase. The eccentricity is the hardest to fit and should be the last parameter to be varied. Fortunately, SYSTEMIC has a nice graphical interface where it is easy for the user to vary parameters and see how they fit the data. Once you have a good estimates of P, K, e, and ω you can optimize all parameters simultaneously. It is best to start with known planet systems that are easy, and work your way up the difficulty scale.

1.6 Sources of Errors and False Planets

In this lecture I will address the sources of errors in making RV measurements. These errors will be divided into two broad categories: (1) errors associated with the instrument, image stability, or in the data analysis ("human") and (2) those associated with intrinsic stellar variability ("stellar"). Strictly speaking, errors induced by stellar variations are not really errors. After all, your RV measurements are detecting a real physical phenomenon on the star. However, in the exoplanet business this stellar variation hinders our ability to detect planets and adds an "error" to the RV value caused by the orbiting planet. One person's signal is another person's noise!

The first category of errors are called "human" because it is in our power to minimize these either through more stable spectrographs, better detectors or improved data analyses. We can at least do something about reducing these errors. Some examples of such errors include:

1. Guiding errors
2. Changes in the spectrograph (i.e., IP)
3. Wavelength calibration
4. Detectors
5. Barycentric corrections

It is important to note that the final RV measurement error of an instrument results from a total error budget. Every component starting with the image stability at the slit/fiber, going through the optical components and optical coatings, the thermal and mechanical stability of components of the spectrograph, the detector, and finally the stability and accuracy of the wavelength calibration all contribute to the error. To achieve the best RV precision you need to minimize the errors of every component contributing to the budget. It makes no sense to invest money on, say, a fancy wavelength calibration device only to have instrumental instability (thermal or mechanical) contributing a larger error to the RV measurements.

The second category which I call "stellar" errors result from physical phenomena associated with the star. These errors are more pernicious than the human errors simply because there is basically little we can do to minimize these. We simply cannot tell the stars to behave so that we can detect their planets! Our only hope is that we might be able to correct the measured RV for the stellar variability. One price we have paid for achieving such exquisite RV precision is that now the stellar variability is dominating the error in the RV measurements when it comes to detecting exoplanets.

1.6.1 Instrumental Errors

1.6.1.1 Guiding Errors

Recall that the spectrograph is merely a camera that produces a dispersed image of the slit/fiber at the detector. If there is any image motion at the entrance to the spectrograph, then this would manifest itself as instrumental Doppler shifts of the lines as measured by the detector. This is particularly a problem for slit spectrographs used under good seeing conditions. Telescopes usually guide on starlight that does not enter the slit (or fiber), but that is reflected into the guide camera by pick-off mirrors. If the seeing is too good, then all of the light will go down the slit and light is only reflected into the guide camera when the star has moved off the slit. This results in large image motion.

Some of these guide errors can be minimized by using optical fibers. Fibers have good scrambling abilities so that in spite of image motion at the fiber entrance, the output light beam is reasonably stable. For improved stability astronomers often use double scramblers—having the light go through two separate fibers. The cost for this improved stability is light loss. Recently, hexagonal fibers are coming into use because these have superior scrambling abilities compared to circular fibers.

Even when using optical fibers guiding errors can still come into play in subtle ways. Figure 1.40 shows RV measurements of a star taken with the fiber-fed FIES spectrograph of the Nordic Optical Telescope. Each point represents the RV measured from a single spectral order. These are all from the same star, so the RV should be constant. However, the measured RV systematically decreases by almost $200\,\mathrm{m\,s^{-1}}$ as one goes from red to blue wavelengths.

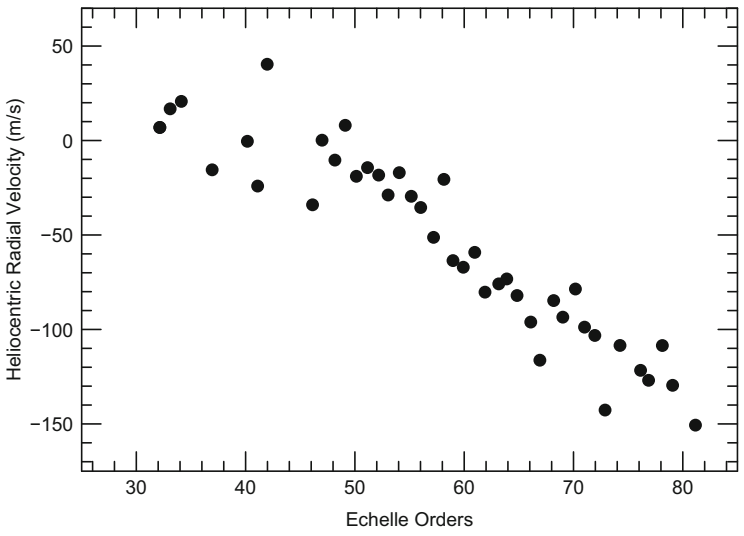

Fig. 1.40 The heliocentric radial velocity of a target star as a function of echelle spectral orders. Red wavelengths are at lower spectral orders (Courtesy of Davide Gandolfi)

The reason for this is the Earth's atmospheric dispersion. When observing at high air mass the star at the entrance fiber to the spectrograph is no longer a white light point source, but rather a dispersed stellar image. The guide camera probably has a filter so it is guiding on one color of the stellar image. The other colors are not hitting the fiber properly resulting in a systematic Doppler shift as a function of wavelength. This effect can be minimized by using an atmospheric dispersion corrector (ADC). This optical device corrects for the dispersion of the atmosphere so a (single) white light image is hitting the entrance slit or optical fiber.

1.6.1.2 Changes in the Instrumental Setup

We have seen how changes in the IP of the spectrograph can introduce instrumental shifts in the RV measurements. Most of these changes are subtle and due to changes in the temperature, alignment of the optics, mechanical shifts, etc., in other words, things that the observer has little control over. It is important when making precise RV measurements that the observer does not change the instrumental setup of the spectrograph thus introducing changes in the IP.

Recall that the image of spectral lines represents an image of the entrance slit/fiber convolved with the IP and the intrinsic broadening due to the star (e.g., rotation). If you make slight changes at the entrance, this will appear at the image (detector).

An example of how a slight change in the instrumental setup can produce a velocity offset is the case of the purported planet around the M-dwarf star VB 10. Astrometric measurements detected a planet with a true mass of 6.5 M_{Jupiter} in a 0.74-yr orbit (Pravdo and Shaklan 2000). The left panel of Fig. 1.41 shows the RV confirmation measurements for this planet phased to the orbital period, along

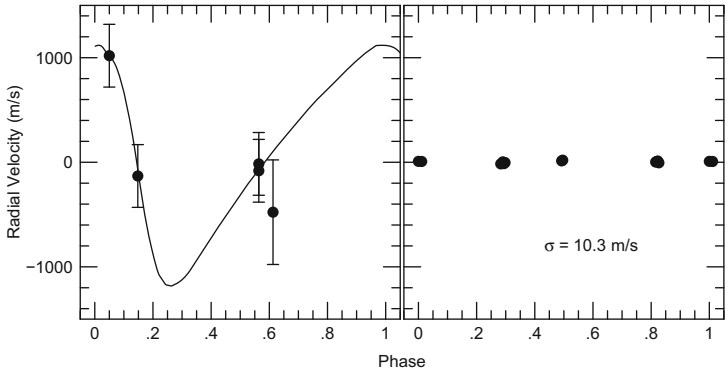

Fig. 1.41 (*Left*) RV measurements of VB 10 phased to the orbital period of the presumed planet (Zapatero Osorio et al. 2009). The outlier (*square*) was a measurement taken with a different slit width. (*Right*) RV measurements taken with the CRIRES spectrograph (Bean et al. 2010), again phased to the orbital period. No planet signal is seen in these data

with the orbital solution (Zapatero Osorio et al. 2009). The RV measurements were made using the telluric method for lines found in the near infrared, so in principle instrumental shifts should be eliminated. However, one should immediately be wary that the orbital solution, particularly the eccentricity, seems to be driven by only one RV measurement. Remove this point and one sees no RV variations. It is thus important to confirm this point with additional measurements. Whenever you see such data in the literature, it "screams" for more data to be taken.

The right panel of Fig. 1.41 shows RV measurements taken by Bean et al. (2010) using the infrared CRIRES spectrograph at the 8.2 m very large telescope (VLT) of the European Southern Observatory. An ammonia gas absorption cell was used to provide the wavelength calibration. So in this case the instrumental shifts should also be minimized. These measurements are also phased to the orbital period and on the same velocity scale as the left panel of Fig. 1.41. The RV measurements are constant to within $10.3\,\mathrm{m\,s^{-1}}$ and there is no indication for the presence of a planetary companion.

So what went wrong with the earlier measurements, particularly the one outlier that defines the presence of a planet? On careful reading of Zapatero Osorio et al. (2009) one finds that the outlier was taken with a slightly different instrumental setup. The observers had widened the entrance slit to the spectrograph. This was probably done to allow more light to enter the spectrograph (possibly observing conditions were marginal). This is a bad idea because by decreasing the resolution of the spectrograph this changed the IP. Large Doppler shifts can also be introduced by the details in how the slit jaws move. Does one or both sides move? And if both sides move, do they do so exactly in unison? The price paid for a higher signal-to-noise ratio spectrum is not worth the degraded precision this causes due to changes in the IP.

When making precise RV measurements one must never, ever make any changes to the spectrograph setup. Keep everything as constant as possible. Maintain the same slit width for all observations. If you are using a fiber-fed spectrograph and have a choice of fibers for different resolution, always use the same fiber diameter for all measurements. If you changed the entrance slit, used a different fiber, replaced a calibration lamp, changed the CCD detector, introduced a filter into the light path, etc., all these will introduce instrumental shifts in your RV measurements. If you make a change in the spectrograph, all subsequent measurements should be considered as being taken with a different instrument that has a different zero point velocity offset. These are independent data sets, you simply cannot throw all the data together and search for planets.

1.6.1.3 Effects of the Detector

The last, and often ignored, component of the RV budget train for an instrument is the CCD detector. Generally these detector systems are provided by observatories for a broad range of uses. The CCD laboratory of an observatory will certainly

have stability in mind, but maybe not always at a level needed for precise RV measurements.

CCD detectors can introduce RV errors in a variety of ways. We have already discussed such things as flat field errors, and errors due to fringing. Other sources can be the structure of the CCD and the inter-pixel gaps. Errors can occur as an absorption line moves across such a gap. As a rule of thumb when making precise RV measurements it is best to always place your stellar spectrum at the same location on the CCD detector. Of course this is not always possible as barycentric motion of the Earth will always displace the spectral lines.

Noise in the CCD detector can also creep in as an RV error in very subtle ways. Figure 1.42 shows the RV measurement error as a function of time for star from the Tautenburg Observatory Planet Search Program. For the most part the RV error shows an RMS scatter of 5–7 $\mathrm{m\,s^{-1}}$, which is what we expect for the typical signal-to-noise level. However, at a time JD = 2454500 the RV showed a factor of 3–4 increase in the error. An inspection of the reduced data showed nothing out of the ordinary, all the data were taken at high-signal-to-noise ratios, in good conditions, and with the same exposure times. What was going on?

The bias level[5] of the CCD gave us the first hint that the problem was in the CCD detector. Rather than being at the normal level of 100 analog-to-digital units

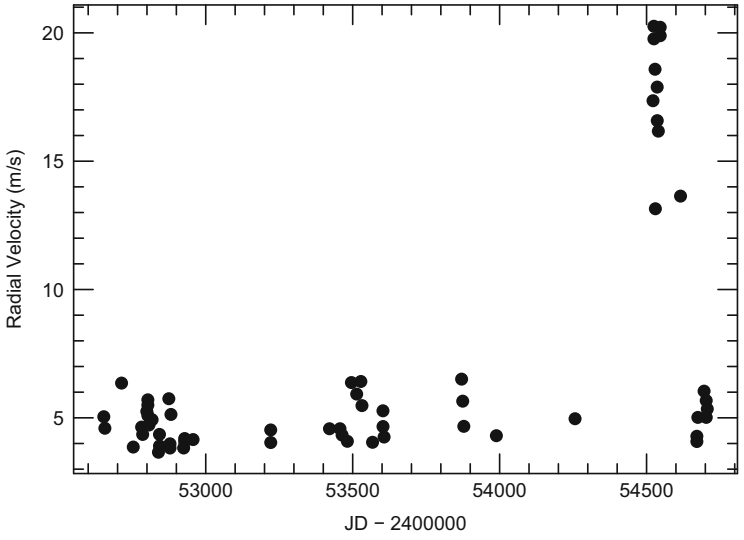

Fig. 1.42 The RV measurement error as a function of time for a star from the Tautenburg Observatory Planet Search Program. The outliers were measurements taken when there was CCD electronics were picking up noise from the signal generators that drove the dome motors

[5]The bias level is a voltage offset applied to the data to ensure that no negative data values enter the analog-to-digital converter.

(ADU), it was a factor of 10 higher. After some investigations by the technical staff of the observatory the problem was traced to noise from telescope motors. The signal generators that were previously used to drive the telescope motors used a sine wave which has a clean Fourier spectrum (a ∂-function). This would not introduce noise into the CCD readout electronics and a good RV measurement was possible. At the recommendation of the manufacturer the signal generators were changed to ones that produced a square wave signal which was better for the motors. As we have seen, a square wave has a very messy Fourier power spectrum with many components over a wide range of frequencies. One of these Fourier frequencies hit a resonance with the CCD electronics and this introduced noise into the CCD control system electronics. Once this was isolated the RV measurements had errors that returned to their normal values as shown by the last measurements.

So, should you as an observer be concerned with what the technical day crew is doing to your telescope far from your spectrograph? The lesson learned is yes! This is an obvious case where the error introduced was large and some detective work could find the source of the error. But what if we want to push the RV error down to a few cm s^{-1}, could we even notice measurable changes in the stability of the CCD? So an important part of the RV error budget is a very stable detector which not always taken into account when pushing RV measurements to higher precision.

1.6.2 Laser Frequency Combs

At the heart of any instrument for precise stellar radial velocity measurements is the wavelength calibration. In order to have an accurate calibration of the wavelength scale one needs densely packed calibration emission or absorption lines. Molecular iodine have a dense forest of lines, but these are of unequal strength. The accuracy of the iodine calibration is set by the accuracy of the FTS used to provide the calibration spectrum. Furthermore, the iodine absorption lines contaminate the stellar spectrum. Th-Ar has a density of emission lines that vary as a function of wavelength and there are some spectral regions where one finds few emission lines for calibration. Plus, the wavelengths of the thorium lines still has to be measured in a laboratory and these also have their own measurement errors. All of these add errors that degrade the quality of the wavelength calibration.

Laser frequency combs (LFC) offer a promising solution to better wavelength calibration. LFCs produce an absolute and repeatable wavelength scale using features that are equally spaced across the spectrum. Femtosecond pulses from a mode-locked laser occur at a pulse repetition rate f_{rep}. In frequency this yields a spectrum $f_n = f_0 + nf_{\mathrm{rep}}$ where f_0 is the carrier offset frequency and n is a large integer, $n \sim 10^5$–10^6. Frequencies are stabilized using an atomic clock.

Figure 1.43 is a spectrum of a star taken with the LFC installed at the HARPS spectrograph. One can see that compared to the Th-Ar spectrum (Fig. 1.17) the laser combs provide a much denser set of calibration lines. First use of the LFC at HARPS indicates that an RV precision of $\sim$ cm s^{-1} is possible (Lo Curto et al.

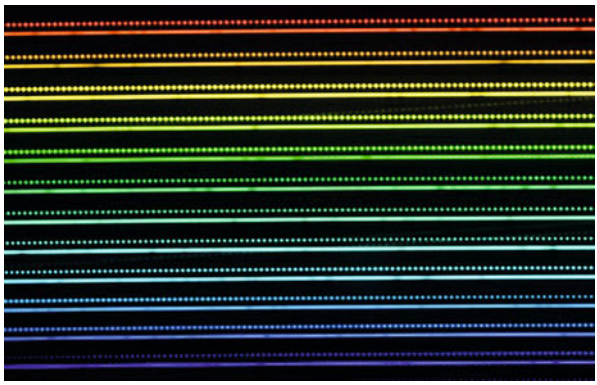

Fig. 1.43 The laser frequency comb at the HARPS spectrograph. The *lower rows* are the orders of the stellar spectrum. Just above these are a series of emission peaks produced by the laser frequency comb. Note the higher density of calibration features compared to Th-Ar (see Fig. 1.17). Figure courtesy of ESO

2012). Currently LFC is rather expensive (≈0.5 million Euros) and requires trained personal for its operation. However, in the near future LFC should be cheaper and "turn key" devices, that is, you buy it, plug it in, and immediately (almost) start making RV measurements.

1.6.3 Telluric Features

As we discussed earlier, the telluric method can be a simple and inexpensive way to improve the precision of RV measurements. However, if you are employing other wavelength calibration, these will contaminate your stellar spectrum and degrade your RV precision. Telluric features mostly from water vapor, are stronger towards the infrared; however, telluric features can be found around ≈5700 Å and these start to become more denser longward of 5900 Å. Thus telluric contamination is a concern even at optical wavelengths.

Telluric features have a more or less fixed wavelength (outside of shifts due to wind a pressure changes) so in principle these can be removed from the spectra before calculating your RV. However, it is best to simply mask off these spectral regions in the analysis process.

1.6.4 Barycentric Corrections

Depending on the time of year and the location of a star in the sky, the orbital motion of the Earth can introduce a Doppler shift of $\pm 30\,\mathrm{km\,s^{-1}}$ to your measured RV. The rotation of the Earth can introduce a shift of $\pm 460\,\mathrm{m\,s^{-1}}$. If you have not removed the barycentric and rotational motion of the Earth to a high accuracy, then this will introduce errors in your RV measurements, or worse create periodic "fake" planets in your data. This is particularly true if you want to use the RV method to find

terrestrial planets in the habitable zone of G-type stars. These will induce a reflex motion of the star of only $10\,\mathrm{cm\,s^{-1}}$ and with a period of 1 year. You do not want our own Earth to be the only habitable planet you find in your RV data!

Most RV programs use the JPL Solar System Ephemeris DE200 (Standish 1990) to calculate the Earth's barycentric motion. This corrects for the Earth's motion to a level of a few $\mathrm{cm\,s^{-1}}$. However, errors in the barycentric correction can creep into your results in subtle ways.

1. *Inaccurate observatory coordinates*
   The latitude, longitude, and altitude of your observatory need to be known very well. For example, an error of 100 m in the height of the observatory can introduce an error at the $1\,\mathrm{cm\,s^{-1}}$ levels (Wright and Eastman 2014). Fortunately, modern GPS system can get these coordinates to very good precision.
2. *Inaccurate time of observations*
   Exposure times for spectral observations for an RV measurement range from a few to 30 min. It is important to have an accurate measurement of the time for these observations. Typically, one simply takes the mid-point in time of your observations. However, what happens if there are transparency changes in the Earth's atmosphere during your exposure, say clouds have moved in for the last half of the exposure? In this case the time of arrival of most of the photons will be different from the "geometric" mid-time of your observation. For this reason most RV programs use an exposure meter to calculate the intensity weighted time of exposure.
3. *Inaccurate position of stars*
   One needs accurate positions of your target stars. For example, if you want to have a barycentric correction better than $10\,\mathrm{cm}^{-1}$ so that you can detect habitable exo-earths, you would need to know the position of your stars to better than a few mill-arcseconds. Fortunately, the astrometric mission GAIA will soon provide us with extremely accurate stellar positions down to V-magnitude = 20. Of course, in accounting for the position of stars you also will have to take into account the precession and nutation of the Earth.
4. *Proper motion of stars*
   One may have an accurate position for your target star, but stars move in space! If you observe your star a year later, it will no longer have the same coordinates due to its proper motion. Barycentric corrections should take into account the changing coordinates of the star due to proper motion. Again, GAIA should give us accurate proper motions for a large number of stars. Another error comes from the secular acceleration due to proper motion which I will discuss below.
5. *Differential barycentric motion*
   During your exposure the Earth's barycentric motion is changing slightly. This means that spectral lines on your detector are moving during the exposure resulting in a slight blurring of the lines. To my knowledge no program takes this into account. This effect probably does not have a large influence for Doppler

measurements at a precision of $\approx 1\,\mathrm{m\,s^{-1}}$, but may be important when pushing measurements to the few $\mathrm{cm\,s^{-1}}$ precision.

An excellent paper discussing the sources of errors to barycentric corrections is Wright and Eastman (2014).

1.6.5 The Secular Acceleration

There is another error resulting from the proper motion of a star and that is the so-called secular acceleration. This acceleration is not actually an error, but arises from the different viewing angle of a high proper motion star (right panel of Fig. 1.44). Imagine a high proper motion star that is approaching you. When it is at a large distance you will measure a blueshifted velocity (−v). As the star approaches you the tangential velocity of the star increases at the expense of the radial component. When it is crossing your line of sight the radial velocity goes through zero and on to positive values. When the star is far away the radial velocity you measure is a red-shifted value, +v. The secular acceleration depends on the proper motion of the star, and your viewing angle.

The left panel of Fig. 1.44 shows the RV measurement of Barnard (Kürster et al. 2000), the star which has the highest proper motion. The predicted secular acceleration of the star (line) fits the observed trend in the RVs quite well (line). If this is not taken properly into account, you would think this linear trend was due to a companion (stellar or planetary) to the star.

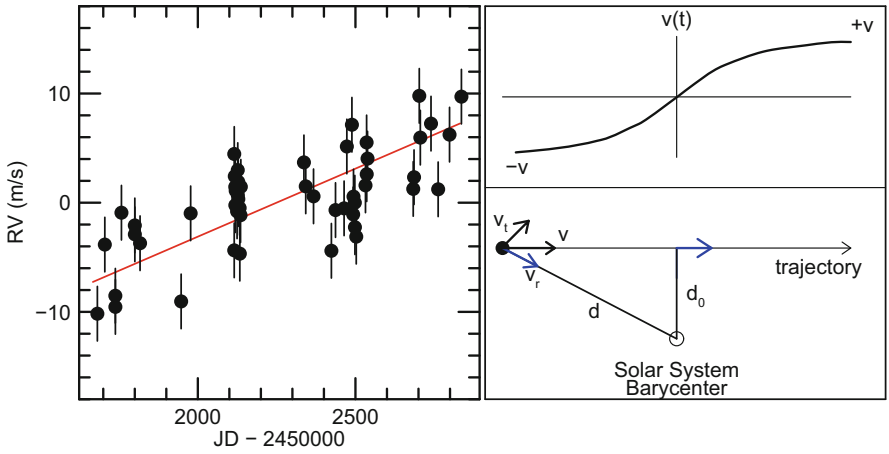

Fig. 1.44 (*Right*) Schematic showing how the secular acceleration arises due to the proper motion of the star. Far from the observer the star has a high radial velocity, $\pm\ v$. As the star crosses the line of sight the radial velocity passes through zero. (*Left*) RV measurements for the high proper motion star Barnard (Kürster et al. 2000). The *line* represents the predicted secular acceleration

1.6.6 Stellar Noise

The instrumentation for precise stellar RV measurements has improved to the point that astronomers can now routinely measure the RV of a star with a precision of $\approx 1\,\mathrm{m\,s^{-1}}$. At this precision the error in the RV measurement is dominated by intrinsic stellar noise rather than instrumental errors. Table 1.3 shows some of the sources of stellar noise, their amplitudes, and time scales.

Stellar oscillations, for the most part, do not represent a real problem for the RV detection of exoplanets. For solar-like stars these amplitudes are small ($\approx 0.5\,\mathrm{m\,s^{-1}}$), but more importantly, the time scales are short (5–15 min). By taking an exposure of your star that is longer than the time scale of the oscillations or co-adding several observations you can "beat down" the oscillation noise.

Changes in the convection pattern of the star have time scales associated with the stellar activity cycle, typically years to decades for sun-like stars (our Sun has an 11-year solar cycle). This is generally a problem when trying to find long period planets, particularly Jovian analogs. (Coincidentally, the orbital period of Jupiter is comparable to the solar cycle.)

The largest source of stellar noise errors for many sun-like stars is the activity in the form of cool spots, hot faculae, and plage. The left panel of Fig. 1.45 shows a simulation of a spot distribution on a sun-like star that is rotating at $2\,\mathrm{km\,s^{-1}}$. The

Table 1.3 Sources of intrinsic stellar RV noise

Phenomenon	RV amplitude ($\mathrm{m\,s^{-1}}$)	Time scales
Solar-like oscillations	0.2–0.5	∼5–15 min
Stellar activity (e.g., spots)	1–200	∼2–50 days
Granulation/Convection pattern	∼ few	∼3–30 years

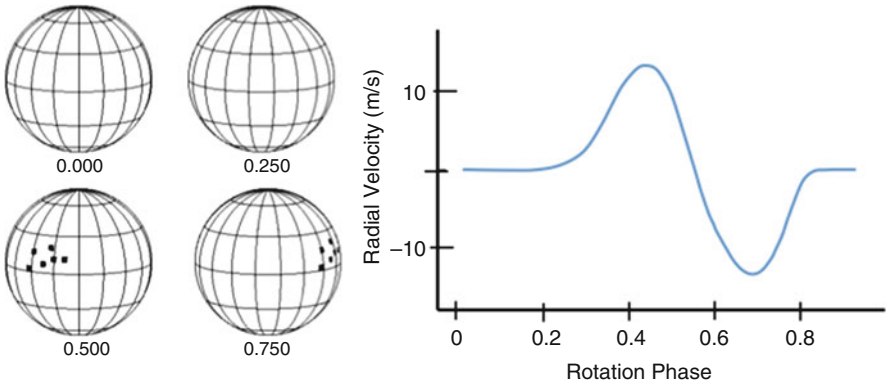

Fig. 1.45 (*Left*) A simulated spot distribution on a solar-like star rotating at $2\,\mathrm{km\,s^{-1}}$. (*Right*) The corresponding RV variations due to this spot distribution

spot coverage is a few percent. The right panel shows the RV variations due this spot distribution and it amounts to a peak-to-peak amplitude of $20\,\mathrm{m\,s^{-1}}$.

The RV amplitude due to spots depends not only on the spot filling factor, but also on the rotational velocity of the star. Saar and Donahur (1997) estimated this to be

$$A_{\mathrm{RV}}[\mathrm{m\,s^{-1}}] \approx 6.5 v \sin i f^{0.9} \tag{1.31}$$

where f is the spot filling factor in percent and $v \sin i$ is the rotational velocity of the star in $\mathrm{km\,s^{-1}}$. The RV was measured using the centroid of the spectral line.

Hatzes (2002) confirmed this behavior, but with a slightly higher amplitude, using simulated data where the RV was calculated using a procedure that mimicked the iodine method.

$$A_{\mathrm{RV}}[\mathrm{m\,s^{-1}}] \approx (8.6 v \sin\, i - 1.6) f^{0.9} \tag{1.32}$$

which is valid for $v \sin\, i > 0.2\,\mathrm{km\,s^{-1}}$.

For long-lived spots ($\Delta T >$ a stellar rotation) the RV variations due to activity will be coherent and appear as a periodic signal. A well-known case for this is the star HD 166435. This star showed RV variations with a period of 4 days that mimicked the signal due to a hot Jupiter. It was shown however, that these variations were actually due to spots (Queloz et al. 2000).

Unfortunately, spots are born, grow, and eventually fade away. They also migrate in both longitude and latitude, the latter which will have a different rotational period due to differential rotation. This will create a very complex time series which greatly hinders our ability to detect planets, particularly ones with small RV amplitudes. One has to be able to distinguish between RV signals due to planets and those due to activity. This requires the use of activity diagnostics.

1.6.6.1 Activity Diagnostics

In the next lecture I will cover "tricks" that can be used for extracting the planetary signal in the presence of noise. In the current lecture I will mention some diagnostics one can use to determine if your RV signal is actually due to activity.

Common diagnostics used for checking if RV variations are due to activity are (1) photometric measurements, (2) Ca II H& K measurements, and (3) spectral line bisectors.

Ca II and Brightness Variations

Starspots are typically ≈2000–3000 K cooler than the surrounding photosphere. Consequently, they will produce intensity and color variations with the rotation period of the star. Active stars show an emission feature in the core of the Ca

II H & K lines that is due to the presence of a chromosphere. This emission is often measured through a so-called S-index (e.g., Baliunas et al. 1985). Since this emission is concentrated in plage regions the S-index will also show variations with the rotation period of the star.

Spectral Line Bisectors

Spots produce distortions in the spectral lines which causes RV variations. A common way of measuring the line shapes is via the spectral line bisector: the loci of the mid-points of a spectral line measured from the core to the continuum (Gray 1982). Bisectors are quantified using either the bisector span, the slope of the bisector between two arbitrarily chosen points on the bisector, or the curvature which is normally the difference in slopes between the upper and lower halves of the bisector (approximately the derivative of the slope). The line bisector has become a common tool for confirming exoplanet discoveries (Hatzes et al. 1998).

Figure 1.46 shows the measurements of all these diagnostic quantities for the spotted star HD 166435 showing that the RV variations are indeed due to a spot.

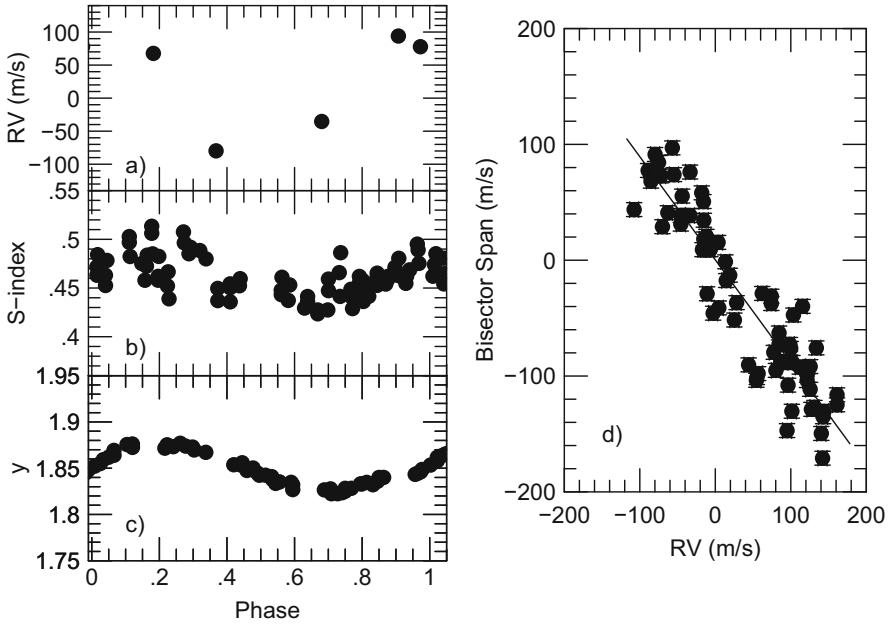

Fig. 1.46 (*Left*) Time series and phased values for the RV (*panel a*), S-index (*panel b*), and y-mag (*panel c*) of HD 166435. (*Right*) The bisector span versus the RV for HD 166435 (Queloz et al. (2000))

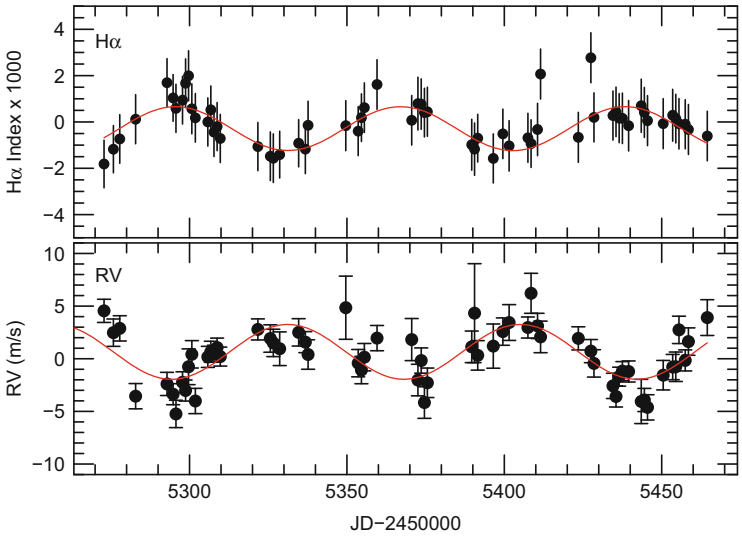

Fig. 1.47 (*Top*) The variations of the Hα index from Robertson et al. (2014) and a sine fit (*curve*). (*Bottom*) The RV variations attributed to the planet GJ 581d. Both data were produced by pre-whitening the original data in order to isolate the variations at the orbital period of GJ 581d (66-day)

Balmer Hα

The Balmer Hα line has emerged as a powerful diagnostic for excluding the RV variations due to stellar activity in M-dwarf stars. Kürster et al. (2000) found that RV variations due to Barnard correlated with the changes in the equivalent width of Hα. Robertson et al. (2014) used Hα index measurements to refute the planet in the habitable zone of GJ 581. They found that Hα correlated with the 66-day RV variations due to the purported planet GJ 581d. Figure 1.47 shows the Hα index measurements of Robertson et al. after applying our old friend pre-whitening to remove all periodic signals in the Hα data except those at the 66-day orbital period of GJ 581d. These Hα variations are 180° out-of-phase with the RV variations attributed to the planet. In this case stellar surface activity produces RV variations that mimicked a low mass planet in the habitable zone of the star.

Infrared RV Measurements

The contrast between cool spots and hot photosphere decreases as one goes to longer wavelengths. This contrast ratio can be estimated using the black body law:

$$F_p/F_s = \frac{e^{hc/k\lambda T_s} - 1}{e^{hc/k\lambda T_p} - 1} \tag{1.33}$$

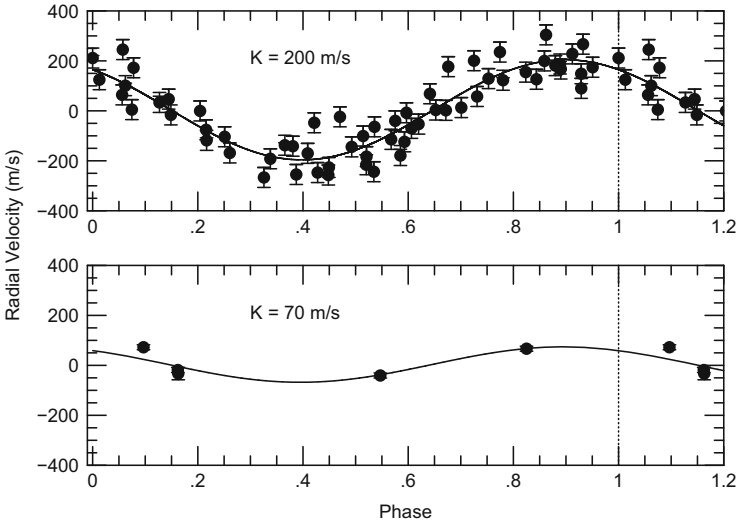

Fig. 1.48 (*Top*) The RV variations of TW Hya measured at optical wavelengths and phased to the orbital period of the purported planet (Setiawan et al. 2008). (*Bottom*) The RV variations of TW Hya measured at near infrared wavelengths (Huélamo et al. 2008) and phased to the planet orbital period. The amplitude in this case is at least a factor of 3 smaller indicating the variations arise from a spot. Measurements are repeated past the vertical line at phase 1.0

where λ is the wavelength of light, h and c the standard constants (Planck and speed of light), T_P is the photospheric temperature, about 5800 K for a solar-like star, and T_s is the sunspot temperature, about 3000–4200 K for sunspots.

For starspots that are ≈2000–3000 K cooler than the surrounding photosphere produce a contrast ratio over a factor of ten less at 1.5 μm compared to observing at 5500 Å. Thus a useful diagnostic is to make RV measurements at different wavelengths. For a planetary companion the RV amplitude should be constant as a function of wavelength. On the other hand, spots would produce an RV amplitude that is smaller in the infrared region. TW Hya provides a nice example of the utility of RV measurements in the IR for confirming planet detections. Setiawan et al. (2008) reported the presence of a 9.8 M_{Jup} mass companion in a 3.56 days orbit based on RV measurements made in optical regions (Fig. 1.48). This was potentially an important discovery since TW Hya is a young T Tauri star that still has its proto-planetary disk. The authors excluded activity as a cause based on no variations in the line bisectors, although this was based on rather low resolution data (R = 50,000).

Infrared RV measurements point to another phenomenon as the source of the RV variations (Huélamo et al. 2008). These show an RV amplitude for the "planet" that is one-third the amplitude of the variations in the optical (Fig. 1.48). The RV variations are almost certainly due to a spot on the surface of the star.

TW Hya highlights the pitfalls of using line bisectors to confirm planets. Many rapidly rotating active stars like the weak T Tauri star V410 Tau have large polar spots (e.g., Hatzes 1995). When viewed from the poles these spots could produce

detectable RV variations, yet small bisector variations that would be hard to detect, particularly at lower spectral resolutions. The reason for this is that the bisector span is mostly influenced by the mid-regions of the spectral line. When measuring the bisector span one avoids the cores of the lines and the wings since these have the largest errors (Gray 1982). Unfortunately, polar spots, especially when viewed at relatively low stellar inclinations, produce most of the distortions to the stellar line near the core. This can still produce a large RV variation, but because the core of the spectral lines is often avoided in calculating the bisector one may see little variation in bisector span. If any variations are to be seen, they might be detected in the bisector curvature. This is certainly the case for TW Hya which has stellar inclination of only about $= 7°$, i.e. we view the rotation pole of the star.

So the rule of thumb should be if you see line bisector variations with the RV period, you do not have a planet. If you do not see any bisector variations, this may not necessarily confirm the planet. In other words, a lack of bisector variations is a necessary, but not necessarily a sufficient condition to confirm a planetary companion.

I should note that rotational modulation can also produce harmonics in the periodogram in much the same way that an eccentric orbit does. Surface structure rarely produces a pure sine-like RV variation. It always is more complicated showing not only the rotational frequency, but also its harmonics. For example, if you have two spots on opposite sides of the star, you will see half the rotational period, or twice the rotational frequency. Three spots would produce three times the rotational frequency and so forth. A complicated spot distribution can produce several harmonics in the periodogram.

1.7 Dealing with Stellar Activity

1.7.1 Pre-Whitening

We have seen in Sect. 1.4 how pre-whitening can be used to extract multi-periodic signals in RV data due to planetary systems. Pre-whitening is also a useful tool for filtering out the RV variations due to activity. Although it is tied to the stellar rotation, the RV variations due to activity are not always strictly periodic. This is because surface features (spots, plage, etc.) are born, evolve, and eventually decay. Spots formed at, or migrating to, different latitudes on the star will have different rotation periods due to differential rotation. The result is that the RV signal due to activity will be complex, appearing periodic possibly for only short time segments.

Because of this complex activity RV variations one might think that there should be no physical basis for fitting these with a series of sine functions, which is what pre-whitening does. There might not be a strong physical basis, but there is a mathematical basis for doing this. This is due to the fact that sines (and cosines) are a set of orthogonal functions that form a basis set. This means that most functions

can be fit by a linear combination of sines (or cosines) even if they do not appear periodic. If you doubt this, use *Period04* to pre-whiten the function $y = x$. With enough terms you can use sine functions to fit a straight line.

CoRoT-7 is one case where pre-whitening worked extremely well at finding planets in the presence of activity signals. CoRoT-7 has a transiting rocky planet that was discovered by the CoRoT space mission (Léger et al. 2009). Radial velocity measurements confirmed that rocky nature of the planet (Queloz et al. 2009), and found an additional two planets that do not transit (Hatzes et al. 2010). CoRoT-7 is a modestly active star as the CoRoT light curve shows an $\approx$2 % modulation with the $\approx$24 days rotation period of the star (Léger et al. 2009).

The top panel of Fig. 1.49 shows the RV measurements for CoRoT-7 taken with the HARPS spectrograph. The line shows the multi-sine component fit coming from both activity and planets found by the pre-whitening process (Table 1.4). The largest variations are due to spots on the stellar surface.

The middle panel of Fig. 1.49 shows the activity-only RV signal after removing the contribution of the planets and the fit. The lower panel shows the planets only RV variations and the fit after removing the activity signal. Figure 1.50 shows the planets of CoRoT-7 phased to their respective orbital periods.

So in this case pre-whitening has uncovered eight significant periodic signals. How do we know which is due to a planet and which is due to stellar activity?

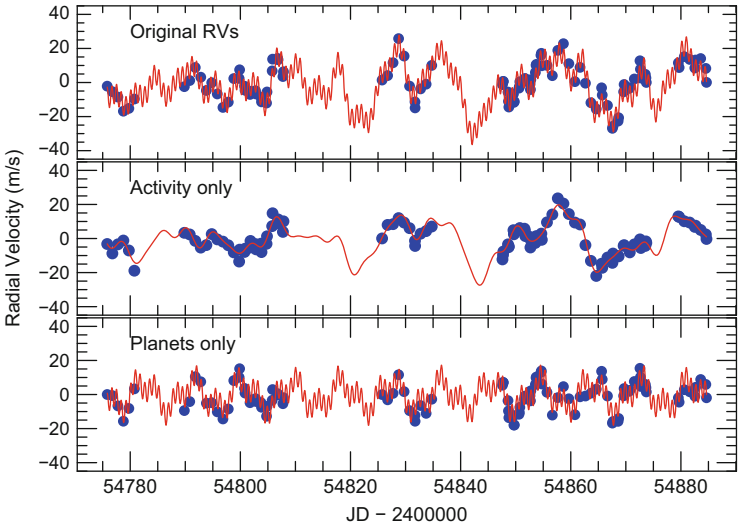

Fig. 1.49 (*Top*) The RVs for CoRoT-7 measured by HARPS. The *curve* represents the multi-sine component fit found by pre-whitening. (*Center*) The RV variations and fit to the activity-only variations for CoRoT-7. (*Bottom*) The RV variations and fit to the three planets of CoRoT-7

Table 1.4 Frequencies found in the CoRoT-7 RV data with pre-whitening

	ν (day^{-1})	P (days)	K (m s^{-1})	Comment
f_1	0.0448	22.32	9.18	Activity (rotation period)
f_2	0.1108	9.03	7.02	Planet
f_3	0.0959	10.43	6.07	Activity
f_4	0.2708	3.69	5.21	Planet
f_5	1.1704	0.854	5.75	CoRoT-7b
f_6	0.1808	5.53	3.28	Activity
f_7	0.0353	28.33	5.49	Activity
f_8	0.0868	11.52	3.05	Activity

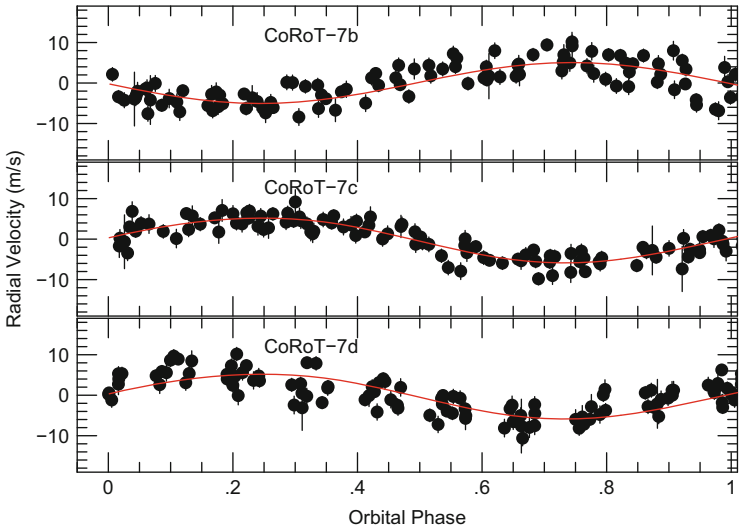

Fig. 1.50 The three planets to CoRoT-7 found after removing the activity signal using pre-whitening. The RVs are phased to the respective orbital period for each planet

One test is to compare the periodogram of your RVs to those of standard activity indicators. Figure 1.51 compares the LSP of the RV, line bisectors, Ca II S-index, and full width at half maximum (FWHM) of the cross-correlation function. Activity signals appear as peaks in all four quantities. The planet signals (vertical dashed lines) only appear in the RV periodogram. Note that for clarity we have only plotted to a frequency of 0.3 day^{-1}, so the orbital frequency of CoRoT-7b (1.17 days^{-1}) does not appear. However, one can see its alias (marked by a vertical line) at 0.17 day^{-1}.

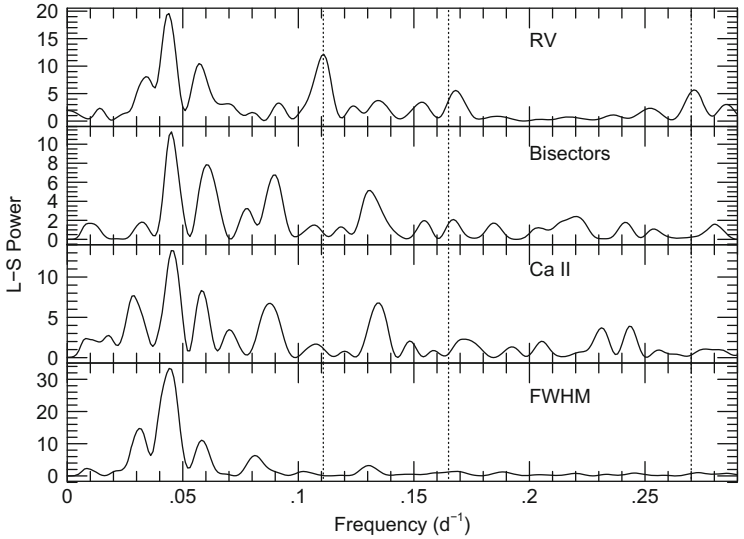

Fig. 1.51 The LSP of the RV variations (*top*), bisector velocity span (*center top*), Ca II S-index (*center bottom*), and FWHM of the CCF (*bottom*). *Vertical dashed lines* mark the positions of the three planets in the system. The orbital frequency of CoRoT-7b does not appear on this scale, however its alias at 0.17 day^{-1} can be seen in the RV periodogram

1.7.2 Trend Fitting

Fitting Fourier components (pre-whitening) to the activity RV signal can produce good results, but the user should always use caution. Removing dominant peaks from the Fourier amplitude spectrum will *always* make the remaining peaks look more significant and it is easy to isolate a noise peak and make it look like a significant signal.

This is demonstrated in Fig. 1.52. This shows the LSP of the residual RV from simulated data after applying the pre-whitening process. It shows a peak at ν = 0.309 c day^{-1} (P = 3.2 days) that is statistically significant with a FAP = 0.0004. The simulated data mimicked the HARPS RV data for α Cen B (Dumusque et al. 2012). A synthetic RV signal due to activity was generated using sine functions from the dominant Fourier components from the RV activity signal in α Cen B. These synthetic data were sampled in the same manner as the real data and noise at a level of 2 m s^{-1} was also added. There was no periodic signal at ν = 0.309 day^{-1} that was used to generate the data, only the simulated activity signal was present. In this case pre-whitening produced a significant looking signal at a frequency not present in the data. The sampling of the activity signal, noise, and the pre-whitening process conspired to make a noise signal look like variations due to a planetary companion.

It is therefore important, if possible, to check the results of pre-whitening with independent methods. One such method is trend filtering. Figure 1.53 compares

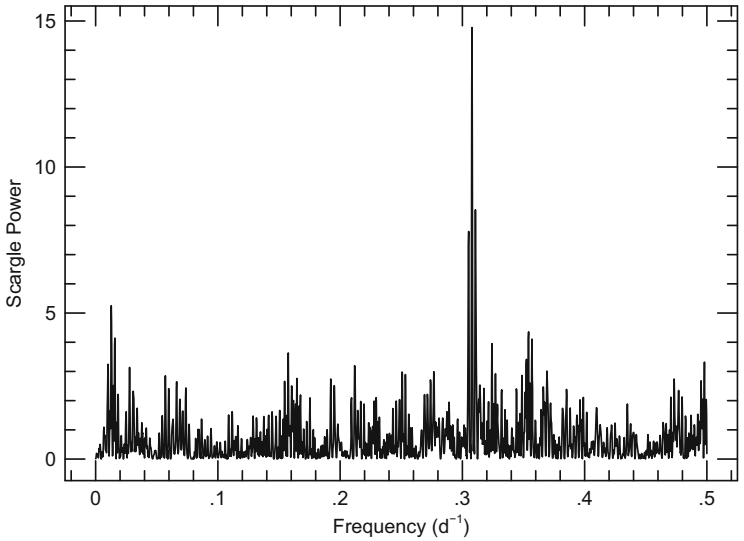

Fig. 1.52 LSP of a synthetic data set consisting of a simulated activity signal that has been pre-whitened. The peak at $\nu = 0.309$ day^{-1} is an artifact due to pre-whitening as no signal at this frequency was inserted in the data

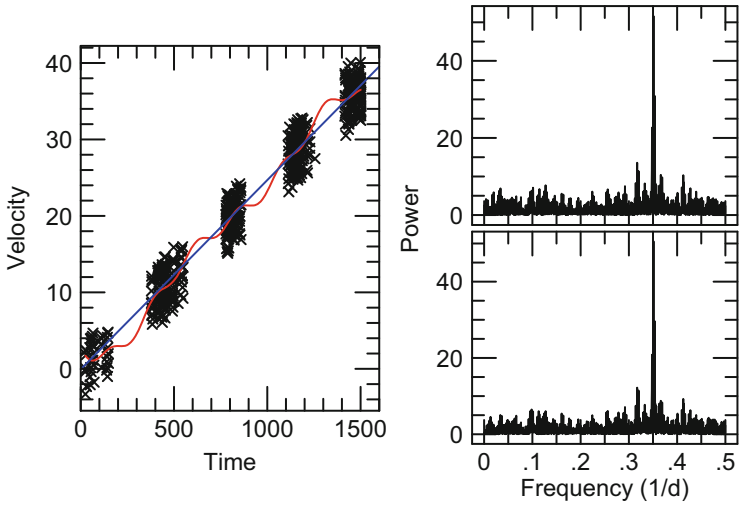

Fig. 1.53 (*Left*) Synthetic RVs consisting of a planet signal superimposed on a long term trend. The *blue line* shows the fit to the trend from a linear fit, the *red line* from the pre-whitening sine coefficients. The LSP of the RV residuals after removing the trend by fitting a straight line (*top*) or by pre-whitening the data (*bottom*). Both methods for removing the trend produce consistent results

how the two methods, trend filtering and pre-whitening, work. The left panel shows simulated RV measurements consisting of a planet signal superimposed on a linear trend. Random noise at a level of $2\,\mathrm{m\,s^{-1}}$ has also been added to these data. Clearly, one needs to remove this trend to detect the signal of the underlying planet.

The obvious method is simply to fit a straight line to the trend (blue line in figure), subtract it, and then perform a period analysis on the residual RV data. The less obvious method is to use pre-whitening to find the dominant sine components to the trend (red line), remove it, and look at the residual RVs. In this case both methods find the signal of the planet (right panels Fig. 1.53). Note that pre-whitening has one disadvantage in that if the planet orbital frequency is near that of one of the sine components needed to fit the trend, then it will be hidden.

What if the activity signal over a long time span is more complicated than a linear trend? Local trend fitting breaks the time series into chunks small enough that the underlying trend can be fit with simple functions. The length of each time chunk is defined by two time scales. The first is the orbital period of the planet P_p. The second time scale is that of activity signal which we take as the rotation period of the star, P_{rot}. One then divides the RV time series in chunks of time length, ΔT, such that

$$P_p < \Delta T < P_{\mathrm{rot}} \tag{1.34}$$

Over the time span ΔT the RV due to activity should be slowly changing so that one can fit it with a low order polynomial. If the activity signal looks coherent over one or two rotation periods, one can even use sine functions to fit the activity. One then removes this *local* trend in the chunk, adds all the residuals from the individual chunks, and use these to search for periodic signals due to planets.

This can also be understood in terms of Fourier components. We are removing low frequency components due to activity to search for much higher frequency components due to the orbit of a companion.

This local trend filtering was used to cast doubt on the planetary companion to α Cen B (Hatzes 2013a). Figure 1.54 shows a subset of the RV measurements for α Cen B. The solid line represents the local trend fit to the variations. The variations of the planet are shown by the dashed line. Note that the variations of the presumed planet should change more rapidly than the changes in the RV due to activity.

The top panel of Fig. 1.55 shows the LSP of the α Cen B RVs after pre-whitening the data leaving only the signal due to the planet. One can see a peak at the frequency of the planet ($\nu = 0.309\ \mathrm{day^{-1}}$) that is reasonably significant and consistent with the Dumusque et al. (2012) result. The lower panel of Fig. 1.55 shows the LSP of the RV residuals after removing the activity variations with local trend fitting. The signal of the planet has nearly disappeared and the weak signal has a FAP of $\approx$40 %. In this case pre-whitening of the activity signal gives a vastly different result than using trend fitting. Tests of the trend fitting method indicate that it should have found the planet at a much higher significance (Hatzes 2013a). Unlike in Fig. 1.53 where the two methods produce consistent results, for α Cen B there is a discrepancy. This does not mean that the planet is not there, just that its presence is not as significant as previously thought.

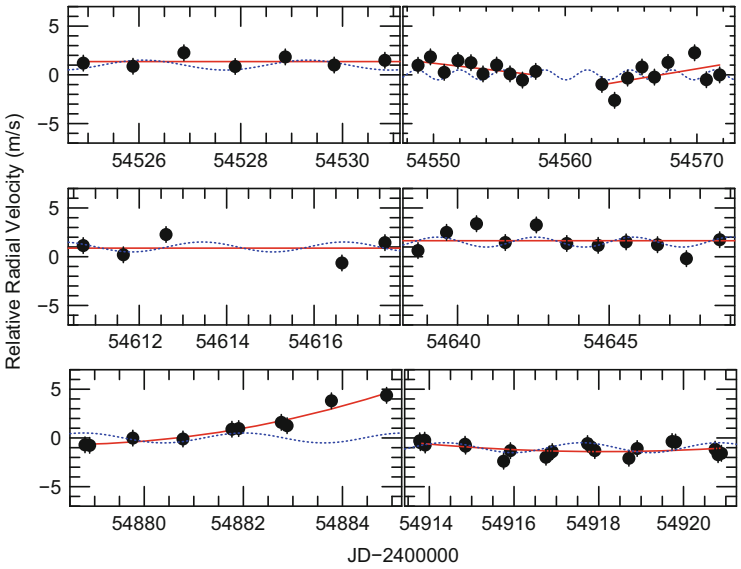

Fig. 1.54 Subsets of the α Cen B RVs showing local trend fits (*solid line*) to the activity variations. The *dashed line* shows the expected RV variations from the planet α Cen Bb

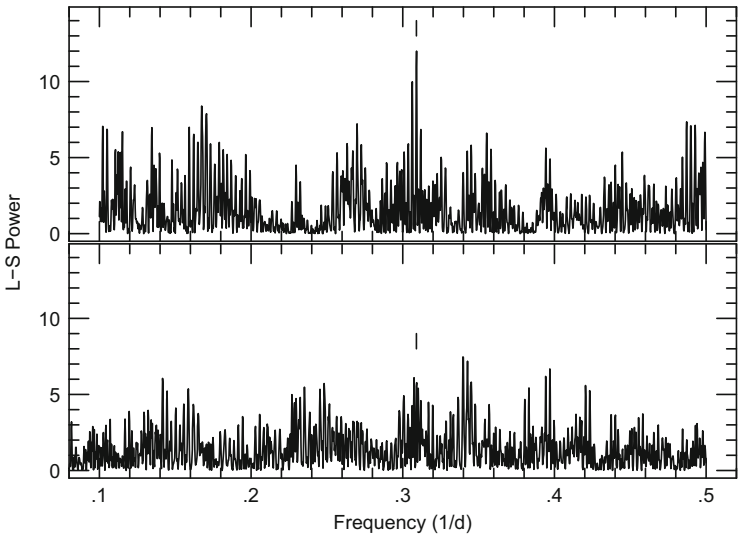

Fig. 1.55 (*Top*) The LSP periodogram of the α Cen B residual RVs produced by filtering the activity using pre-whitening. (*Bottom*) The LSP of the residual RVs for α Cen B after using local trend filtering. The power of the planet signal (*vertical mark*) is greatly reduced using local trend filtering

Hatzes (2013a) proposed that the RV signal due to the planet was an artifact of the activity signal combined with the sampling window. Filtering the data could result in the presence of a planet-like signal like the one shown in Fig. 1.52. Indeed, recent work by Rajpaul et al. (2016) indicates that the 3.24 days planet signal is a ghost of signal that was present in the window function. In interpreting a signal in a time series one must be careful not only of signals coming from other phenomena, such as activity, but also how these interact with the sampling window.

1.7.3 Floating Chunk Offset

Ultra-short period planets with periods less than one day offer us another way to filter out the activity signal. Such short period planets were unexpected until the discovery of CoRoT-7b (Léger et al. 2009) and Kepler-10b (Batalha et al. 2011), both with periods of 0.82 day. The shortest period planet discovered to date is Kepler-78b with an orbital period of a mere 0.35 day or 8 h (Sanchis-Ojeda et al. 2013)!

For these short period planets we can exploit the fact that the orbital period of the planet is much shorter than the rotation period of the star and thus the time scale for stellar activity. Figure 1.56 shows how we can we can exploit this short orbital period. If the planet has a short-period orbit, over the course of one night ($\Delta T \approx 8$ h) you will observe a significant fraction of its orbit. Assuming a circular orbit, this will be a short segment of a sine wave (blue segments in figure). If the rotation period of the star is much longer than planet's orbital period, then the RV

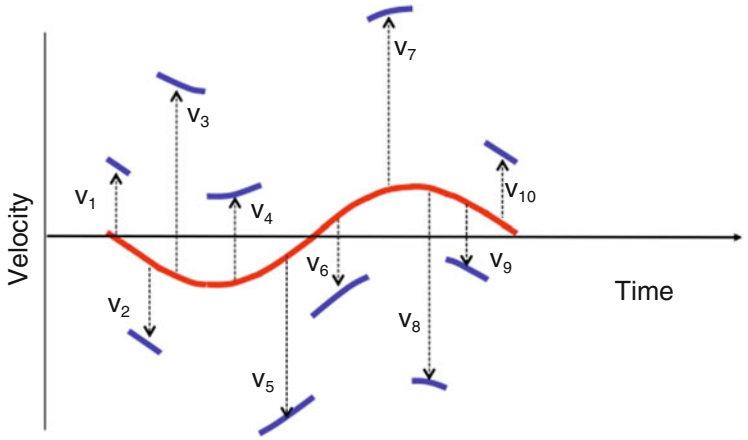

Fig. 1.56 Schematic of the floating chunk offset (FCO) method. Observations of a short period planet on a given night appear as short segments of a sine function for circular orbits. The segment on a given night has a velocity (v_i) offset due to activity, long period planets, systematic errors, etc. By finding the appropriate offset one can line up the segments to recover the planet's orbit (*curve*)

contribution from spots is constant since there has not been enough time for the star to rotate significantly, or for spots to evolve. The spot distribution is essentially frozen-in on the stellar surface and it creates a velocity offset, v_1, for that first night.

Note the we do not care what is causing this velocity offset. It could be spots, it could also be systematic errors, even additional long period planets. All we care about is that during the course of the night the RV contribution from all these other phenomena remains more or less constant.

We then observe the star on another night when stellar rotation has moved the spots or these have evolved so that we have a different view of the stellar activity. This will create a different velocity offset v_2. If we do this for several nights and phase the data to the orbital period of the planet, we will see segments of sine waves each having a different velocity offset v_i. All we have to do is calculate the best offset v_i for each segment, in a least squares sense, that causes all these segments to align on the orbital RV curve. This method was used to provide a refined measurement of the mass of CoRoT-7b (Hatzes et al. 2011). Since the offsets in each time chunk is allowed to "float" I refer to this technique as the floating chunk offset (FCO) method.

Figure 1.57 shows the results of applying the FCO method to the RV measurements used to determine the mass of the transiting Earth-sized planet Kepler-78b (Hatzes 2014). These data were taken with different instruments, the Keck HIRES (Howard et al. 2013) and the HARPS-N spectrograph (Pepe et al. 2013). Because these are different instruments with different RV techniques (iodine method for the

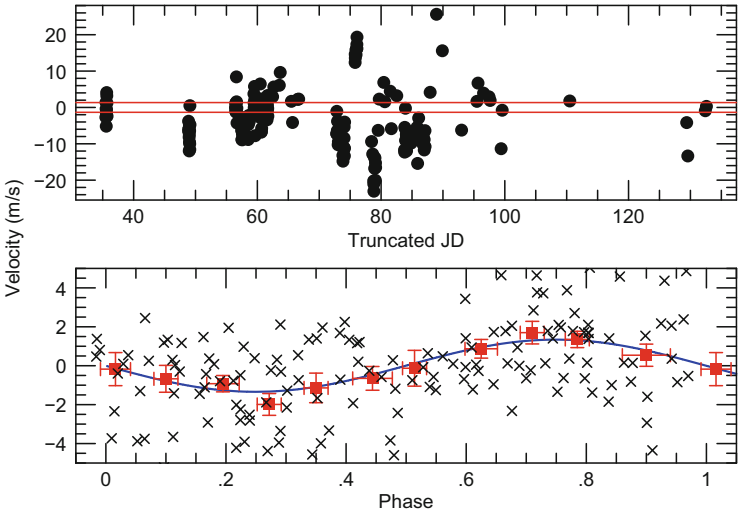

Fig. 1.57 (*Top*) The RV measurements for Kepler-78. The two *horizontal lines* show the RV extrema of Kepler-78. Most variations are due to stellar activity. (*Bottom*) The RV measurements of Kepler-78b phased to the orbital period after using the FCO method to remove the activity signal. *Squares* represent binned values

Keck data and simultaneous Th-Ar for the HARPS-N data) each data set has a different zero point offset. The top panel shows the combined RV data after putting both sets on the same zero point scale. Most of the RV variations ($\pm 20\,\mathrm{m\,s^{-1}}$) are due to the stellar activity. The horizontal lines mark the extrema of the RV variations due to Kepler-78b which are about a factor of ten smaller than the activity variations. The orbital period of Kepler-78b which is 0.35 days is about a factor of 30 smaller than the rotational period of the star at 10.4 days.

The lower panel of Fig. 1.57 shows the RV orbital curve due to Kepler-78b after applying the FCO method to the combined RV data sets. The FCO method is very effective at filtering out the low frequency "noise" due to activity (Fig. 1.58)

The FCO method can also be used as a periodogram to search for unknown, short period planets in your RV data. Basically, you take a trial period and find the velocity offsets that provide the best sine fit to the data. Look at the reduced χ^2. Try another trial period. You then plot the reduced χ^2 as a function of input periods to find the one that best fits the data.

The FCO periodogram for the Kepler 78 RV data is shown in Fig. 1.59. The reduced χ^2 is plotted with decreasing values along the ordinate so that the minimum value appears as a peak, like in the standard periodogram. The best fit period to the Kepler 78 data is indeed at 0.35 day. Even if we did not know a transiting planet was present, the FCO periodogram would have detected it.

It is also possible to use the FCO periodogram on eccentric orbits. One simply uses a Keplerian orbit with non-zero eccentricity rather than a sine wave (zero eccentricity).

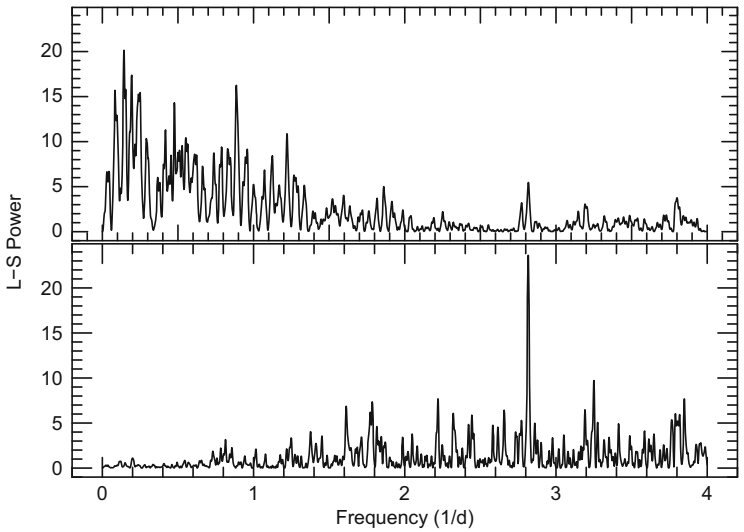

Fig. 1.58 The LSP of the RV measurements for Kepler-78b before (*top*) and after (*bottom*) applying FCO filtering. Note that all the low frequency components due to stellar activity have been filtered out which enhances the planet signal

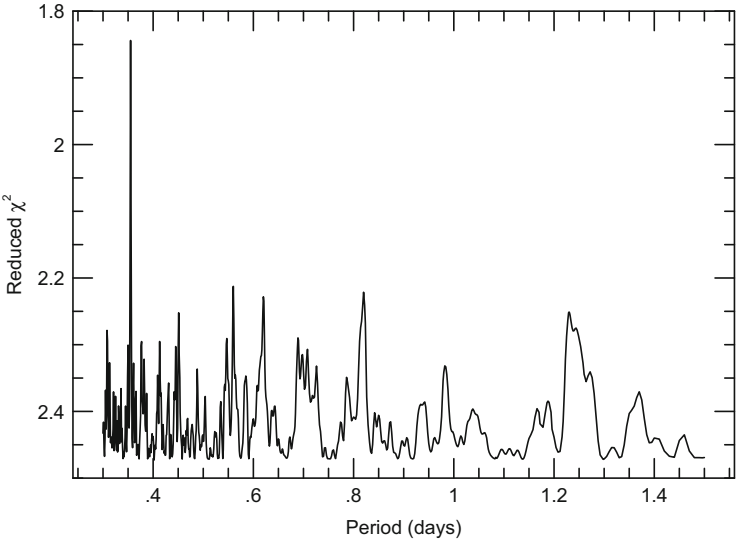

Fig. 1.59 The FCO periodogram of the Kepler-78 RV measurements. Note how the filtering enhances the detection of Kepler-78b

1.8 Closing Remarks

This lecture series has covered all aspects of the radial velocity method as it pertains to the detection of exoplanets. As we have seen it involves a broad range of techniques that involve instrumentation, analyses, and interpretation. The problem of detecting exoplanets is basically one of detecting small signals in the presence of noise. One can rightfully argue that the demands of exoplanet detection "pushes the envelope" of all aspects of the method.

It was not possible in these short lecture notes to cover adequately all aspects of the method. So much has been left out for lack of space. A detailed and proper discussion of each topic would require its own chapter in a book, not just one small section. There were other "hot" topics in the RV methods such as the detection of planets in RV data using Bayesian estimation, Gaussian processes for filtering out activity, and Monte Carlo Markov chains for estimating errors that were also not included. There were too few lectures and simply too little space to include these here. This is best left for future advanced schools given by lecturers with more expertise on these subjects than I have. These lectures should therefore not be treated as the final word, but rather a starting point where the reader can gather more information on the subject.

Acknowledgements I thank Michael Perryman for his wonderful handbook on exoplanets, the first real textbook that covers all aspects of exoplanet science. It made preparation of some of my lectures much easier. I would like to thank the organizers V. Bozza, L. Mancini, and A. Sozzetti

for putting on a great school on exoplanet detection methods and for being such wonderful hosts during my time in Vietri sul Mare. I also thank my fellow lecturers for teaching me some things and for the time spent together, in particular Andrew Collier Cameron. It is always good to spend some time with a long time friend and colleague, especially on the Amalfi Coast! And last but not least I thank the students for their attention, patience, and enthusiasm. Because of them I will watch more Lewis Black videos in order to improve my lecture style.

References

Anglada-Escudé, G., Butler, R.P.: Astrophys. J. Suppl. Ser. **200**, 15 (2012)
Baliunas, S., Horne, J.H., Porter, A., et al.: Astrophys. J. **294**, 310 (1985)
Batalha, N.M., Borucki, W.J., Bryson, S.T. et al.: Astrophys. J. **729**, 27 (2011)
Bean, J.L., McArthur, B.E., Benedict, G.F., et al.: Astron. J. **134**, 749 (2007)
Bean. J.L., Seifahrt, A., Hartman, H., et al.: Astrophys. J. **711**, 19 (2010)
Benedict, G.F., Henry, T.J., McArthur, B.E., et al.: Astrophys. J. **581**, 115 (2002)
Benedict, G.F., McArthur, B.E., Gatewood, G., et al.: Astron. J. **132**, 2206 (2006)
Borucki, W., Koch, D., Batalha, N., et al.: In Transiting Planets, Proceedings of the International Astronomical Union, IAU Symposium, vol. 253, p. 289 (2009)
Butler, R.P., Marcy, G.W., Williams,E., McCarthy, C., Dosanjh, P., Vogt, S.S.: Publ. Astron. Soc. Pac. **108**, (1996)
Campbell, B., Walker, G.A.H.: Publ. Astron. Soc. Pac. **91**, 540 (1979)
Campbell, B., Walker, G.A.H., Yang, S.: Astrophys. J. **331**, 902 (1988)
Cochran, W.D., Hatzes, A.P., Hancock, T.J.: Astrophys. J. Lett. **380**, 35L (1991)
Demory, B.-O., Gillon, M., Deming, D., et al.: Astron. Astrophys. **533**, 114 (2011)
Dumusque, X., Pepe, F., Lovis, C., et al.: Nature **491**, 207 (2012)
Endl, M., Kürster, M., Els, S.: Astron. Astrophys. **362**, 585 (2000)
Gray, D.F.: Publ. Astron. Soc. Pac. **95**, 252 (1982)
Griffin, R., Griffin, R.: Mon. Not. R. Astron. Soc. **162**, 255–260 (1973)
Guenther, E.W., Wuchterl, G.: Astron. Astrophys. **401**, 677 (2003)
Hatzes, A.P.: Astrophys. J. **451**, 784 (1995)
Hatzes, A.P.: Astron. Nachr. **323**, 392 (2002)
Hatzes, A.P.: Astrophys. J. **770**, 133 (2013a)
Hatzes, A.P.: Astron. Nachr. **335**, 616 (2013b)
Hatzes, A.P.: Astron. Astrophys. **568**, 84 (2014)
Hatzes, A.P., Cochran, W.D.: Astrophys. J. **413** 339 (1993)
Hatzes, A.P., Cochran, W.D., Bakker, E.J.: Nature **392**, 154 (1998)
Hatzes, A.P., Cochran, W.D., Endl, M., et al.: Astrophys. J. **599**, 1383 (2003)
Hatzes, A.P., Dvorak, R., Wuchterl, G., et al.: Astron. Astrophys. **520**, 93 (2010)
Hatzes, A.P., Fridlund, M., Nachmani, G.: Astrophys. J. **743**, 75 (2011)
Hatzes, A.P., Zechmeister, M., Matthews, J., et al.: Astron. Astrophys. **543**, 98 (2012)
Howard, A.W., Sanchis-Ojeda, R., Marcy, G.W., et al.: Nature, **503**, 381 (2013)
Horne, J.H., Baliunas, S.L.: Astrophys. J. **302**, 757 (1986)
Huélamo, N., Figuerira, P., Bonfils, X., et al.: Astron. Astrophys. **489**, 9 (2008)
Jenkins, J.S., Yoma, N., Becerra, R.P, Mahu, R., Wuth, J.: Mon. Not. R. Astron. Soc. **441**, 2253 (2014)
Kürster, M., Endl, M., Els, S., Hatzes, A.P., Cochran, W.D., Döbereiner, S., Dennerl, K.: Astron. Astrophys. **353**, L33 (2000)
Kuschnig, R., Weiss, W.W., Gruber, R., Bely, P.Y., Jenkner, H.: Astron. Astrophys. **328**, 544 (1997)
Latham, D., Stefanik, R.P., Mazeh, T., Mayor, M., Burki, G.: Nature **339**, 38L (1989)
Latham, D.W., Rowe, J.F., Quinn, S.N., et al.: Astrophys. J. **732**, 24 (2011)
Léger, A., Rouan, D., Schneider, J., et al.: Astron. Astrophys. **506**, 287 (2009)

Lenz, P., Breger, M.: Commun. Asteroseismol. **146**, 53 (2005)
Lo Curto, G., Manescau, A., Holzwarth, R., et al.: Proc. SPIE **8446**, 84461W (2012)
Lomb, N.R.: Astrophys. Space Sci. **39**, 447 (1976)
Marcy, G.W., Butler, R.P., Vogt, S.S., et al.: Astrophys. J. **505**, 147 (1998)
Mayor, M., Queloz, D.: Nature **378**, 355 (1995)
McArthur, B., Endl, M., Cochran, W.D., et al.: Astrophys. J. **614**, 81 (2004)
Meschiari, S., Wolf, S.A., Rivera, E., Laughlin, G., Vogt, S., Butler, P.: In: du Foresto, V.C., Gelino, D.M., Ribas, I. (eds.) Pathways Towards Habitable Planets, p. 503. Astronomical Society of the Pacific, San Francisco (2010)
Murdoch, K.A., Hearnshaw, J.B., Clark, M.: Astrophys. J. **413**, 349 (1993)
Pepe, F., Mayor, M., Delabre, B., et al.: Proc. SPIE **4008**, 582 (2000)
Pepe, F., Cameron, A.C., Latham, D.W., et al.: Nature **503**, 377 (2013)
Perryman, M.: The Exoplanet Handbook. Cambridge University Press, Cambridge (2014)
Pravdo, S.H., Shaklan, S.B.: Astrophys. J. **700**, 623 (2009)
Queloz, D., Mayor, M., Weber, L., et al.: Astron. Astrophys. **354**, 99 (2000)
Queloz, D., Bouchy, F., Moutou, C., et al.: Astron. Astrophys. **506**, 303 (2009)
Rajpaul, V., Aigrain, S., Roberts, S.: Mon. Not. R. Astron. Soc. **456**, 6 (2016)
Robertson, P., Mahadevan, S., Endl, M., Roy, A.: Science **345**, 440 (2014)
Saar, S.H., Donahur, R.A.: Astrophys. J. **485** 319 (1997)
Sanchis-Ojeda, R., Rappaport, S., Winn, J.N.: Astrophys. J. **774**, 54 (2013)
Scargle, J.: Astrophys. J. **263**, 865 (1982)
Setiawan, J., Henning, Th., Launhardt, R., Müller, A., Weise, P., Kürster, M.: Nature **451**, 38 (2008)
Standish, E.M.: Astron. Astrophys. **233**, 252 (1990)
Struve, O.: Observatory **72**, 199 (1952)
Valenti, J.A., Butler, R.P., Marcy, G.W.: Publ. Astron. Soc. Pac. **107**, 966 (1995)
Vogel, H.: Astron. Nachr. **82**, 291 (1872)
Vogt, S.S., Butler, R.P., Marcy, G.W., Fischer, D.A., Pourbaix, D., Apps, K., Laughlin, G.: Astrophys. J. **568**, 352 (2002)
Vogt, S.S., Butler, R.P., Rivera, E., et al.: Astrophys. J. **723**, 954 (2010)
Vogt, S.S., Butler, R.P., Haghighipour, N.: Astron. Nachr. **333**, 561 (2012)
Winn, J.N., et al.: Astrophys. J. **737**, 18 (2011)
Wright, J.T., Eastman, J.D.: Publ. Astron. Soc. Pac. **126**, 838 (2014)
Zapatero Osorio, M.R., Martin, E.L., del Burgo, C., Deshpande, R., Rodler, F., Montgomery, M.M.: Astron. Astrophys. **505**, L5 (2009)
Zechmeister, M., Kürster, M.: Astron. Astrophys. **491**, 531 (2009)

Part II
The Transit Method

Chapter 2
Extrasolar Planetary Transits

Andrew Collier Cameron

Abstract An extrasolar planet will transit the visible hemisphere of its host star if its orbital plane lies sufficiently close to the observer's line of sight. The resulting periodic dips in stellar flux reveal key system parameters, including the density of the host star and, if radial-velocity observations are available, the surface gravitational acceleration of the planet. In this chapter I present the essential methodology for modelling the time-dependent flux variation during a transit, and its use in determining the posterior probability distribution for the physical parameters of the system. Large-scale searches for transiting systems are an efficient way of discovering planets whose bulk densities, and hence compositions, can be accessed if their masses can also be determined. I present algorithms for detrending large ensembles of light curves, for searching for transit-like signals among them. I also discuss methods for identifying diluted stellar eclipsing binaries mimicking planetary transit signals, and validation of transit candidates too faint for radial-velocity follow-up. I review the use of time-resolved spectrophotometry and high-resolution spectroscopy during transits to identify the molecular constituents of exoplanetary atmospheres.

2.1 Introduction to Exoplanetary Transits

If the orbital planes of extrasolar planetary systems are randomly oriented in space, a subset of them must lie in planes close enough to the line of sight that one or more planets in a system will transit the disk of the host star. Struve (1952) proposed the idea that the resulting periodic, temporary drops in stellar flux could be used as a planet detection method. Although Struve's idea pre-dated the technology needed to detect transits by half a century, his estimates of the likelihood of transits occurring, and of their depth and duration, have been borne out by observation over the last 15 years. Large-scale surveys from the ground have revealed that roughly one main-sequence star in 1000 hosts a gas-giant planet in the kind of orbit that Struve

A.C. Cameron (✉)
SUPA, School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, KY16 9SS, UK
e-mail: acc4@st-andrews.ac.uk

V. Bozza et al. (eds.), *Methods of Detecting Exoplanets*, Astrophysics and Space Science Library 428, DOI 10.1007/978-3-319-27458-4_2

envisaged. NASA's *Kepler* mission (Borucki et al. 2010) has extended the detection domain into the realm of planets down to terrestrial size, providing the first insights into the size and occurrence distributions of rocky, icy and gaseous planets.

In the first of this series of lectures I discuss the transit probability and its dependence on system geometry, then discuss the physical information that can be deduced in a model-independent way from transits.

2.1.1 Transit Probability

In the absence of any prior knowledge of the system's inclination, the probability of transits being visible over interstellar distances is given by the fraction of the celestial sphere swept out by the planet's shadow (Fig. 2.1).

From the observer's point of view, the apparent separation of the planet and star at mid-transit is conveniently expressed as a dimensionless impact parameter b, expressed in units of the host star's radius R_*:

$$b = \frac{a \cos i}{R_*}. \tag{2.1}$$

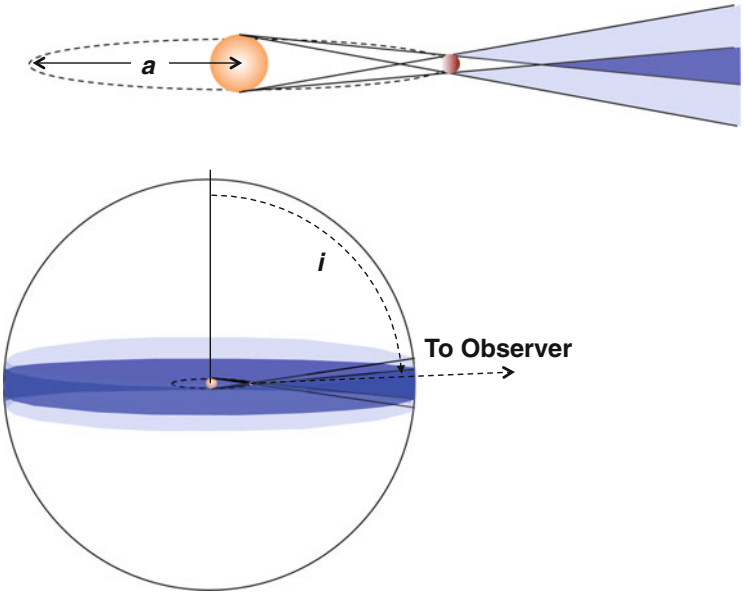

Fig. 2.1 Schematic view of the regions of the celestial sphere from which full (*dark shadow*) and grazing (*light shadow*) transits are visible

Here a is the semi-major axis of the orbit (assumed for the moment to be circular), and i is the angle between the angular-momentum vector of the planet's orbit and the line of sight. A grazing transit of a planet with radius R_p will have $R_* + R_\mathrm{p} > a\cos i > R_* - R_\mathrm{p}$. For a full transit to occur, the inequality $a\cos i > R_* - R_\mathrm{p}$ must be satisfied.

The fraction of the solid angle on the celestial sphere enclosing cones with opening angles in the range i to $i + di$ is given by

$$\frac{d\Omega}{4\pi} = \frac{2\pi \sin i\, di}{4\pi} = \frac{d(\cos i)}{2}. \tag{2.2}$$

For randomly oriented orbits, the probability that grazing or full transits will occur is

$$\Pr\left(\cos i < \frac{R_* + R_\mathrm{p}}{a}\right) = \frac{1}{2}\int_{-(R_*+R_\mathrm{p})/a}^{(R_*+R_\mathrm{p})/a} = \frac{R_* + R_\mathrm{p}}{a}. \tag{2.3}$$

In the typical case where $R_\mathrm{p} \ll R_*$, the probability of transits occurring is simply R_*/a.

$$\Pr\left(\cos i < \frac{R_*}{a}\right) = \simeq 0.0046\left(\frac{R_*}{R_\odot}\right)\left(\frac{1\mathrm{au}}{a}\right). \tag{2.4}$$

This clearly favours the discovery of hot planets in close orbits around their host stars. Transits of Earth are visible from only 0.46 % of the celestial sphere. For Jupiter, orbiting 5.2 AU from the Sun, the probability is only 0.09 %. For this reason, transit surveys conducted with the goal of discovering planets at distances of order 1 AU from their host stars must monitor many thousands of objects for several years.

2.1.2 Early Detections

The first successful detections of extrasolar planets orbiting main-sequence stars were made via the radial-velocity method, which favours discovery of massive planets in close orbits around their host stars. The early radial-velocity discoveries such as 51 Peg b (Mayor and Queloz 1995) had minimum masses characteristic of gas-giant planets. Their radii were expected to be comparable to that of Jupiter, implying transit depths of order 1 %. The hot Jupiters typically lay about ten stellar radii from their host stars, giving roughly a one-in-ten chance that transits would be observable. Once orbital solutions were published, intensive high-precision photometric follow-up campaigns were conducted around the time of inferior conjunction.

Indeed HD 209458b, the 14th radial-velocity planet discovery to be published (Charbonneau et al. 2000; Henry et al. 2000), was the first system in which transits

were found to occur. The first detections were made on the nights of 1999 September 9 and 16 with the prototype STARE instrument, a small f/2.9 Schmidt camera of focal length 289 mm, operated on a tripod in the parking lot of the High-Altitude Observatory in Boulder, Colorado by then-graduate student David Charbonneau. Analysis of the transits yielded a fractional flux deficit $\Delta f/f = 1.7\,\%$ and duration from first to fourth contact of about 3 h.

2.1.3 Transit Depths and Durations

The fractional flux deficit at mid-transit corresponds approximately to the ratio of the projected areas of the planet and star (Fig. 2.2):

$$\frac{\Delta f}{f} \simeq \left(\frac{R_\mathrm{p}}{R_*}\right)^2 = 0.0105 \left(\frac{R_\mathrm{p}}{R_\mathrm{Jup}}\right)^2 \left(\frac{R_*}{R_\odot}\right)^{-2}. \tag{2.5}$$

In practice, the transit depth overestimates the ratio of areas if the limb darkening of the stellar photosphere is not taken into account. For a star with specific intensity I_0 at disk centre and a linear limb-darkening law with limb-darkening coefficient u, the specific intensity of the point behind the centre of the planet at mid-transit is

$$I = I_0(1 - u(1 - \mu)) \tag{2.6}$$

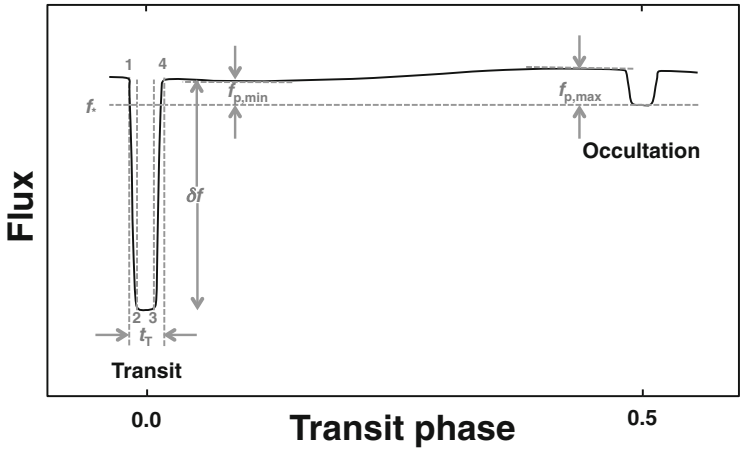

Fig. 2.2 Anatomy of the thermal-infrared light curve of a planet in a circular orbit. The first, second, third and fourth contact points of the transit are labelled; the total transit duration t_T is the duration from first to fourth contact. Outside transit and secondary occultation, the flux from the planet varies quasi-sinusoidally between $f_{p,\mathrm{max}}$ when the hottest part of the dayside hemisphere faces the observer, and $f_{p,\mathrm{min}}$ for the coolest part of the nightside. The flux level f_* of the star alone is seen only during occultation, when the planet is behind the star

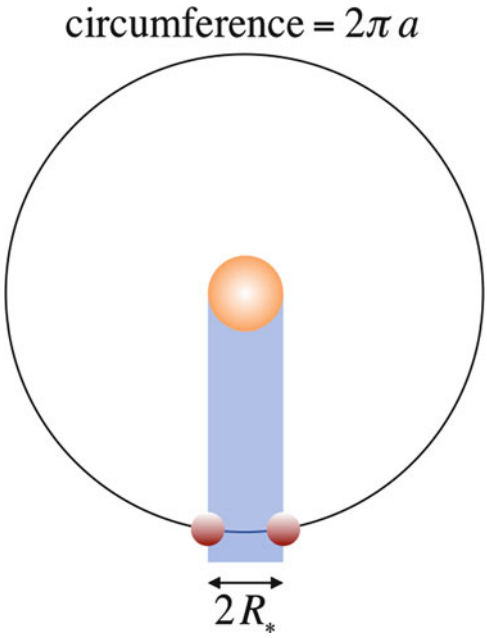

Fig. 2.3 Schematic showing that the ratio of transit duration to orbital period is proportional to the ratio of the stellar radius to the orbital separation. The planet is shown at the mid-ingress and mid-egress positions where its centre lies on the stellar limb, as seen by an observer in the orbital plane

where the direction cosine $\mu = \cos\theta$ between the line of sight and the stellar surface normal at the planet's position is related to the impact parameter b by $\cos\theta = \sqrt{1-b^2}$. If $R_\mathrm{p} \ll R_*$ and the planet is completely dark, the ratio of the flux blocked by the planet to the total flux from the visible stellar hemisphere is

$$\frac{\Delta f}{f} = \frac{\pi R_\mathrm{p}^2 I_0 (1-u+u\cos\theta)}{2\pi R_*^2 I_0 \int_0^{\pi/2} (1-u+u\cos\theta)\sin\theta\cos\theta d\theta} = \frac{3(1-u+u\sqrt{1-b^2})}{3-u}\left(\frac{R_\mathrm{p}}{R_*}\right)^2 . \tag{2.7}$$

More realistic limb-darkening laws are discussed in Sect. 2.4.

For the simplified case of a planet in a circular orbit with inclination $i = 90°$, the transit duration T from mid-ingress to mid-egress is related to the orbital period P by simple geometry:

$$\frac{T}{P} = \frac{1}{\pi}\sin^{-1}\frac{R_*}{a} \tag{2.8}$$

where a is the orbital semi-major axis. Using Kepler's third law to substitute for a (Fig. 2.3),

$$\frac{T}{P} = \frac{1}{\pi}\sin^{-1} R_* \left(\frac{4\pi^2}{GM_* P^2}\right)^{1/3} . \tag{2.9}$$

In the more general case where the inclination is less than 90°, the transit duration is reduced. Seager and Mallén-Ornelas (2003) generalise the time t_T from first to last contact as a function of inclination:

$$\frac{t_T}{P} = \frac{1}{\pi} \sin^{-1} \left(\frac{R_*}{a} \left\{ \frac{[1 + (R_\mathrm{p}/R_*)]^2 - [(a/R_*) \cos i]^2}{1 - \cos^2 i} \right\}^{1/2} \right). \tag{2.10}$$

For $\cos i \ll 1$, and noting that $b = (a/R_*) \cos i$ for a circular orbit, this becomes

$$\frac{t_T}{P} = \frac{R_*}{\pi a} \sqrt{\left(1 + \frac{R_\mathrm{p}}{R_*}\right)^2 - b^2}. \tag{2.11}$$

2.1.4 Model-Independent System Parameters

For the case where $R_* \ll a$, this leads to a model-independent relationship between the transit duration, the orbital period and the stellar bulk density ρ_*,

$$T \simeq 3\mathrm{h} \left(\frac{P}{4\mathrm{d}}\right)^{1/3} \left(\frac{\rho_*}{\rho_\odot}\right)^{-1/3}. \tag{2.12}$$

If an orbital radial-velocity solution is available, it is also possible to measure the planetary surface gravity g_p using the radial acceleration of the star at conjunction and the transit duration. Again assuming a circular orbit with $i = 90°$, the stellar orbital acceleration at conjunction is

$$\frac{dv_\mathrm{r}}{dt} = \frac{2\pi K}{P} = \frac{GM_\mathrm{p}}{a^2} = g_\mathrm{p} \frac{R_\mathrm{p}^2}{a^2} = g_\mathrm{p} \frac{R_\mathrm{p}^2}{R_*^2} \frac{R_*^2}{a^2}, \tag{2.13}$$

where K is the amplitude of the star's orbital reflex motion about its centre of gravity with the planet. Southworth et al. (2007) pointed out that since $(R_\mathrm{p}/R_*)^2$ is related to the transit depth by Eq. (2.7) and (R_*/a) is related to the transit duration by Eq. (2.8), this provides a model-independent measure of the planet's surface gravitational acceleration,

$$g_\mathrm{p} = \frac{2\pi K}{P} \left(\frac{R_*}{R_\mathrm{p}}\right)^2 \left(\frac{a}{R_*}\right)^2. \tag{2.14}$$

To determine the planetary bulk density ρ_p requires a precise estimate of the stellar radius, as well as knowledge of the transit depth:

$$\rho_\mathrm{p} = \frac{3g_\mathrm{p}}{4\pi G R_\mathrm{p}} = \frac{3g_\mathrm{p}}{4\pi G R_*} \left(\frac{R_*}{R_\mathrm{p}}\right). \tag{2.15}$$

If precise measures of the host star's angular diameter θ and parallax $\hat{\pi}$ (or distance d) are available, the planetary bulk density can also be derived independently of stellar models, since $R_* = \theta d = \theta/\hat{\pi}$:

$$\rho_\mathrm{p} = \frac{3 g_\mathrm{p} \hat{\pi}}{4\pi G \theta} \left(\frac{R_*}{R_\mathrm{p}} \right). \tag{2.16}$$

2.2 Transit Surveys

By the start of the twenty-first century, radial-velocity surveys of F, G, K and M stars had established that the occurrence rate of gas-giant planets orbiting within 0.1 AU of their host stars was about 1 % (Marcy et al. 2005). Such planets orbit close enough to their host stars to have transit probabilities of order 2–10 %. The transits of such planets have durations of order 3 h, short enough for complete events to be detected in the course of a single night from the ground. Detection of more distant planets is more difficult. At 1 AU from a star of 1 solar mass, the transit duration is 13 h and the probability of transits occurring is greatly reduced.

Despite being intrinsically rare, large close-orbiting planets have a sufficiently high transit probability that roughly one star in every thousand should host a transiting hot Jupiter. To achieve a yield of order 1000 hot Jupiters at a detection efficiency of 100 % requires at least one million stars to be monitored with a photometric precision better than 1 %.

2.2.1 Ground-Based Surveys

Transit searches became a high-priority goal for the optical gravitational lensing experiment survey (OGLE, Udalski et al. 1992), which was already monitoring millions of stars in the galactic bulge region using the 1.3-m Warsaw telescope at Las Campanas, Chile. Many dozens of transit candidates with V magnitudes between 15 and 16 were published by Udalski et al. (2002a,b,c), but efforts to establish their planetary nature via radial-velocity follow-up proved challenging even with the UVES instrument on the VLT (e.g. Bouchy et al. 2004, 2005) owing to the faintness of the host stars. A shallower, wider-field approach was needed to produce brighter candidates for which radial-velocity follow-up could be conducted efficiently on smaller telescopes.

The brightest known stars hosting transiting hot Jupiters are HD 209458 (magnitude $V = 7.6$) and HD 189733 ($V = 7.7$). Using the fact that a 5-magnitude increase in limiting magnitude accesses a volume of space 1000 times greater, and assuming these to be both representative and the only examples of their kind brighter than $V = 8.0$, simple extrapolation suggests that there ought to be at least 2000 similar objects brighter than $V = 13.0$. This is encouraging, because 1.2 m-class telescopes

with high-precision radial-velocity spectrometers such as CORALIE on the 1.2-m Euler telescope at La Silla and SOPHIE in the 1.9-m telescope at Haute-Provence are capable of performing the essential radial-velocity follow-up at magnitudes brighter than $V = 12.0$ or so. The whole sky comprises 41,253 square degrees, so there should be at least one transiting hot Jupiter brighter than $V = 12.0$ per 82 square degrees of sky.

The image scale of a telescope of focal length f is

$$3600 \times \frac{180}{\pi} \frac{1}{f} = \frac{206265}{f} \text{ arcsec/mm}, \tag{2.17}$$

if f is expressed in mm. To image 82 square degrees of sky on to an affordable science-grade charge-coupled device (CCD) of 2048^2 13.5 μm pixels requires a telescope with a focal length of 174 mm.

These considerations inspired several groups to embark on ground-based surveys employing commercial 200-mm camera lenses on robotic mounts backed by science-grade CCDs. The image scale of these 200-mm lenses is

$$3600 \times \frac{180}{\pi} \frac{1}{f} = 1031 \text{ arcsec/mm}, \tag{2.18}$$

yielding a pixel scale

$$0.0135 \times 1031 = 13.9 \text{ arcsec/pixel} \tag{2.19}$$

and hence a field of view whose angular extent is

$$\frac{2048 \times 13.9}{3600} = 7.9 \text{ degrees}, \tag{2.20}$$

with a solid angle of 62 square degrees. The expected planet catch is therefore of order one hot Jupiter brighter than $V = 12.0$ per camera. The most successful among these surveys have been the transatlantic exoplanet survey (TrES, Alonso et al. 2004); the wide-angle search for planets (WASP, Pollacco et al. 2006), the Hungarian automated telescope network (HATNet, Bakos et al. 2004) and the XO survey (McCullough et al. 2005). More recently, HATSouth (Bakos et al. 2013), the Qatar exoplanet survey (QES, Alsubai et al. 2013) and the kilodegree extremely little telescope (KELT, Pepper et al. 2007) have entered service and started publishing new discoveries of transiting planets. Together they have surveyed about 80 % of the sky (Fig. 2.4), and published over 180 confirmed discoveries of transiting gas-giant and ice-giant planets brighter than $V = 13.0$ with periods less than 10 days. The mass–radius diagram for these ground-based discoveries is shown in Fig. 2.5.

The next-generation transit survey (NGTS, Wheatley et al. 2014) builds on the experience of the WASP survey, with the goal of detecting transits 0.001 magnitude

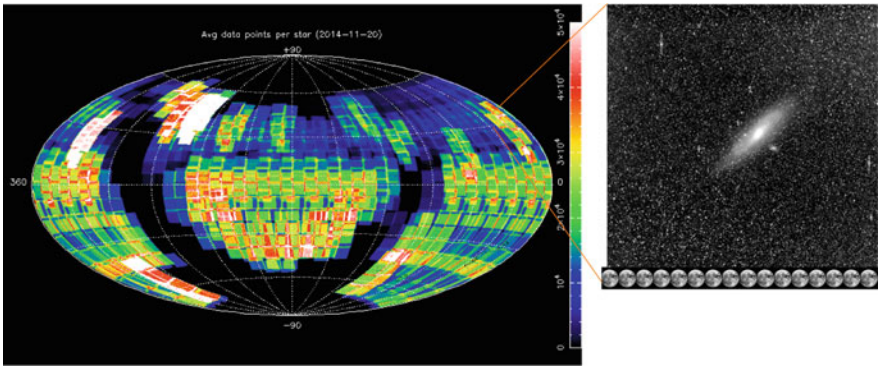

Fig. 2.4 Sky coverage of the WASP survey. The *colour scale* denotes the number of exposures in the archive for a given field, from *black* (0) to *white* (50,000 images). The average field has been observed 25,000 times over 2 or 3 seasons. The 8-camera "footprint" of the instrument is apparent. The zoomed image at right illustrates the 7.8-degree square field of view of a single camera

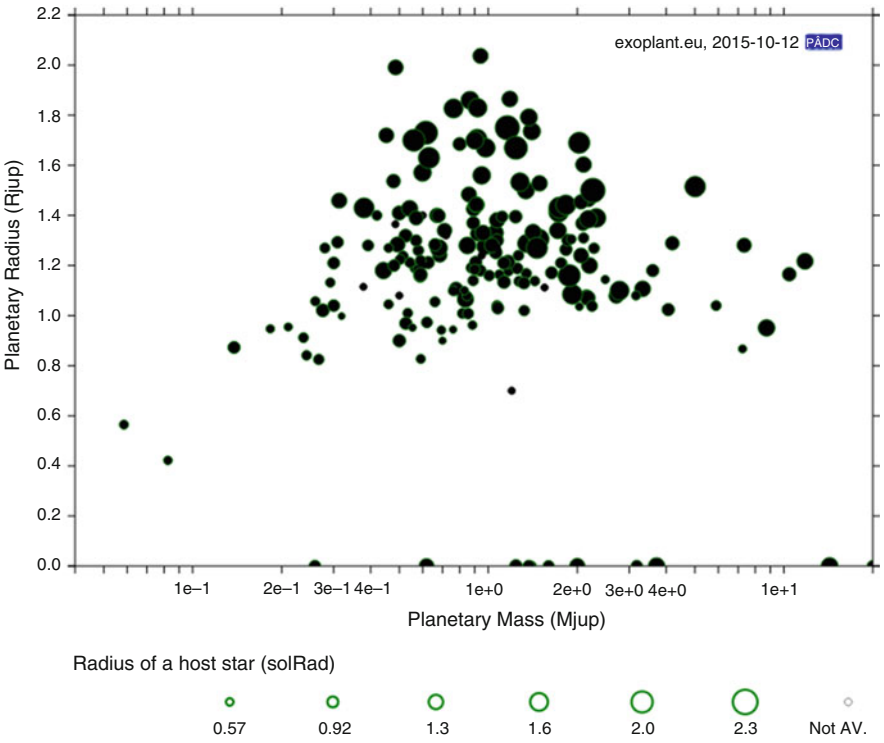

Fig. 2.5 Planetary mass–radius diagram for ground-based discoveries from the TrES, WASP, HATNet, XO, HATSouth, QES and KELT surveys. *Symbol size* denotes host star radius, illustrating that inflated planets tend to orbit large stars, and that small planets are more easily detected around smaller host stars

deep from the ground. It comprises 12 small robotic telescopes located at Paranal, Chile. As discussed in Sect. 2.3.4, colour-dependent flat-fielding errors dominate the WASP error budget at bright magnitudes. Experiments with the prototype NGTS instrument have verified that the required noise levels can be achieved with sub-pixel guiding precision. The 600–900 nm passband of NGTS is designed to probe the short-period population of large super-Earths and mini-Neptunes orbiting K and early M dwarfs. The resulting discoveries will be significantly brighter than their *Kepler* counterparts, making them viable targets for radial-velocity mass determination, and ultimately for transmission and occultation spectroscopy studies with the *James Webb Space Telescope* (JWST, Gardner et al. 2006).

2.2.2 Space-Based Surveys

Ground-based surveys are limited in their ability to detect small planets and planets in long-period orbits. Atmospheric transparency fluctuations and scintillation limit the relative photometric precision attainable from the ground to 0.1 % or so from the very best sites, while the Earth's rotation precludes reliable detection of transits of more than 5 or 6 h duration.

Earth-sized planets have radii that are an order of magnitude smaller than Jupiters, with transit depths of order:

$$\left(\frac{R_\oplus}{R_\odot}\right)^2 \simeq 10^{-5}. \tag{2.21}$$

Space-based CCD photometry with sub-pixel pointing precision can attain a relative precision of order 10^{-5} with rigorous ensemble photometry and decorrelation (see Sect. 2.3). The long stare times needed to detect planets in long-period orbits can be achieved either by placing the spacecraft in independent orbit around the Sun or (more cheaply) in a near-polar Sun-synchronous low-Earth orbit, pointing away from the Sun.

The *CoRoT* mission, which remained operational for 6 years from launch in December 2006–November 2012, comprised a 27-cm telescope with a field of view of 2.7×3.05 degrees in a 900-km polar orbit (Auvergne et al. 2009). The orbit plane drifted only very slowly in right ascension, allowing a series of 5-month campaigns to be conducted in fields located close to the two intersections of the galactic plane with the ecliptic. Among the 28 planets published to date from the *CoRoT* mission, CoRoT-7b (Queloz et al. 2009) is distinguished as being the first example of a "super-Earth", with a radius of 1.7 $R_\oplus$ and an Earth-like density suggesting a predominantly iron-silicate composition (Haywood et al. 2014).

NASA's *Kepler* mission (Borucki et al. 2010) is a Schmidt telescope with an effective aperture of 0.95 m, whose focal plane is tiled with 42 CCDs giving a field of view of 115 square degrees at an image scale of 4 arcsec per pixel. *Kepler*

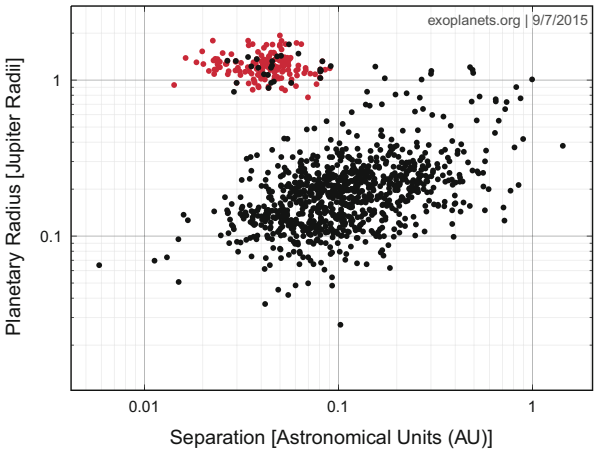

Fig. 2.6 Planetary radius-separation diagram for hot Jupiters found in ground-based surveys (*red*) and planet candidates with measured radii from the *Kepler* archive (*black*). In both cases, the *lower right-hand* boundary represents the transit-detection threshold. Hot Jupiters are intrinsically rare, even though they dominate ground-based surveys

was launched in 2009 May into an Earth-trailing heliocentric orbit. The baseline mission was to observe a single field, roughly midway between the bright stars Deneb and Vega, for 3.5 years. The focal plane detector configuration has fourfold symmetry, to allow for a 90° rotation around the boresight every 3 months, to maintain illumination of the solar panels.

Figure 2.6 shows clearly that the transit candidates with measured radii in the archive from *Kepler's* baseline mission occupy a very different part of mass-separation space to that occupied by the hot Jupiters from the ground-based surveys. Hot Jupiters are easily detected, so the *Kepler* points give a clearer impression of their paucity in comparison to the ice-giant and super-Earth planets that dominate the *Kepler* population.

Shortly after the baseline mission was extended in 2013, the *Kepler* spacecraft suffered a failure of its third remaining gyroscope. A new mode of operation, the K2 mission, is currently in progress (Howell et al. 2014). In this mode, *Kepler* points to targets near the ecliptic plane, slowing the rate of drift caused by unbalanced radiation pressure on its solar panels. A different field is selected each quarter. Although the pointing is less stable, careful decorrelation restores photometric precision roughly a factor 2 poorer than was achieved in the original mission. In addition to transit hunting, K2 permits a wider range of astrophysical investigations to be conducted than was possible with the original mission.

2.3 Ensemble Photometry and Transit Detection

Wide-field imaging photometry at optical wavelengths is performed using CCD detectors. In this section, I describe the key steps in the data processing for a ground-based wide-field photometry survey. The procedures described here follow closely those used in the WASP project, as described by Collier Cameron et al. (2006). The instrumentation, observing principles and data-reduction strategies of the HAT and WASP projects are described by Bakos et al. (2004, 2013) and Pollacco et al. (2006), respectively.

2.3.1 Image Preprocessing

Successful detection of transits with depths of 1 % or less requires meticulous care at every stage in the image processing (Fig. 2.7). At the start of each night, the cameras record sequences of zero-exposure frames, to map any spatial variation in the bias signal applied to the amplifier while the image is being read out. Dark frames, taken with the shutter closed and exposure times equal to the science frames, map the thermal noise pattern of the CCD. Once bias-subtracted, they may be subtracted from the science frames.

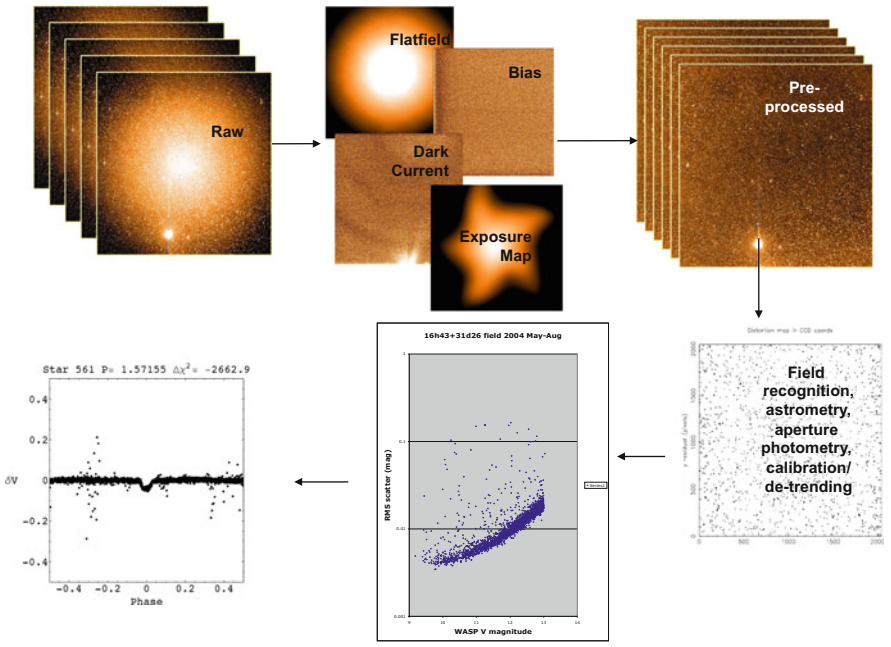

Fig. 2.7 Flow diagram illustrating the main steps in the WASP data-reduction pipeline

The relative sensitivities of individual pixels in the image are mapped using a sequence of "flat-field" exposures of the sky near the zenith, taken in dusk and dawn twilight. Exposure times are adjusted for each exposure according to a model of the sky brightness as a function of solar zenith angle at mid-exposure. Between exposures, the camera mount is moved slightly, to ensure that stellar images do not fall on the same pixels.

The resulting stack of images contains information about the brightness gradient of the twilight sky, the vignetting pattern of the optical system and the variation in exposure over the frame resulting from the finite travel time of the mechanical CCD shutter. The shutter travel time is measured from the linear dependence of the signal (relative to the image median) in a given pixel, as a function of inverse exposure time. The signal in each pixel is then scaled to give a uniform exposure time across the image, for both flat field and science frames. The large-scale variations in the individual flat fields are decomposed into a linear sky illumination gradient and a centro-symmetric vignetting pattern. The gradient is removed and the flat fields are normalised to their median signal values. Finally, the median of the individual flat-field images is computed, to eliminate any stellar images.

In practice, clouds or technical problems prevent acquisition of the flat-field sequences on some nights. Flat fields evolve with time. In the WASP and HAT cameras, for example, dust particles on the optical surfaces come and go, producing shadow discs with finite lifetimes. WASP employed a weighted average of historical flat-field images, using a weighting function which decayed exponentially with a time constant of 2 weeks.

The WASP data-reduction pipeline employs catalogue-driven aperture photometry. For each frame, an automated object-detection routine determines the positions of all point sources on the frame. The catalogue of object positions is cross-matched against the TYCHO-2 catalogue (Høg et al. 2000), and establishes an astrometric frame solution with an RMS precision of 0.1–0.2 pixel. The pipeline carries out aperture photometry at the positions of all objects brighter than magnitude 15.0 in the red bandpass of the USNO-B1.0 catalogue. The aperture photometry is carried out in three concentric circular apertures with radii of 2.5, 3.5 and 4.5 pixels. The flux ratios between pairs of apertures contain information about image morphology. They are used to detect and flag blended pairs of images. The relationship between raw instrumental magnitude and TYCHO-2 V magnitude is established via a colour-dependent transformation. Further details of the image preprocessing are given in Sect. 4 of Pollacco et al. (2006).

2.3.2 Decorrelation

At this stage, the raw instrumental magnitudes are subject to various sources of systematic error. Foremost among these is atmospheric extinction. The attenuation of light from each object depends on the wavelength dependence of scatterers in the Earth's atmosphere and the path length through the atmosphere. The airmass X is

the path length scaled to that at the zenith, and varies with zenith angle z as $\sec z$. The intensity along the incoming beam decays exponentially with X. The observed magnitude $m_{\rm obs}$ is related to the magnitude m_0 observed above the atmosphere by

$$m_{\rm obs} = m_0 + kX. \tag{2.22}$$

The extinction coefficient k is itself dependent on the colour of the star. This arises because k increases with decreasing wavelength, due to Rayleigh and other scattering processes. If a wide bandpass is used, intrinsically red stars will suffer less extinction than bluer ones. Expressed in terms of a star's colour index c, the extinction equation becomes

$$m_{\rm obs} = m_0 + k'X + k''cX, \tag{2.23}$$

where k' and k'' are referred to as the primary and secondary extinction coefficients.

Correcting for secondary extinction is problematic if, as is generally the case, the colour of the star is not known in advance. To make matters worse, the extinction coefficient varies slowly across the sky at many sites, and can change with time-dependent factors such as local wind speed, direction and humidity. To combat this, Tamuz et al. (2005) devised a generalised iterative scheme, known as SysRem, for the correction of extinction and other systematics that vary smoothly across the field of view as a function of time.

SysRem operates on a rectangular array of light curves of N objects, each of which is observed at M different epochs. Before applying theSysRem algorithm, the WASP pipeline carries out a maximum-likelihood coarse decorrelation to determine the mean magnitude and variance of each of the N stars, and the zero-point correction and its variance for each of the M images. The mean magnitude of each star is

$$\hat{m}_i = \frac{\sum_j m_{i,j} w_{i,j}}{\sum_j w_{i,j}}, \tag{2.24}$$

where the weight $w_{i,j} = 1/(\sigma_{i,j}^2 + \sigma_{t(j)}^2)$ includes both the estimated variance of the observed magnitude and an additional intra-frame variance $\sigma_{t(j)}^2$ which serves to down-weight images degraded, for example, by drifting cloud.

Similarly, the zero-point correction for each frame is

$$\hat{z}_j = \frac{\sum_i m_{i,j} u_{i,j}}{\sum_i u_{i,j}}, \tag{2.25}$$

where the weight $u_{i,j} = 1/(\sigma_{i,j}^2 + \sigma_{s(i)}^2)$ includes both the estimated variance of the observed magnitude and an additional per-star variance $\sigma_{s(i)}^2$ which serves to down-weight stars with high intrinsic variability.

To solve for $\sigma^2_{s(i)}$, consider the likelihood of obtaining a data vector $\boldsymbol{D} = \{m_{i,j}, j = 1 \ldots M\}$ for star i conditioned on a model $\boldsymbol{\mu} = \{\hat{m}_i + \hat{z}_j, j = 1 \ldots M\}$ assuming Gaussian noise:

$$L(\boldsymbol{D}|\boldsymbol{\mu}) = (2\pi)^{-M/2} \prod_j \left[\frac{1}{\sigma^2_{i,j} + \sigma^2_{t(j)} + \sigma^2_{s(i)}} \right] \times \exp\left\{ -\frac{1}{2}\chi^2_i \right\} \tag{2.26}$$

where

$$\chi^2_i = \sum_j \frac{(m_{i,j} - \hat{m}_i - \hat{z}_j)^2}{\sigma^2_{i,j} + \sigma^2_{t(j)} + \sigma^2_{s(i)}}. \tag{2.27}$$

To obtain the maximum-likelihood solution, we first solve iteratively for $\sigma^2_{s(i)}$ holding $\sigma^2_{t(j)}$ constant:

$$\sum_j \frac{1}{\sigma^2_{i,j} + \sigma^2_{t(j)} + \sigma^2_{s(i)}} - \sum_j \frac{(m_{i,j} - \hat{m}_i - \hat{z}_j)^2}{\left[\sigma^2_{i,j} + \sigma^2_{t(j)} + \sigma^2_{s(i)}\right]^2} = 0. \tag{2.28}$$

Analogously, we solve iteratively for $\sigma^2_{t(j)}$ holding $\sigma^2_{s(i)}$ constant:

$$\sum_i \frac{1}{\sigma^2_{i,j} + \sigma^2_{t(j)} + \sigma^2_{s(i)}} - \sum_i \frac{(m_{i,j} - \hat{m}_i - \hat{z}_j)^2}{\left[\sigma^2_{i,j} + \sigma^2_{t(j)} + \sigma^2_{s(i)}\right]^2} = 0. \tag{2.29}$$

The entire system of equations (2.24), (2.25), (2.28) and (2.29) is then iterated to convergence to give $\hat{m}_i$, $\hat{z}_j$, $\sigma^2_{s(i)}$ and $\sigma^2_{t(j)}$.

We start by subtracting the inverse variance-weighted average magnitude of each star and the zero point of each frame to obtain a residual array

$$r_{i,j} = m_{i,j} - \hat{m}_i - \hat{z}_j. \tag{2.30}$$

SysRem is applied to this residual array. The goal is to minimise the misfit S^2 between the residuals and a model consisting of the product of the extinction coefficient c_i for each star and the airmass a_i at the time of observation. The misfit statistic for star i is

$$S^2_i = \sum_j (r_{i,j} - a_i c_j)^2 w_{i,j}, \text{ where } w_{i,j} = \frac{1}{\sigma^2_{i,j} + \sigma^2_{t(j)} + \sigma^2_{s(i)}}. \tag{2.31}$$

Since the average airmass a_j of each frame is known, we can now determine an effective "extinction coefficient" for each star by optimal scaling:

$$c_i = \frac{\sum_j r_{i,j} a_j w_{i,j}}{\sum_j a_j^2 w_{i,j}}. \tag{2.32}$$

Similarly, the problem can be re-cast to determine the optimal effective airmass a_j for each image, given the effective extinction coefficients c_i for the individual stars:

$$a_j = \frac{\sum_i r_{i,j} c_i w_{i,j}}{\sum_i c_i^2 w_{i,j}}. \tag{2.33}$$

Once again, the system is iterated to convergence to obtain optimal sets of c_i and a_j. Tamuz et al. (2005) find that the algorithm converges to the same values irrespective of the values used for the initial airmasses. The final c_i and a_j are not therefore necessarily related to stellar extinction coefficients and airmasses. For example, steadily declining temperature during the night can induce a drift in focus which affects stars in different parts of the field to a greater or lesser extent. The same approach can be used to model further systematics with this kind of linear behaviour in the residuals from the initial calculation:

$${}^{(1)}r_{i,j} = r_{i,j} - {}^{(1)}c_i^{(1)} a_j, \tag{2.34}$$

by minimising

$${}^{(1)}S_i^2 = \sum_{i,j} ({}^{(1)}r_{i,j} - {}^{(2)}c_i^{(2)} a_j)^2 w_{i,j} \tag{2.35}$$

The procedure is very similar to principal-component analysis (PCA), and indeed reduces to PCA in the case where all the data weights are identical. The WASP pipeline computes and corrects for the four most significant sets of basis vectors ${}^{(k)}a_j$ and ${}^{(k)}c_i$. This is sufficient to remove the most significant sources of instrumental and environmental systematic error that are common to all stars in the field, while preserving genuine astrophysical variability (Fig. 2.8). Other decorrelation methods to which the reader is refereed include the trend filtering analysis (TFA, Kovács et al. 2005), which is used in both HAT and WASP transit searches. External parameter decorrelation (EPD, Bakos et al. 2010) is a parametric decorrelation method using basis functions constructed from factors known or suspected to cause systematic error such as pixel coordinates and sub-pixel phase, sky background level, point-spread function shape parameters, hour angle and zenith distance. EPD and SysRem are generally applied before TFA. A more rigorous Bayesian approach analogous to SysRem and TFA (Smith et al. 2012) is used to generate the cotrending basis vectors for the maximum a posteriori pre-search conditioning (PDC-MAP) data products from the *Kepler* mission. More recently, Gibson (2014)

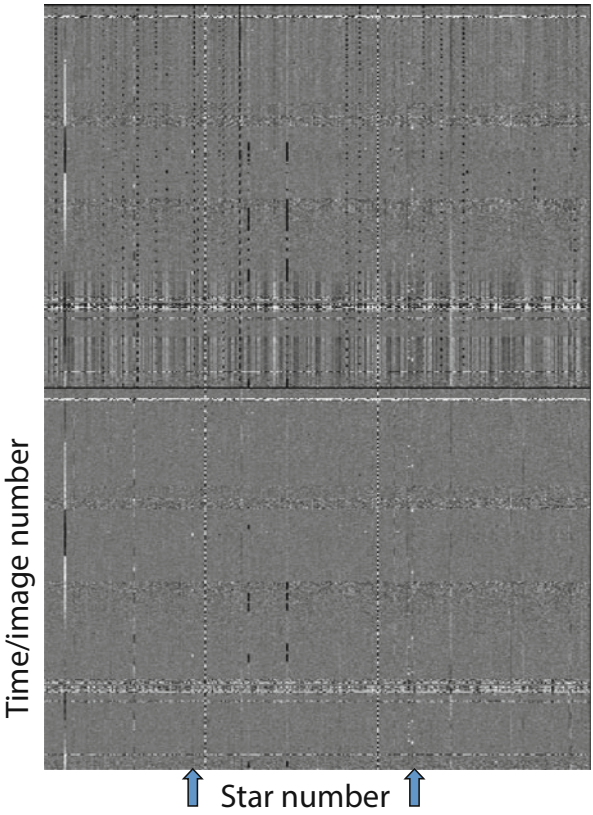

Fig. 2.8 Application of SysRem to the ensemble of light curves from WASP field 16h30+28. The *upper panel* shows the residual array in *greyscale* form following subtraction of the mean magnitude of each star and the zero point of each frame. Each column is the light curve of one star; each row holds data from a single image. The *lower panel* shows the same data after removal of four SysRem basis functions. The *arrows* denote two objects exhibiting occasional transit-like events

and Aigrain et al. (2015) have developed non-parametric equivalents of EPD employing Gaussian-process regression to model the systematics in HST/NICMOS, Spitzer and K2 photometry.

2.3.3 Transit Detection

The most widely used transit-detection method is the box least-squares (BLS) method of Kovács et al. (2002). Following decorrelation, we compute and subtract the optimal average magnitude of the observations $\hat{x}_j$ of a given star as

$$x_j = \bar{x}_j - \frac{\sum_j \bar{x}_j w_j}{\sum_j w_j} \tag{2.36}$$

where w_j includes the stellar-variability and spatial-transparency variances as in Eq. (2.31).

We also define the global summations

$$t = \sum_j w_j \text{ and } \chi_0^2 = \sum_j x_j^2 w_j, \tag{2.37}$$

assuming the noise to be uncorrelated.

The transit search is conducted on a frequency grid. The frequency spacing must satisfy the requirement that the phase of each observation must change by less than the transit duration between adjacent frequencies. At each frequency the data are sorted in phase and partitioned into blocks of in-transit (ℓ, low) and out-of-transit (h, high) points. Different transit phases are explored by repartitioning with the low block at a succession of locations along the phase-sorted dataset, and summing within the low block:

$$s = \sum_{i\in\ell} x_i w_i, \quad r = \sum_{i\in\ell} w_i, \quad q = \sum_{i\in\ell} x_i^2 w_i. \tag{2.38}$$

The mean light level in transit (L) and its variance are given by

$$L = \frac{s}{r}, \quad \text{Var}(L) = \frac{1}{r}, \tag{2.39}$$

while outside transit

$$H = \frac{-s}{t-r}, \quad \text{Var}(H) = \frac{1}{t-r}. \tag{2.40}$$

The fitted transit depth and its associated variance are

$$\delta = L - H = \frac{st}{r(t-r)}, \quad \text{Var}(\delta) = \frac{t}{r(t-r)}, \tag{2.41}$$

yielding the signal-to-noise ratio of the transit depth

$$\text{S/N} = s\sqrt{\frac{t}{r(t-r)}}. \tag{2.42}$$

The improvement in the fit to the data when compared with that of a constant light-curve model is, assuming uncorrelated noise,

$$\Delta\chi^2 = \frac{s^2 t}{r(t-r)}. \tag{2.43}$$

The goodness of fit outside transit is then,

$$\chi_h^2 = \chi_0^2 - \frac{s^2}{(t-r)} - q. \tag{2.44}$$

At each frequency, the values of the transit depth, $\Delta\chi^2$ and χ^2_h, are stored for the transit phase that yields the best fit.

When transits are clearly detected, the resulting periodogram of $\Delta\chi^2$ usually shows a clear minimum at the orbital period. Harmonics of the orbital period are usually present at multiples and submultiples of the true period. For observations made from a single site, their relative strengths depend on the window function arising from the visibility of transits during the day/night cycle (Fig. 2.9). In practice, however, simple algorithms such as BLS are susceptible to false alarms, and human inspection of the light curves phase-folded at the dominant frequency is essential to verify whether or not genuine transits are present.

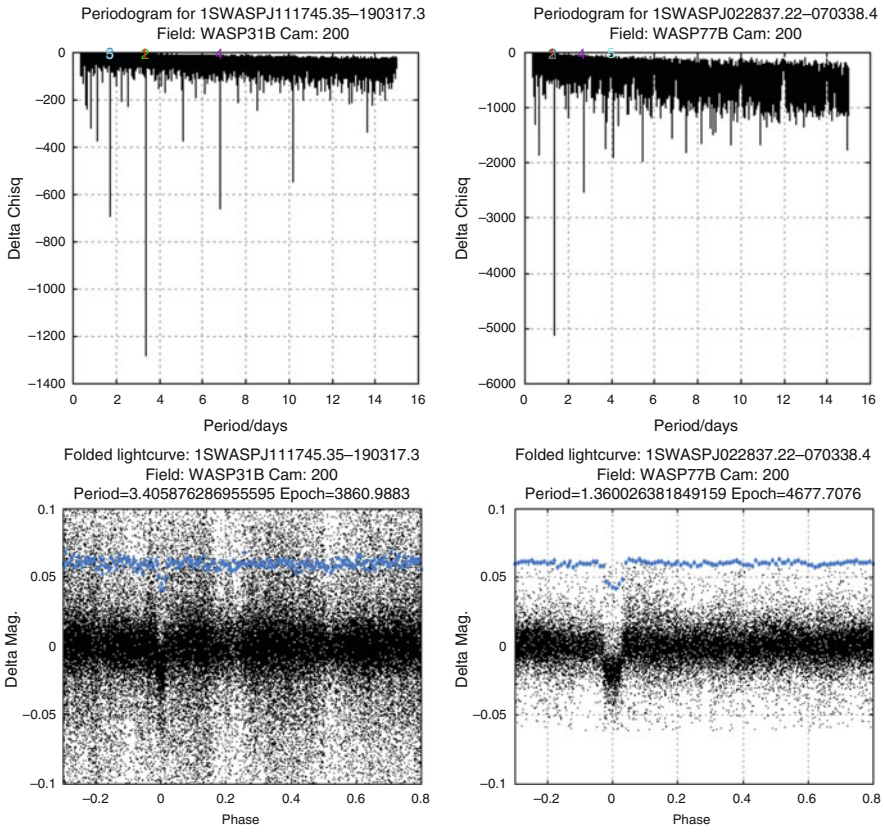

Fig. 2.9 BLS periodograms (*upper*) and folded transit light curves (*lower*) for two WASP planets: WASP-31b ($P = 3.40$ days, *left*) and WASP-77b ($P = 1.36$ days, *right*). Note spacing of harmonics at multiples of P and $P/2$, indicating that multiple transits are clearly detected

2.3.4 *Correlated Noise*

The treatment in Sect. 2.3.3 above yields the improvement in χ^2 relative to a constant-flux model with no transits. If the residual noise after decorrelation were Gaussian and independent, Bayesian model-comparison tools could be used to determine whether transits are present. In practice, however, correlated noise remains. At the instrumental level, imperfect flat fielding arising from transient dust particles on the optics and the colour dependence of the CCD pixel sensitivity pattern introduces complex position-dependent flux variations on a variety of length scales. At the astrophysical level, stellar p-modes, granulation and magnetic activity introduce correlated brightness fluctuations on a wide variety of timescales. A realistic assessment of the likelihood function ideally requires an understanding of the structure of the full covariance matrix.

Methods such as Gaussian-process regression provide powerful tools for modelling correlated noise sources in a non-parametric way. Unfortunately they are computationally inefficient for datasets of more than a few hundred data points. Pont et al. (2006) developed a simple approach for assessing the contributions of correlated noise on different timescales. The data are boxcar-smoothed on a succession of box lengths L. The empirical power-law dependence of the RMS scatter in the binned light curve on L takes the form of a power law:

$$\sigma_{\rm binned} = \sigma_{\rm unbinned} L^b. \tag{2.45}$$

If the noise is uncorrelated, we expect $b = -1/2$; completely correlated noise arising from variability on timescales intermediate between the longest smoothing length considered and the length of the data train gives $b = 0$.

Carter and Winn (2009) pointed out that when correlated noise is present whose power spectral density varies with frequency f as an inverse power law $1/f^\gamma$, the covariance matrix of the noise process is nearly diagonal when the data are transformed into a wavelet basis. For more or less evenly sampled data, this gives a fast and efficient method for calculating relative likelihoods and obtaining reliable estimates of parameter uncertainties.

2.4 Transit Parameter Fitting

As discussed in Sect. 2.1.3, the fraction of the host star's light blocked during a transit depends on the planet-to-star area ratio, the inclination of the orbit to the line of sight and the form of the stellar photospheric limb-darkening profile. The duration of ingress and egress also depends on the relative radii of the planet and star, and on the inclination of the orbit to the line of sight.

The two key parameters that determine the flux deficit at any given moment during the transit are the projected separation $z \equiv d/R_*$ of the centres of the planet

and the star, and the planet/star radius ratio $p \equiv R_p/R_*$. The apparent separation of the star and planet projected on the plane of the sky is $d = r \sin\alpha$, where r is the instantaneous distance of the planet from the star and α is the star–planet–observer phase angle. Calculation of the flux deficit involves integrating the surface brightness $I(\mu)$ of the photosphere over the solid angle of the part of the photosphere obscured by the planet. In the limit where the planet is much smaller than the star, a fast approximation involves using the photospheric brightness at the centre of the planet. Either approach requires a limb-darkening model, which is described by one or more auxiliary parameters u_n whose values depend on the temperature, pressure and opacity profile of the photosphere. A number of authors have published fast analytic algorithms for computing the flux deficit for a variety of limb-darkening models. The simplest of these is the very simplistic linear model

$$\frac{I(\mu)}{I_0} = 1 - u(1-\mu), \tag{2.46}$$

where the parameters I_0, u_n and μ are defined as in Eq. (2.6). More sophisticated treatments range from the quadratic approximation

$$\frac{I(\mu)}{I_0} = 1 - \sum_{n=1}^{2} u_n(1-\mu^n) \tag{2.47}$$

to a 4-coefficient nonlinear model which reproduces the intensity profile near the limb in a much more satisfactory manner:

$$\frac{I(\mu)}{I_0} = 1 - \sum_{n=1}^{4} u_n(1-\mu^{n/2}). \tag{2.48}$$

The most widely used formulation is that of Mandel and Agol (2002), though more recent treatments by Pál (2008) and Giménez (2006) offer improved numerical stability and precision in some circumstances. Numerous compilations of theoretical limb-darkening coefficients are available in a range of common photometric bandpasses, computed from model atmospheres on grids of stellar effective temperature, surface gravity and metallicity, and fitted with the linear, quadratic and 4-coefficient nonlinear approximations. The most widely used compilations at present are those of Claret (2003, 2004) and Sing (2010).

Given a set of photometric measurements and suitable implementation of a transit model in subroutine form, the procedure for computing a single realisation of a transit model involves adopting a set of limb-darkening coefficients for the relevant passband and a value for the planet/star radius ratio p, and computing the projected separation of centres $z(t)$ at each of the times t to be considered. The output is an array of flux ratios $F(t)/F_0$ where F_0 is the stellar flux outside transit. For parameter-fitting purposes, the model is evaluated at the times of observation, though denser sampling is often used for graphical presentation of the light-curve model.

Denser time sampling is also necessary when the exposure time is long enough that the relative flux changes significantly during the exposure. This is an important issue with, for example, the long-cadence observing modes of the *CoRoT* and *Kepler* missions (Kipping 2010). Satisfactory modelling of the recorded flux in these circumstances requires integration of the instantaneous model flux, necessitating denser sampling over the exposure duration.

2.4.1 Orbital Elements

A dynamical model of the planet's orbit is needed to compute the separation of centres $z(t)$ at a sequence of times t. In addition to the scaling parameters $z_{\rm max} = a/R_*$ and $p = R_{\rm p}/R_*$, the orbital elements of the planet must be known. For the simplest possible system comprising only a star and a planet, the epoch of periastron t_0, orbital period P, inclination i, eccentricity e and argument of periastron ω provide a full description of the orbital motion.

As illustrated in Fig. 2.10, the true anomaly of the planet at the times of mid-transit and mid-occultation is

$$\nu_{\rm tr} = \frac{\pi}{2} - \omega \qquad \text{and} \qquad \nu_{\rm occ} = \frac{3\pi}{2} - \omega, \tag{2.49}$$

respectively.

The time delay from periastron to mid-transit for a planet in an eccentric orbit is computed from Eq. (2.49) for the true anomaly at mid-transit. Substituting $\nu_{\rm tr}$ for ν in Eq. (2.51) yields the eccentric anomaly $E_{\rm tr}$ at mid-transit. The time delay is then given by

$$t_{\rm tr} - t_0 = \frac{P}{2\pi} M_{\rm tr} = \frac{P}{2\pi}(E_{\rm tr} - e \sin E_{\rm tr}). \tag{2.50}$$

A similar procedure yields the true anomaly at mid-occultation.

For the purposes of determining the scaled separation of centres z at any time, we need to calculate the mean anomaly ν as a function of time. At any time t, the true anomaly ν is computed from the mean anomaly $M = 2\pi(t - t_0)/P$ via the eccentric anomaly

$$E = 2 \tan^{-1} \left[\sqrt{\frac{1-e}{1+e}} \tan \frac{\nu}{2} \right], \tag{2.51}$$

which is related to the mean anomaly by $M = E - e \sin E$. A simple, fast iterative solution starts by estimating $E_1 = M$, then iterating

$$E_{i+1} = M + e \sin E_i. \tag{2.52}$$

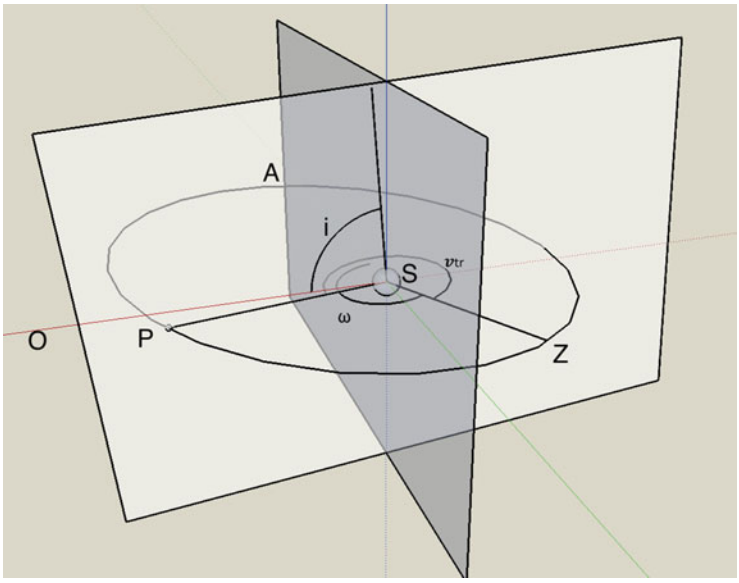

Fig. 2.10 Orbital elements of an extrasolar planet P orbiting a star S. The *red axis points* to the observer. The *green and blue axes* lie in the (*grey*) plane of the sky; the *red and blue axes*, and the orbital angular-momentum vector, lie in the *white plane*. Mid-transit occurs when the planet passes through the *white plane*. The argument of periastron ω (angle ASZ) is measured from the intersection of the planes of the orbit and the sky on the side of the orbit where the planet's motion is toward the observer O, around the star in the plane of the orbit and to the direction Z of periastron. The true anomaly ν is the angle ZSP; here, the configuration is shown at mid-transit

On convergence,

$$\nu = 2\tan^{-1}\left[\sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2}\right]. \tag{2.53}$$

The instantaneous star–planet distance is

$$r = a(1 - e\cos E). \tag{2.54}$$

Seen from the observer's viewpoint, the planet's position along the green axis in Fig. 2.10 on the plane of the sky is

$$x_{\rm p} = r\sin(\nu + \omega - \pi/2). \tag{2.55}$$

Along the projection of the orbital axis in the plane of the sky (the blue axis in Fig. 2.10),

$$z_{\rm p} = -r\cos(\nu + \omega - \pi/2)\cos i. \tag{2.56}$$

The third component of the planet's cartesian position vector is toward the observer along the red axis in Fig. 2.10:

$$y_{\rm p} = r\cos(\nu + \omega - \pi/2)\sin i. \tag{2.57}$$

The star–planet–observer phase angle, which determines the fractional illumination of the planet's visible hemisphere, is $\cos\alpha = y_{\rm p}/r$. The apparent star–planet separation is thus

$$r\sin\alpha = \sqrt{x_{\rm p}^2 + z_{\rm p}^2}, \tag{2.58}$$

and the scaled separation is

$$z = r\sin\alpha/R_*. \tag{2.59}$$

By differentiating $y_{\rm p}$ with respect to time we obtain the planet's velocity toward the observer relative to the star:

$$\begin{aligned} v_{\rm p} &= \frac{dy_{\rm p}}{d\nu}\frac{d\nu}{dM}\frac{dM}{dt} \\ &= \frac{2\pi a}{P}\frac{\sin i}{\sqrt{1-e^2}}(e\cos\omega + \cos(\nu+\omega)). \end{aligned} \tag{2.60}$$

The component of the star's reflex motion away from the observer (along the red axis in Fig. 2.10) is then

$$v_{\rm r} = K(e\cos\omega + \cos(\nu+\omega)) + \gamma, \tag{2.61}$$

where γ is the rate of change of distance of the system's centre of mass away from the solar-system barycentre. Note that $\sin i$ is implicit in the value of K:

$$K = \frac{2\pi a}{P}\frac{m_{\rm p}}{m_* + m_{\rm p}}\frac{\sin i}{\sqrt{1-e^2}}. \tag{2.62}$$

The eccentricity of the orbit modifies the durations of transits and occultations through Kepler's second law. The transverse velocity of the planet at mid-transit is obtained by differentiating Eq. (2.55) for $x_{\rm p}$ and substituting Eq. (2.49) to obtain

$$v_t = \frac{2\pi a}{P}\frac{e\sin\omega + 1}{\sqrt{1-e^2}}. \tag{2.63}$$

At first and third contact, the separation of centres is $h = 1 + R_\mathrm{p}/R_*$. The impact parameter at mid-transit is

$$b = \frac{z_\mathrm{p}}{R_*} = -\frac{a}{R_*}\frac{1-e^2}{1+e\sin\omega}. \tag{2.64}$$

For the case where $R_* \ll a$, the approximate transit duration is

$$\frac{t_\mathrm{tr}}{P} \simeq \frac{R_*}{a}\frac{\sqrt{(1+R_\mathrm{p}/R_*)^2 - b^2}}{\pi}\frac{1+e\sin\omega}{1-e^2}. \tag{2.65}$$

2.4.2 *Bayesian Parameter Fitting*

Equations (2.59) and (2.61) provide a complete orbital model for generating synthetic fluxes and radial velocities at a set of times of observation. In the early stages of investigation immediately following detection of transits, however, radial velocities are seldom available. At this stage, we want to know if the transits are real. If they are, we also want to know the relative radii of the star and the transiting companion, and to determine whether the density of the host star is consistent with being on the main sequence.

The data comprise a sequence of observed relative fluxes or magnitudes D_i with estimated variances σ_i^2, and a sequence of model fluxes $\mu(\boldsymbol{\theta}; t_i)$ evaluated at the times t_i of observation. The model depends on a set $\boldsymbol{\theta}$ of model parameters. Assuming the orbit to be circular and the limb-darkening coefficients to be fixed, a photometric data sequence containing multiple transits is fully described by a model with five free parameters: the epoch t_tr of mid-transit, the orbital period P, the scale parameters R_p/R_* and a/R_* and the impact parameter b of the orbit. If the observational errors are assumed to be independent and Gaussian, the joint probability density function (or likelihood) of obtaining the observations conditioned on the model is

$$\mathscr{L} = P(\boldsymbol{D}|\boldsymbol{\mu}(\boldsymbol{\theta})) = \prod_i \frac{1}{2\pi\sigma_i}\exp\left(-\frac{1}{2}\frac{(D_i - \mu(\boldsymbol{\theta}; t_i))^2}{\sigma_i^2}\right). \tag{2.66}$$

Evaluating the product and taking logs, we obtain

$$\ln\mathscr{L} = -\frac{n}{2}\ln(2\pi) - \sum_i \ln\sigma_i - \frac{1}{2}\chi^2 \quad \text{where} \quad \chi^2 = \sum_i \frac{(D_i - \mu(\boldsymbol{\theta}; t_i))^2}{\sigma_i^2}. \tag{2.67}$$

Over the last 10 years, Markov-chain Monte Carlo (MCMC) methods have become popular as a means of determining the joint posterior probability distribution of the parameter set $\boldsymbol{\theta}$ describing this problem (Holman et al. 2006; Collier

Cameron et al. 2007; Burke et al. 2007). Given a vector of state variables $\boldsymbol{\theta}$ from which the model parameters can be calculated, the log likelihood is calculated. Each element θ_j of the state vector is then perturbed by a small amount, usually a Gaussian random deviate scaled to the estimated width σ_j of the posterior probability distribution for that parameter:

$$^{(k+1)}\theta_j = ^{(k)}\theta_j + \sigma_j G[0,1]. \tag{2.68}$$

The log likelihood for this $(k+1)$th state vector is evaluated for the new parameter set. The decision to accept or reject the proposal is made according to the Metropolis–Hastings rule (Metropolis et al. 1953; Hastings 1970). If the log likelihood of the proposal exceeds that of its predecessor, the new state vector is accepted and written into the $(k+1)$th step of the chain. If, however, the log likelihood has decreased, a random number $U[0,1]$ is drawn from the uniform distribution. If

$$\frac{^{(k+1)}\mathscr{L}}{^{(k)}\mathscr{L}} > U[0,1], \tag{2.69}$$

then the proposal is accepted and written into the chain as before. If the proposal fails this test, the proposal is rejected and the (k)th state vector is copied into the $(k+1)$th step of the chain.

If the parameter uncertainties are estimated correctly and the parameters are mutually uncorrelated, the algorithm should converge quickly to a stationary state, exploring the joint posterior probability distribution of the state vector of model parameters in the vicinity of the maximum-likelihood solution. In this ideal state of affairs the proposal acceptance rate should be about 25 %, leading to a correlation length of a few steps for the chains. Unless the maximum-likelihood solution has already been found by other means, there is likely to be a "burn-in" period as the algorithm moves toward the optimal solution. Similarly, the acceptance rate may not be ideal if the jump lengths for any of the parameters are under- or over-estimated. The acceptance rate can be tuned by running the chain for a few thousand steps after achieving the stationary state, discarding the burn-in sequence and re-determining the standard deviations of the chains. The chain can then be continued with the new jump lengths, and the poorly tuned part discarded.

It is best practice to carry out formal tests such as that of Gelman and Rubin (1992) to demonstrate that the chain has reached a stationary state, and to measure and publish the correlation lengths of the chains for each of the parameters. Excessively long correlation lengths may indicate a poor choice of jump parameters, such that two or more parameters are strongly correlated with each other. When two parameters are strongly correlated, a 1-σ jump in either of them inevitably results in a strong decrease in likelihood, leading to very low acceptance rates unless the step size is reduced.

For this reason a/R_* is a poor choice as a state variable in the transit problem, as illustrated on Fig. 2.11. The orbital separation a is strongly constrained by the stellar

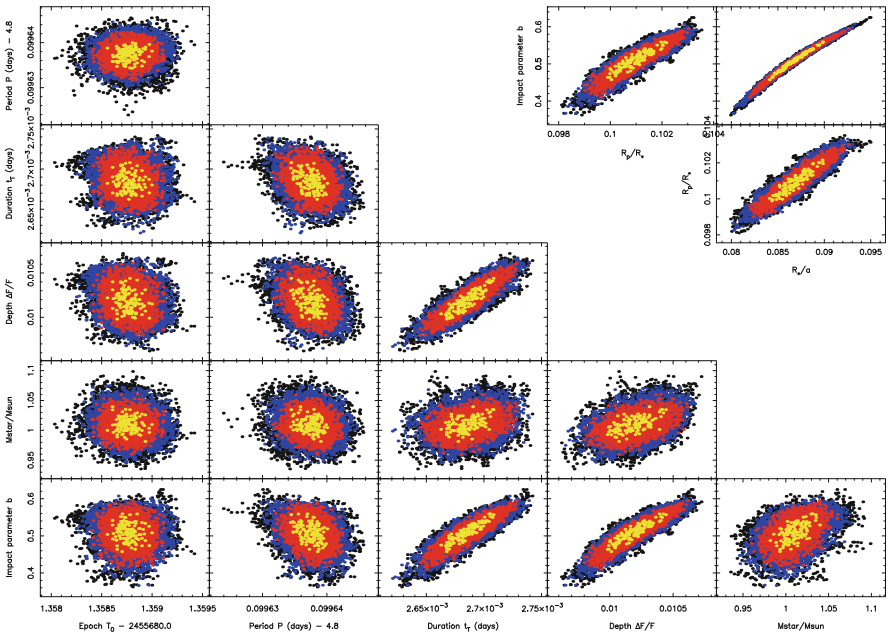

Fig. 2.11 Phase plots showing posterior probability distributions from an MCMC analysis of the transits of WASP-20b, marginalised onto all possible pairs of the six MCMC state variables T_0, P, t_T, $\frac{\Delta f}{f}$, M_* and b. The correlations between the three physical parameters R_*/a, R_p/R_* and b governing the transit profile are also shown. The correlation between scaled stellar radius R_*/a and impact parameter b (*top right*) is seen to be much more extreme than that between transit duration t_T and b. *Yellow, red and blue points* have $2\Delta \ln \mathscr{L} > -2.30,\ -6.17 and\ -11.8$ relative to the maximum-likelihood value (Colour figure online)

mass and orbital period. The transit duration depends on both a/R_* and the impact parameter b. As b increases, R_* must also increase to preserve a good fit to the width of the transit. A better choice of state variable $a/R*$ is the approximation to the fractional transit duration given in Eq. (2.65), which is more nearly independent of b. The value of a/R_* is recovered trivially by inverting Eq. (2.65). Similarly, when fitting eccentric orbits, Ford (2005) recommends $e\cos\omega$ and $e\sin\omega$ in preference to e and ω. Many authors now use $\sqrt{e}\cos\omega$ and $\sqrt{e}\sin\omega$, which imposes a uniform implicit prior on the eccentricity over the range $0 < e < 1$.

Another approach, known as affine-invariant MCMC, is to determine the principal axes of the posterior probability distribution in parameter space. Jumps are then made along the principal axes, whose local directions can be determined efficiently using an ensemble of Markov chains (Goodman and Weare 2010; Gregory 2011). This method is well suited to more complex, high-dimensional problems that are

difficult to orthogonalise. A publicly available code employing this method, EMCEE, is described by Foreman-Mackey et al. (2013).[1]

2.5 Candidate Validation and False-Positive Winnowing

There is an important distinction between false alarms and false positives. False alarms are objects in which transit-detection software produces a signal that passes the detection threshold, but in which no transits are in fact present. False positives, on the other hand, display genuine transit-like events caused by phenomena other than planetary transits (Fig. 2.12).

There are very strong motivations for eliminating astrophysical false positives as early as possible in the discovery process. In ground-based surveys, confirmation is almost invariably made via radial-velocity observations, and a cost of many hours of 2 m-class telescope time per target. For space-based surveys such as NASA's *Kepler* mission, there are simply too many faint candidates to be followed up spectroscopically. The influence of false positives on the statistics of planet occurrence needs to be understood.

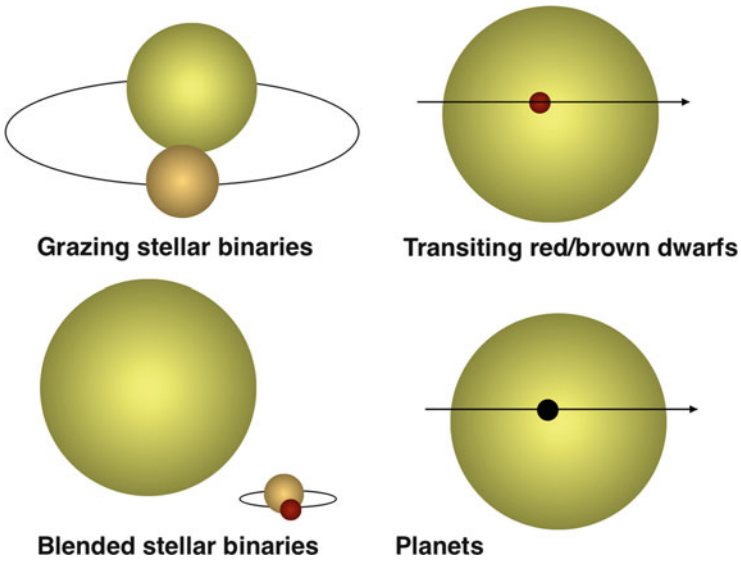

Fig. 2.12 Four types of astrophysical systems giving rise to transits or transit-like events. The blended stellar binaries may be either chance alignments or gravitationally bound hierarchical triples

[1] EMCEE and many other publicly available codes may be found in the GitHub code repository at `https://github.com`.

Validation of transit candidates involves using the transit light curve itself in conjunction with existing catalogue data to determine whether the probability that the transits are caused by a planet exceeds the probability of transit-like events arising from other astrophysical causes, by a large margin.

2.5.1 Astrophysical False Positives

Brown (2003) identified five main classes of objects that can exhibit transit-like events in ground-based surveys. The first are the "good" ones: main-sequence stars with transiting planets, undiluted by the light of nearby objects. This leaves four classes of astrophysical false positive: undiluted main-sequence binaries that are inclined at such an angle to the line of sight that they exhibit grazing stellar eclipses, or with very low-mass stellar companions; main-sequence eclipsing binaries diluted by the light of a brighter foreground star; main-sequence binaries diluted by the light of a brighter, physically associated tertiary star and giant primary stars with main-sequence companions.

2.5.1.1 Grazing Binaries

Grazing binaries are systems in which two stars of roughly equal mass and radius are inclined at such an angle to the line of sight that they exhibit grazing stellar eclipses. This causes small, periodic dips in brightness whose depths are comparable to those expected for planets. The light curves of grazing eclipses are invariably V-shaped, lacking the quasi-flat total phase of a planetary transit. For this reason, model fits to grazing eclipses invariably yield high impact parameters. If the two components of a grazing stellar binary have different effective temperatures, the effective temperature derived for the system from colour indices in optical passbands may differ appreciably from that measured in the near infrared, e.g. using the 2MASS JHK colours. If follow-up photometry in multiple passbands is available, a colour-dependent eclipse depth is a good indicator that the cooler object is self-luminous. Finally, MCMC analysis of the light curve often yields an abnormally low stellar density for the primary, inconsistent with the mass estimated from the system colour.

2.5.1.2 Low-Mass Stellar or Substellar Companions

More plausible-looking flat-bottomed transits can occur in close stellar binaries with very unequal mass ratios. Low-mass stellar or substellar (i.e. brown dwarf) companions have radii comparable to or in some cases smaller than those of gas-giant planets. Their transits show the same rapid ingress and egress seen in planetary transits. If the companion is sufficiently self-luminous, secondary eclipses may be detectable if the discovery photometry is of sufficiently high precision. Rowe et al.

(2014) point out that it is straightforward to estimate the surface brightness of the planet's dayside in both thermal and reflected light. A secondary eclipse whose depth exceeds the expected value indicates a self-luminous object.

A very low-mass star or brown dwarf is sufficiently massive to give significant tidal elongation of the primary star. The strong tidal interaction also leads to rapid synchronisation of the primary's rotation, which may give rise to optical modulation by starspot activity arising from rotationally enhanced dynamo action. The ellipsoidal variation manifests itself as a sinusoidal modulation at twice the orbital frequency, with minima at the times of transit and secondary eclipse. Starspot activity also gives quasi-sinusoidal variability on the orbital frequency, often with a contribution at twice the orbital frequency. Unlike ellipsoidal variability, however, starspot modulation evolves with time. With light curves of sufficient precision (e.g. *Kepler*) relativistic Doppler beaming may also be detectable (Faigler and Mazeh 2011). If the effects of ellipsoidal variation and Doppler beaming are built into the transit model, detectable signals will yield a mass estimate for the secondary, obviating the need for radial-velocity follow-up.

2.5.1.3 Blended Eclipsing Binaries

A chance alignment of a bright, isolated star with a fainter stellar eclipsing binary produces diluted eclipses that can mimic planetary transits. In ground-based surveys, such impostors are often revealed by MCMC analysis, which yields a stellar density inconsistent with the overall colour of the system. Secondary eclipses are often present, and short-period systems may also exhibit ellipsoidal variations. Such systems may also yield inconsistent effective temperatures from optical and 2MASS colours, and exhibit wavelength-dependent eclipse depths. Space-based surveys for smaller planets are vulnerable to a related type of blend, in the form of transiting giant-planet systems diluted by a brighter, nearby star to mimic a much smaller transiting planet.

In ground-based surveys, which have notoriously poor angular resolution, such blends can be identified efficiently with follow-up photometry with larger telescopes having superior angular resolution. "On-Off" photometry, in which sequences of CCD frames are taken within and outside transit, allow faint, resolved stars showing deep eclipses on the transit ephemeris to be identified.

Batalha et al. (2010) found that many false positives caused by background eclipsing binaries display a characteristic astrometric signature in *Kepler* images. If the angular separation is non-zero, the light centroid will move toward the brighter, isolated star when the fainter binary goes into eclipse. Rowe et al. (2014) developed this method further to improve sensitivity to the very small astrometric shifts that betray this type of system. If the two stars in the faint binary have similar effective temperatures, the primary and secondary eclipses will be of comparable depth. Modelling the depths of the odd- and even-numbered transits separately is an effective way to detect primary and secondary eclipses with subtly different depths.

2.5.1.4 Hierarchical Triples

Hierarchical triples bear many similarities to blended eclipsing binaries, except that the faint eclipsing binary and the brighter diluting star are in a physically associated triple system. Their angular separations tend to be small. For example, consider a close binary with eclipses intrinsically 0.5 magnitude deep, diluted by a companion 5 magnitudes brighter at 1 AU separation. The diluted transits would appear 0.005 mag deep. At a distance of 250 pc, the angular separation would be 0.004 arcsec. During eclipse, the centroid would shift 19.5 μas toward the brighter companion. To detect such a shift would be challenging even for the ESA *Gaia* mission (de Bruijne 2012), whose single-measurement precision is of order 30 μas.

2.5.2 False-Positive Winnowing

2.5.2.1 Dwarf–Giant Separation

Proper-motion measurements convey valuable information about the luminosity class of the host star, allowing Bayesian estimation of the relative likelihood that the host star is on the main sequence and not a more distant giant. Gould and Morgan (2003) found that a plot of the reduced proper motion (RPM) against $(B_T - V_T)$ colour from the TYCHO-2 catalogue was effective at separating dwarfs from giants in the planning of transit surveys. Collier Cameron et al. (2007) adapted the method to the 2MASS JHK photometric system, defining the RPM as

$$H_J = J + 5\log(\mu) \tag{2.70}$$

and plotting it against $(J - H)$ colour. Figure 2.13 illustrates how the location of a WASP host star (in this case the slightly evolved WASP-12) in this diagram confirms its main-sequence status. The publication of the first data release from the *Gaia* mission will supersede the RPM method. With an anticipation parallax precision of 8 μas at magnitudes $V < 13$, *Gaia* will determine the distances to the existing WASP host stars, whose estimated parallaxes are all greater than 1500 μas, to a precision better than 0.5 %.

2.5.2.2 Main-Sequence Prior

The stellar density derived from the transit duration provides an independent test of the host star's main-sequence status. The relationship between stellar density and effective temperature arises from the main-sequence mass–radius relation. Tingley and Sackett (2005) pointed out that a simple power-law approximation to the mass–radius relation could be used to define a joint prior probability distribution on the mass and radius. Collier Cameron et al. (2007) describe a Bayesian approach in

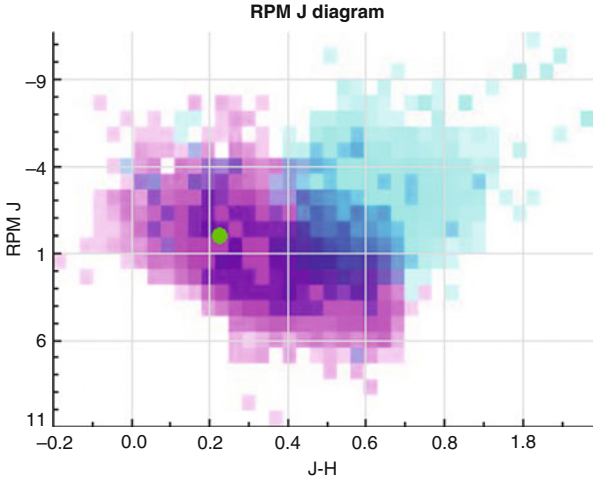

Fig. 2.13 Reduced proper-motion diagram for WASP-12 (*green point*). The magenta density plot represents the probability density for dwarfs among 2000 FGK stars for which high-resolution spectroscopic analyses have been published in the catalogues Valenti and Fischer (2005) and Cayrel de Strobel et al. (2001), cross-matched to the USNO-B1.0 catalogue for proper motions and the 2MASS catalogue for $J - H$. The *cyan plot* gives the probability density for giants from the same catalogues

which the results of MCMC analyses using priors on both mass and radius are compared with those obtained using a prior on the mass alone. The prior on the mass is obtained from the stellar effective temperature, which in turn is estimated from the $J - H$ colour. If the imposition of the main-sequence prior results in a significantly worse fit to the transit duration, the target priority is downgraded as being a likely false positive.

More recent approaches have used the stellar density derived from the transit duration directly. Sozzetti et al. (2007) pioneered the use of ρ_* as a luminosity indicator for stellar evolution models, as a means of determining the physical dimensions of the host star. The "asterodensity profiling" method of Kipping (2014) and Sliski and Kipping (2014) builds on this technique as a false-positive winnowing method.

2.5.3 Validation

Although the methods described above are suitable for selecting ground-based transit candidates for radial-velocity follow-up, the majority of *Kepler* targets are too numerous, too faint, and have expected radial-velocity amplitudes too small, for radial-velocity follow-up to be effective as a means of confirmation. In order to derive the probability density function of planets in the planet radius–orbital

separation plane, it is also necessary to understand the distributions of astrophysical false positives in the same parameter space.

The most problematic contaminants are diluted eclipsing binaries, including diluted main-sequence stars with giant planets. For an individual system, the problem can be posed in terms of fitting a candidate light curve with an eclipsing-binary model plus "third light" from an unresolved, brighter star. The parameter space for hierarchical triples is more restrictive than that for blended eclipsing binaries. Assuming the binary to be detached (as it must usually be to mimic a planetary transit without ellipsoidal variation), all three stars must lie on a single isochrone, and yield a combined colour and light curve consistent with observation.

For blended eclipsing binaries, two of the components must lie along a single isochrone. The probability of finding an eclipsing binary within a given angular separation from a brighter star must be derived from a model of the galactic stellar population in the direction of interest.

Efforts to understand the false-positive rate follow a general approach similar to the BLENDER technique developed by Torres et al. (2005, 2011). BLENDER combines follow-up spectroscopy and imaging at high angular resolution with light-curve fitting, to determine whether a given candidate is more likely to be a planetary system or a false positive. BLENDER is, however, computationally expensive, and the necessary follow-up observations are impractical for the thousands of *Kepler* candidates. This led Morton (2012) to develop an accelerated approach using a simplified light-curve fitting model in conjunction with an arbitrary combination of photometric colours, spectroscopy and adaptive-optics (AO) imaging.

The PASTIS method (Díaz et al. 2014) adopts a fully Bayesian approach to determining whether an individual transit candidate is more likely to be a planetary system or a false positive. This entails modelling the light curve and follow-up observations with a set of different scenarios (planetary system, diluted planetary system, blended or hierarchical eclipsing binary) using MCMC to determine the joint posterior probability distribution for the model parameters. It uses the marginal likelihood of each scenario to determine whether the probability of the planet hypothesis exceeds the combined probabilities of all false-positive hypotheses considered.

A by-product of this kind of validation approach is that for each object considered, it yields the relative probabilities for each of the competing hypotheses, allowing corrections to be made for the various false-positive contaminants in determining the underlying planet population (e.g. Fressin et al. 2013).

2.5.4 Multiple Transiting Systems

Among the major discoveries of the *Kepler* mission is the existence of several thousand systems in which two or more transit signatures are present. A false positive in a multiple transiting system involves either a chance alignment of a single-planet system with a blended eclipsing binary, a hierarchical triple with a

planet orbiting the bright, single component planets transiting both components of an unresolved binary star or else multiple chance alignments.

Lissauer et al. (2014) estimated the expected number of chance alignments of distant eclipsing binaries within *Kepler*'s photometric aperture. Assuming that false positives are randomly distributed among the targets, the probability of a single target displaying j false positives should be described by the Poisson distribution:

$$p(j) = \frac{\lambda^j e^{-\lambda}}{j!}, \tag{2.71}$$

for a population mean of λ false positives per target. In a sample of N targets the expected number of false positives should be $E(j) = Np(j)$. We know that the number of planet candidates per target is less than 2 %, and that the number of false positives must be even less than this, so it is reasonable to set $\lambda \ll 1$ and hence $e^{-\lambda} \approx 1$. If the probabilities that a planet displays one of more false positives, and that it hosts one or more transiting planets, are independent, the probability that a planet displays one or more false positives and hosts one or more transiting planets is simply the product of the two probabilities.

Lissauer et al. (2014) concluded that the vast majority of *Kepler* targets showing multiple transit signatures were true planetary systems. Rowe et al. (2014) used the same framework to validate a sample of 340 planetary systems comprising 851 planets with more than 99 % confidence, without having to resort to spectroscopy or high-resolution imaging for validation.

2.6 Planet Characterisation

What can we learn about the composition of an exoplanet and its atmosphere? Knowledge of the most abundant dusty and molecular constituents of protoplanetary disks and of the planets in our own solar system leads us to believe that the main refractory constituents should be iron and silicates. Water is the dominant volatile likely to be found in the solid or liquid state. Planets with sufficiently high escape velocities and low temperatures should be able to retain atmospheres of essentially primordial gaseous composition. If smaller planets retain atmospheres at all, they are likely to be of relatively high mean molecular weight.

2.6.1 Planet Radius Distribution

The first step in determining the compositions of planets is to examine the histogram of the radii of validated or confirmed planet candidates (Fig. 2.14). The abrupt drop in planet frequency above 3–4 Earth radii is thought to correspond to the threshold

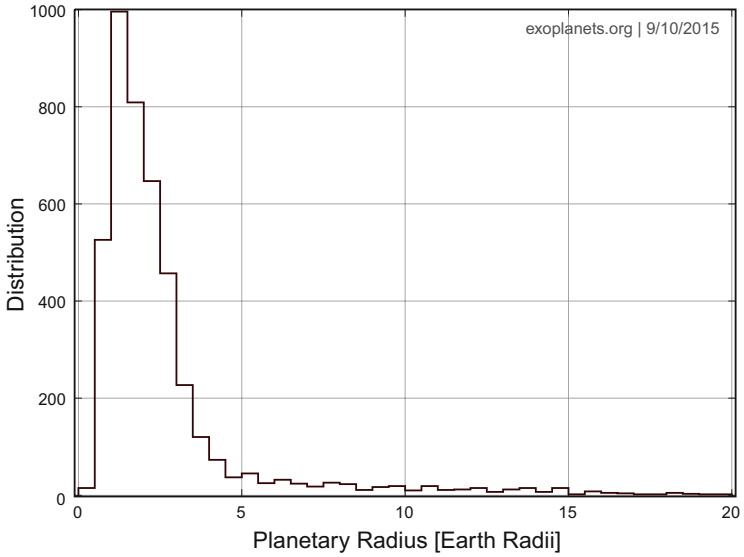

Fig. 2.14 The distribution of radii for confirmed or validated planets from the *Kepler* mission. The paucity of planets with radii less than that of the Earth reflects *Kepler*'s detection threshold rather than any feature of the underlying planet population

mass above which runaway gas accretion occurs, leading to the formation of gas-giant planets (Marcy et al. 2014).

2.6.2 Mass–Radius Relation and Planetary Composition

When the reflex orbital motion of the host star is measurable, and a reliable estimate of the stellar radius is available, the planetary density (and hence mass) can be determined directly using Eq. (2.15) or (2.16). At masses below that of Saturn, planetary radii generally decrease with increasing mass, but the factor-of-two spread in radii at a given mass implies a wide range of compositions (Fig. 2.15). At first glance the radius of a planet alone does not appear to be a reliable indicator of its composition. The radius does, however, appear to give useful insight into the presence or absence of a gaseous envelope. Lopez and Fortney (2014) computed a series of models with a varied mix of iron/silicate and water in the planetary interior, with overlying envelopes of H and He. They found that planets smaller than 1.6 Earth radii cannot retain significant envelopes of H and He, irrespective of interior composition. Planetary radii increase steeply for 1.6–10 Earth radii as the envelope fraction by mass increases from zero to 100 %.

The mass–radius relation for giant planets discovered in ground-based transit surveys (Fig. 2.5) reveals the existence of a large number of inflated hot Jupiters,

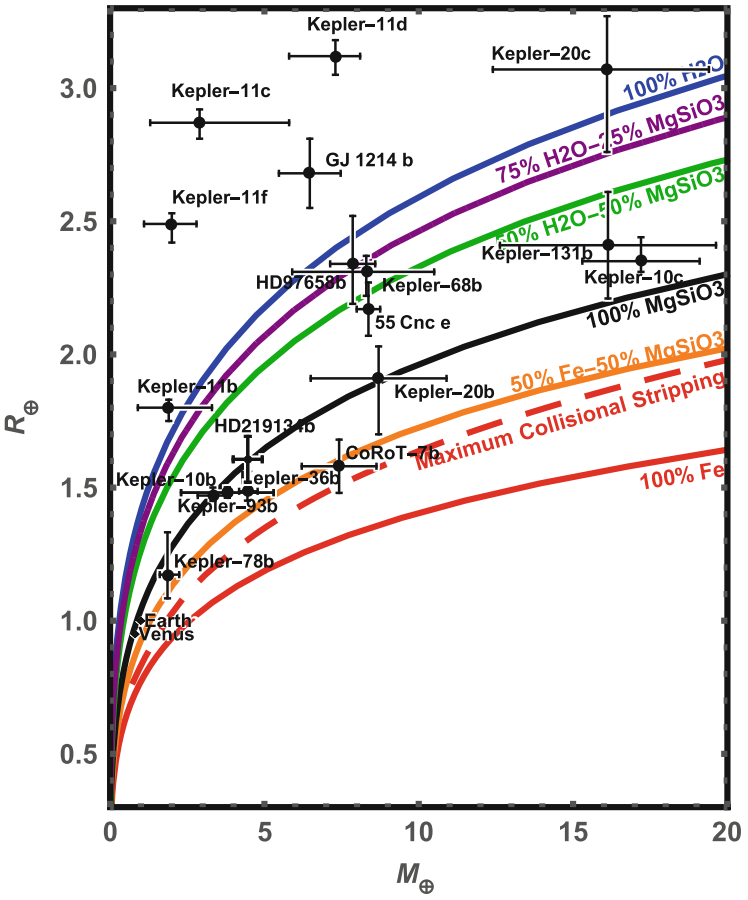

Fig. 2.15 The mass–radius relation for transiting super-Earths and mini-Neptunes whose masses have been determined either by radial-velocity follow-up or by modelling their transit-timing variations. The compositional contours are based on the work of Zeng and Sasselov (2013, 2014). HD 219134b was added with the online tool at `https://www.cfa.harvard.edu/ lzeng/Exoplanet Models.html`

whose radii are greatly in excess of the values predicted by isolated models. A variety of hypotheses have been advanced for the source of internal energy needed to support such inflated radii, ranging from tidal heating during orbit circularisation (Bodenheimer et al. 2001, 2003; Jackson et al. 2008), to ohmic dissipation in partially ionised winds blowing across the planetary magnetic field (Batygin and Stevenson 2010; Batygin et al. 2011; Laughlin et al. 2011) and to irradiation by the host star (Guillot 2005). At present, irradiation appears to be the most compelling explanation. Enoch et al. (2012) found a strong correlation between planet radius and planetary equilibrium temperature for Jupiter-mass planets. Ohmic

dissipation may, however, be important as a means of converting irradiating flux in the atmosphere to internal energy in the deep interior.

2.6.3 Dayside Irradiation and Weather Patterns

The equilibrium temperature of a planet orbiting at distance a from a host star of radius R_* and effective temperature T_* may be estimated by balancing power received against power re-radiated, approximating the planet as a black body radiator:

$$\frac{4\pi R_\mathrm{p}^2}{f}\sigma T_\mathrm{eq}^4 = \frac{4\pi R_*^2 \sigma T_*^4}{4\pi a^2}\pi R_\mathrm{p}^2(1-A). \tag{2.72}$$

This expression makes simple corrections for the Bond albedo A and a factor $1 < f < 2$ representing the efficiency of heat transport from the dayside to the nightside of a tidally locked planet. Isotropic re-radiation is represented by $f = 1$, while re-radiation from the dayside only requires $f = 2$.

In the *Kepler* bandpass, out-of-transit variations are seen in a number of the brighter hot Jupiters, notably HAT-P-7b (Borucki et al. 2009), TrES-2b (Kipping and Bakos 2011) and Kepler-7b (Demory et al. 2011). Being modulated on the planetary orbital period rather than the stellar rotation period, they are attributable to the planetary phase curve rather than stellar activity. At these short wavelengths the phase curve arises mainly from starlight reflected from the planetary dayside, though in very close-orbiting planets far-red thermal emission may be seen as an elevation in the flux at the shoulders of primary transit relative to the purely stellar flux seen at secondary occultation (Hu et al. 2015, and Fig. 2.2).

The temperature contrast between the dayside and the nightside of a planet can be determined directly by observing the full phase curve of the planet around the orbit at thermal-infrared wavelengths. Knutson et al. (2007) made the first phase curve observation of HD 189733b at a wavelength of 8 μm, covering both the primary transit and the secondary eclipse. The difference in planetary flux between the dayside and nightside hemispheres is measured as shown schematically in Fig. 2.2, and yields brightness temperatures of 1212 ± 11 K and 973 ± 33 K, respectively. The phase curve is asymmetric, with minimum brightness occurring slightly after transit and maximum brightness slightly before secondary occultation.

Similar displacements of the hottest part of the dayside from the substellar point have since been observed to varying degrees in the phase curves of WASP-18b (Maxted et al. 2013), HAT-P-2b (Lewis et al. 2013) and WASP-14b (Wong et al. 2015). Theoretical studies of heat transport between the daysides and the nightsides of these planets (Showman et al. 2009; Dobbs-Dixon and Agol 2013) have adapted global atmospheric circulation models to study the consequences of extreme irradiation of the daysides of these tidally locked planets. The simulations

typically produce a super-rotating equatorial jet and large temperature differences between the dayside and nightside.

2.6.4 Transit Spectroscopy

In planets with extended atmospheres, the transit depth is wavelength-dependent. The size of the planet's silhouette is determined by the height at which light passing through the atmosphere at grazing incidence encounters an optical depth of order unity. Brown (2001) developed a simple model for predicting the variation of transit depth with wavelength for atmospheres containing common molecular absorbers such as CO, water and methane. The strength of these absorption features depends on the presence or absence of an opaque cloud deck, and its height in the atmosphere. A high cloud deck produces a nearly featureless spectrum, whereas a clear atmosphere shows molecular absorption or emission, depending on the form of the temperature–pressure structure in the upper atmosphere.

The first successful detection of spectral absorption in a planetary atmosphere was made by Charbonneau et al. (2002), who used time-resolved spectroscopy with the STIS instrument aboard HST to detect the enhancement in the transit depth at the wavelength of the NaI D lines. This suggested that even if a cloud deck was present, atomic sodium was sufficiently abundant in the overlying atmosphere to give a measurable increase in transit depth at the wavelengths of the sodium lines.

Pont et al. (2013) measured the transit depth for HD189733b in several band-passes ranging from the UV to the near infrared, using the STIS, ACS and WFC3 instruments aboard the Hubble Space Telescope. They found an essentially featureless transmission spectrum longward of 1 μm. Shortward of this, the transit depth increased monotonically toward shorter wavelengths from the optical to the near UV. They attributed the featureless infrared spectrum and short-wavelength rise to Rayleigh scattering, presenting grazing-incidence optical depths greater than unity at progressively greater heights above a dusty cloud deck. Pont et al. (2013) pointed out that great care needs to be taken with this type of observation if the host star is magnetically active. Unocculted starspots on the visible stellar hemisphere also deepen the transit by an amount that depends on the spot/photosphere contrast, which increases toward shorter wavelengths. Nonetheless, the Rayleigh scattering slope holds great promise as a means for determining planetary surface gravities (de Wit and Seager 2013), particularly for planets of unspotted early type stars, whose rapid rotation makes it difficult or impossible to measure the reflex orbital motion of the host star.

2.6.5 *Time-Resolved High-Resolution Spectroscopy*

As it transits the face of its host star, a planet in a near-circular orbit exhibits a high radial acceleration. Given Eq. (2.14) for the radial acceleration of the star, the radial acceleration of the planet itself is

$$\dot{v}_{\rm r} \simeq \frac{GM_*}{a^2} = \frac{2\pi K}{P}\frac{M_*}{M_{\rm p}}. \tag{2.73}$$

Using Eq. (2.8) for the approximate transit duration in terms of R_*/a and Eq. (2.73), the range of velocities swept out by the planet during the transit is

$$\delta v_{\rm r} \simeq \frac{P}{\pi}\frac{R_*}{a}\frac{2\pi K}{P}\frac{M_*}{M_{\rm p}}. \tag{2.74}$$

If narrow spectral features originating in the planet's atmosphere can be detected and tracked through transit at high spectral resolution, δv_r becomes directly observable, giving direct and model-independent access to the planet/star mass ratio.

Snellen et al. (2010) used time-resolved spectroscopy during a transit of HD209458b with the CRIRES near-IR echelle spectrograph at the VLT to achieve the first successful observation of this kind. They assumed that CO would be present in the planet's atmosphere, and that the transit depth would therefore increase slightly at the wavelengths of CO absorption features. After careful correction for telluric absorption, and subtraction of the mean stellar spectrum, they cross-correlated the residuals with a model spectrum of CO. The expected feature was weakly detected in the cross-correlation functions of the individual spectra. The radial acceleration of the CCF peak yielded a mass estimate $M_* = 1.00 \pm 0.22 {\rm M}_\odot$ for the host star. In addition, the peak showed a constant velocity offset of 2 km s^{-1} with respect to the system centre of mass, suggesting the presence of supersonic winds blowing toward the planetary nightside.

In addition to revealing the mass of the host star, this technique allows individual molecular species to be identified unambiguously from their unique absorption-line patterns. Moreover, a variant of this method has been applied successfully to the molecular signatures of CO and/or water in the dayside thermal spectra of the non-transiting planets τ Boo b, 51 Peg b and HD 170049b (Brogi et al. 2012, 2013, 2014), as well as the transiting planet HD 189733b (Birkby et al. 2013).

2.7 The Future

Planets that transit their host stars are readily detected in surveys of large numbers of stars. Transiting systems brighter than about 12th magnitude are particularly valuable, because radial-velocity observations allow the planetary mass and bulk density to be determined with 4 m-class telescopes.

New wide-field surveys are currently either in progress or in preparation, whose goal is to increase the number of bright stars hosting small planets, enabling reliable mass determination. NGTS (Sect. 2.2.1) and the NASA *TESS* mission (Ricker et al. 2015) (whose launch is anticipated in 2018) aim to detect large numbers of planets in the 1–4 Earth-radius range orbiting bright K and early M stars. The small radii of the host stars yield comparatively deep transits for small planets, enabling future atmospheric characterisation with larger instruments.

Most valuable of all, however, are the handful of transiting systems bright enough for atmospheric characterisation using either transmission spectroscopy during transit or spectral subtraction at secondary eclipse. The very brightest among these are objects like the super-Earth HD 219134b (Motalebi et al. 2015), and the hot Jupiters HD 209458b and HD 189733b. All of these were discovered in radial-velocity surveys, with the transits subsequently being detected with dedicated space-based or ground-based follow-up photometry. One of the key goals of the Swiss-led ESA S-class mission *CHEOPS* (Broeg et al. 2014), also due for launch in 2018, will be to carry out this type of follow-up on bright stars with low-mass planets, in order to provide targets for subsequent atmospheric characterisation with future large facilities such as *JWST* and the new generation of 20–40 m-class ground-based telescopes.

Although *TESS* will cover the whole sky during its 2-year baseline mission, the stare time on each field of view will be restricted to a month or so. While this allows exploration of the region around M dwarfs where planetary equilibrium temperatures might allow liquid water to exist, the search for "Earth twins" orbiting solar-type stars must await the launch of the ESA M3 mission *PLATO2.0* (Rauer et al. 2014) in 2024 or so. *PLATO2.0*'s 34 small-aperture telescopes will conduct two long pointed campaigns over a field of view 20 times greater than that of *Kepler*, giving access to a much brighter population. Asteroseismology of the host stars of transiting planets will yield precise stellar parameters and ages, opening up the possibility of studying the evolution of planetary atmospheres over the nuclear-burning lifetimes of solar-type stars.

Acknowledgements Andrew Collier Cameron acknowledges the support of the meeting organisers for travel and accommodation at the meeting, and thanks Dr. Raphaëlle Haywood for insightful proof-reading and scientific input.

References

Aigrain, S., Hodgkin, S.T., Irwin, M.J., Lewis J.R., Roberts S.J.: Mon. Not. R. Astron. Soc. **447**, 2880 (2015)
Alonso, R., et al.: Astrophys. J. **613**, L153 (2004)
Alsubai, K.A., et al.: Anal. Chim. Acta **63**, 465 (2013)
Auvergne, M., et al.: Astron. Astrophys. **506**, 411 (2009)
Bakos, G., Noyes, R.W., Kovács, G., Stanek, K.Z., Sasselov, D.D., Domsa, I.: Publ. Astron. Soc. Pac. **116**, 266 (2004)

Bakos, G.Á., et al.: Astrophys. J. **710**, 1724 (2010)
Bakos, G.Á., et al.: Publ. Astron. Soc. Pac. **125**, 154 (2013)
Batalha, N.M., et al.: Astrophys. J. **713**, L103 (2010)
Batygin, K., Stevenson, D.J.: Astrophys. J. **714**, L238 (2010)
Batygin, K., Stevenson, D.J., Bodenheimer, P.H.: Astrophys. J. **738**, 1 (2011)
Birkby, J.L., de Kok, R.J., Brogi, M., de Mooij, E.J.W., Schwarz, H., Albrecht, S., Snellen, I.A.G.: Mon. Not. R. Astron. Soc. **436**, L35 (2013)
Bodenheimer, P., Lin, D.N.C., Mardling, R.A.: Astrophys. J. **548**, 466 (2001)
Bodenheimer, P., Laughlin, G., Lin, D.N.C.: Astrophys. J. **592**, 555 (2003)
Borucki, W.J., et al.: Science **325**, 709 (2009)
Borucki, W.J., et al.: Science **327**, 977 (2010)
Bouchy, F., Pont, F., Santos, N.C., Melo, C., Mayor, M., Queloz, D., Udry, S.: Astron. Astrophys. **421**, L13 (2004)
Bouchy, F., Pont, F., Melo, C., Santos, N.C., Mayor, M., Queloz, D., Udry, S.: Astron. Astrophys. **431**, 1105 (2005)
Broeg, C., Benz, W., Thomas, N.: *CHEOPS* team. Contrib. Astron. Observ. Skalnaté Pleso **43**, 498 (2014)
Brogi, M., Snellen, I.A.G., de Kok, R.J., Albrecht, S., Birkby, J., de Mooij, E.J.W.: Nature **486**, 502 (2012)
Brogi, M., Snellen, I.A.G., de Kok, R.J., Albrecht, S., Birkby, J.L., de Mooij, E.J.W.: Astrophys. J. **767**, 27 (2013)
Brogi, M., de Kok, R.J., Birkby, J.L., Schwarz, H., Snellen, I.A.G.: Astron. Astrophys. **565**, A124 (2014)
Brown, T.M.: Astrophys. J. **553**, 1006 (2001)
Brown, T.M.: Astrophys. J. **593**, L125 (2003)
Burke, C.J., et al.: Astrophys. J. **671**, 2115 (2007)
Carter, J.A., Winn, J.N.: Astrophys. J. **704**, 51 (2009)
Cayrel de Strobel, G., Soubiran, C., Ralite, N.: Astron. Astrophys. **373**, 159 (2001)
Charbonneau, D., Brown, T.M., Latham, D.W., Mayor, M.: Astrophys. J. **529**, L45 (2000)
Charbonneau, D., Brown, T.M., Noyes, R.W., Gilliland, R.L.: Astrophys. J. **568**, 377 (2002)
Claret, A.: Astron. Astrophys. **401**, 657 (2003)
Claret, A.: Astron. Astrophys. **428**, 1001 (2004)
Collier Cameron, A., et al.: Mon. Not. R. Astron. Soc. **373**, 799 (2006)
Collier Cameron, A., et al.: Mon. Not. R. Astron. Soc. **380**, 1230 (2007)
de Bruijne, J.H.J.: Astrophys. Space Sci. **341**, 31 (2012)
de Wit, J., Seager, S.: Science **342**, 1473 (2013)
Demory, B.-O., et al.: Astrophys. J. **735**, L12 (2011)
Díaz, R.F., Almenara, J.M., Santerne, A., Moutou, C., Lethuillier, A., Deleuil, M.: Mon. Not. R. Astron. Soc. **441**, 983 (2014)
Dobbs-Dixon, I., Agol, E.: Mon. Not. R. Astron. Soc. **435**, 3159 (2013)
Enoch, B., Collier Cameron, A., Horne, K.: Astron. Astrophys. **540**, A99 (2012)
Faigler, S., Mazeh, T.: Mon. Not. R. Astron. Soc. **415**, 3921 (2011)
Ford, E.B.: Astron. J. **129**, 1706 (2005)
Foreman-Mackey, D., Hogg, D.W., Lang, D., Goodman, J.: Publ. Astron. Soc. Pac. **125**, 306 (2013)
Fressin, F., et al.: Astrophys. J. **766**, 81 (2013)
Gardner, J.P., et al.: Space Sci. Rev. **123**, 485 (2006)
Gelman, A., Rubin, D.B.: Stat. Sci. **7**, 457 (1992)
Gibson, N.P.: Mon. Not. R. Astron. Soc. **445**, 3401 (2014)
Giménez, A.: Astron. Astrophys. **450**, 1231 (2006)
Goodman, J., Weare, J.: Commun. Appl. Math. Comput. Sci. **5**, 65 (2010)
Gould, A., Morgan, C.W.: Astrophys. J. **585**, 1056 (2003)
Gregory, P.C.: Mon. Not. R. Astron. Soc. **410**, 94 (2011)
Guillot, T.: Annu. Rev. Earth Planet. Sci. **33**, 493 (2005)
Hastings, W.K.: Biometrika **57**, 97 (1970)

Haywood, R.D., et al.: (2014) Mon. Not. R. Astron. Soc. **443**, 2517
Henry, G.W., Marcy, G.W., Butler, R.P., Vogt, S.S.: Astrophys. J. **529**, L41 (2000)
Høg, E., et al.: Astron. Astrophys. **355**, L27 (2000)
Holman, M.J., et al.: Astrophys. J. **652**, 1715 (2006)
Howell, S.B., et al.: (2014) Publ. Astron. Soc. Pac. **126**, 398
Hu, R., Demory, B.-O., Seager, S., Lewis, N., Showman, A.P.: Astrophys. J. **802**, 51 (2015)
Jackson, B., Greenberg, R., Barnes, R.: Astrophys. J. **681**, 1631 (2008)
Kipping, D.M.: Mon. Not. R. Astron. Soc. **408**, 1758 (2010)
Kipping, D.M.: Mon. Not. R. Astron. Soc. **440**, 2164 (2014)
Kipping, D., Bakos, G.: Astrophys. J. **733**, 36 (2011)
Knutson, H.A., et al.: Nature **447**, 183 (2007)
Kovács, G., Zucker, S., Mazeh, T.: Astron. Astrophys. **391**, 369 (2002)
Kovács, G., Bakos, G., Noyes, R.W.: Mon. Not. R. Astron. Soc. **356**, 557 (2005)
Laughlin, G., Crismani, M., Adams, F.C.: Astrophys. J. **729**, L7 (2011)
Lewis, N.K., et al.: Astrophys. J. **766**, 95 (2013)
Lissauer, J.J., et al.: Astrophys. J. **784**, 44 (2014)
Lopez, E.D., Fortney, J.J.: Astrophys. J. **792**, 1 (2014)
Mandel, K., Agol, E.: Astrophys. J. **580**, L171 (2002)
Marcy, G., Butler, R.P., Fischer, D., Vogt, S., Wright, J.T., Tinney, C.G., Jones, H.R.A.: Prog. Theor. Phys. Suppl. **158**, 24 (2005)
Marcy G.W., et al.: Astrophys. J. Suppl. Ser. **210**, 20 (2014)
Maxted, P.F.L., et al.: Mon. Not. R. Astron. Soc. **428**, 2645 (2013)
Mayor, M., Queloz, D.: Nature **378**, 355 (1995)
McCullough, P.R., Stys, J.E., Valenti, J.A., Fleming, S.W., Janes, K.A., Heasley, J.N.: Publ. Astron. Soc. Pac. **117**, 783 (2005)
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: J. Chem. Phys. **21**, 1087 (1953)
Morton, T.D.: Astrophys. J. **761**, 6 (2012)
Motalebi, F., et al.: (2015) arXiv:1507.08532
Pál, A.: Mon. Not. R. Astron. Soc. **390**, 281 (2008)
Pepper, J., et al.: Publ. Astron. Soc. Pac. **119**, 923 (2007)
Pollacco, D.L., et al.: Publ. Astron. Soc. Pac. **118**, 1407 (2006)
Pont, F., Zucker, S., Queloz, D.: Mon. Not. R. Astron. Soc. **373**, 231 (2006)
Pont, F., Sing, D.K., Gibson, N.P., Aigrain, S., Henry, G., Husnoo, N.: Mon. Not. R. Astron. Soc. **432**, 2917 (2013)
Queloz, D., et al.: Astron. Astrophys. **506**, 303 (2009)
Rauer, H., et al.: Exp. Astron. **38**, 249 (2014)
Ricker, G.R., et al.: J. Astron. Telesc. Instrum. Syst. **1**, 014003 (2015)
Rowe, J.F., et al.: Astrophys. J. **784**, 45 (2014)
Seager, S., Mallén-Ornelas, G.: Astrophys. J. **585**, 1038 (2003)
Showman, A.P., Fortney, J.J., Lian Y., Marley, M.S., Freedman, R.S., Knutson, H.A., Charbonneau, D.: Astrophys. J. **699**, 564 (2009)
Sing, D.K.: Astron. Astrophys. **510**, A21 (2010)
Sliski, D.H., Kipping, D.M.: Astrophys. J. **788**, 148 (2014)
Smith, J.C., et al.: Publ. Astron. Soc. Pac. **124**, 1000 (2012)
Snellen, I.A.G., de Kok, R.J., de Mooij, E.J.W., Albrecht, S.: Nature **465**, 1049 (2010)
Sozzetti, A., Torres, G., Charbonneau, D., Latham, D.W., Holman, M.J., Winn, J.N., Laird, J.B., O'Donovan, F.T.: Astrophys. J. **664**, 1190 (2007)
Southworth, J., Wheatley, P.J., Sams, G.: Mon. Not. R. Astron. Soc. **379**, L11 (2007)
Struve, O.: Org. Biomol. Chem. **72**, 199 (1952)
Tamuz, O., Mazeh, T., Zucker, S.: Mon. Not. R. Astron. Soc. **356**, 1466 (2005)
Tingley, B., Sackett, P.D.: Astrophys. J. **627**, 1011 (2005)
Torres, G., Konacki, M., Sasselov, D.D., Jha S.: Astrophys. J. **619**, 558 (2005)
Torres, G., et al.: Astrophys. J. **727**, 24 (2011)

Udalski, A., Szymanski, M., Kaluzny, J., Kubiak, M., Mateo, M.: Anal. Chim. Acta **42**, 253 (1992)
Udalski, A., et al.: Anal. Chim. Acta **52**, 1 (2002a)
Udalski, A., Zebrun, K., Szymanski, M., Kubiak, M., Soszynski, I., Szewczyk, O., Wyrzykowski, L., Pietrzynski, G.: Anal. Chim. Acta **52**, 115 (2002b)
Udalski, A., Szewczyk, O., Zebrun, K., Pietrzynski, G., Szymanski, M., Kubiak, M., Soszynski, I., Wyrzykowski, L.: Anal. Chim. Acta **52**, 317 (2002c)
Valenti, J.A., Fischer, D.A.: Astrophys. J. Suppl. Ser. **159**, 141 (2005)
Wheatley, P.J., et al.: Exploring the formation and evolution of planetary systems. Proc. Int. Astron. Union Symp. **299**, 311–312 (2014)
Wong, I., et al.: Astrophys. J. **811**, 122 (2015)
Zeng, L., Sasselov, D.: Publ. Astron. Soc. Pac. **125**, 227 (2013)
Zeng, L., Sasselov, D.: Astrophys. J. **784**, 96 (2014)

Part III
The Microlensing Method

Chapter 3
Microlensing Planets

Andrew Gould

Abstract The theory and practice of microlensing planet searches is developed in a systematic way, from an elementary treatment of the deflection of light by a massive body to a thorough discussion of the most recent results. The main concepts of planetary microlensing, including microlensing events, finite-source effects, and microlens parallax, are first introduced within the simpler context of point-lens events. These ideas are then applied to binary (and hence planetary) lenses and are integrated with concepts specific to binaries, including caustic topologies, orbital motion, and degeneracies, with an emphasis on analytic understanding. The most important results from microlensing planet searches are then reviewed, with emphasis both on understanding the historical process of discovery and the means by which scientific conclusions were drawn from light-curve analysis. Finally, the future prospects of microlensing planets searches are critically evaluated. Citations to original works provide the reader with multiple entry points into the literature.

3.1 Introduction

Microlensing is an extraordinarily difficult and inefficient means of finding planets. If one conducted a microlensing campaign toward a very favorable star field continuously year after year, then it would require 10 Myr to discover essentially all the planets that could be discovered. That is, in a single year, a fraction 10^{-7} of all the planets would be discovered. Students interested in easy answers would be well advised to skip to the next chapter. In adopting this route, they would travel a well-worn path, first charted by Einstein.

Speculation about gravity's impact on light, based on applying Newtonian gravity to a poorly understood substance, dates to the late eighteenth century, when Michell and Laplace even made the first suggestions of black holes. However, it was not until 1912 that Einstein first explicitly wrote down equations showing that light-bending by massive bodies could magnify the light from more distant sources.

A. Gould (✉)
Department of Astronomy, Ohio State University, 140 W 18th Ave, Columbus, OH, USA
e-mail: gould@astronomy.ohio-state.edu

V. Bozza et al. (eds.), *Methods of Detecting Exoplanets*, Astrophysics and Space Science Library 428, DOI 10.1007/978-3-319-27458-4_3

Although he was already working on general relativity, his calculations were still based on Newtonian gravity. For reasons closely related to the first paragraph, above, these calculations were confined to his notebooks (Renn et al. 1997) and were not published for another 24 years. Einstein began with the impulse approximation, using the assumption (that he himself would prove false within a few years) that the magnitude of the transverse acceleration was $a_\perp = GM\cos^3\phi/b^2$ where M is the mass of the body, b is the impact parameter, and $b\sec\phi$ is the instantaneous distance between the body and the light ray. Then the light deflection is given by

$$\alpha = \frac{\Delta v_\perp}{c} = \frac{1}{c}\int_{-\infty}^{\infty} dt\, a_\perp(t) = \frac{2GM}{bc^2} \to \frac{4GM}{bc^2}, \tag{3.1}$$

where the impulse approximation implies $t = b\tan\phi/c$, and where the last expression takes account of Einstein's own correction due to general relativity.

Einstein next calculated the angular scale of this phenomenon (today called the "Einstein radius"), which in modern notation is written

$$\theta_{\rm E} = \sqrt{\kappa M \pi_{\rm rel}}; \qquad \kappa \equiv \frac{4G}{c^2{\rm AU}} \simeq 8.14\,\frac{\rm mas}{M_\odot}, \tag{3.2}$$

where $\pi_{\rm rel} = {\rm AU}(D_L^{-1} - D_S^{-1})$ is the lens-source relative parallax. In Einstein's derivation, which I will recapitulate below, this appears as an angular normalization in the equation for magnification, but if one is first interested just in the scale, one can consider the simple case of co-linear source, lens, and observer, as illustrated in Fig. 3.1. Then by the small-angle approximation, $b = D_L\theta_{\rm E}$, and by the exterior angle theorem, $\alpha = b/D_L + b/(D_S - D_L)$. Combining these two expressions with Eq. (3.1) yields Eq. (3.2).

While there is no evidence that Einstein worked out the precise probability of gravitational lensing, we know that he did conclude that there was "no great chance of observing this phenomenon," as he explicitly stated when sustained harassment by a Czech engineer led him to write up these results in Einstein (1936). In modern terms, the optical depth (probability of lensing of a given star) is

$$\tau = \int dD_L \pi (D_L\theta_{\rm E})^2 n(D_L) \sim \frac{4\pi GMn}{c^2}D^2 = \frac{4\pi G\rho}{c^2}D^2 \sim \frac{GM_{\rm tot}}{Dc^2} \sim \frac{v^2}{c^2} \tag{3.3}$$

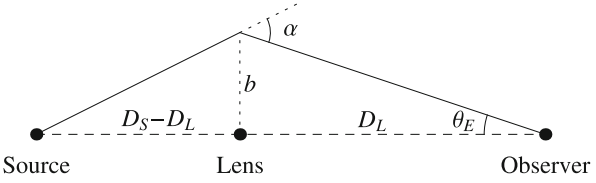

Fig. 3.1 Gravitational lensing geometry in the co-linear case. The image of the source is a ring with radius θ_E

where n and ρ are the number density and mass density of the lenses, D is the generic size of the system, and v is the characteristic velocity of the system (derived from the virial theorem). That is, if we look toward the densest star fields of our Galaxy (the Galactic Bulge) we can expect an optical depth of only $(v_{\rm rot}/c)^2 \sim 10^{-6}$ where $v_{\rm rot}$ is the rotation speed of the Galaxy. Now, the crossing time for a microlensing event is about $D_L\theta_{\rm E}/v_{\rm rot} \sim 30$ days, where I have assumed roughly $M \sim 0.5\,M_\odot$ lenses and lenses at roughly the Galactocentric distance, which implies an event rate of $\Gamma \sim 10^{-5}\,{\rm yr}^{-1}$. If we then note that typical planets are less massive than their hosts by a factor $\sim 10^{-4}$ and that the Einstein radius scales $\theta_{\rm E} \propto M^{1/2}$, we find that planets should be detected in only about 1/100 of events, and so arrive at the depressing conclusion with which this chapter began: only 10^{-7} of all planets in a well-monitored microlensing field will be detected in any given year. If this was enough to discourage Einstein, why should we persist?

The answer one might expect is that we can discover planets and things about planets that are inaccessible to any other technique. And indeed, this ultimately turns out to be the case. However, at the beginning, when microlensing planet searches were proposed by Mao and Paczyński (1991) and Gould and Loeb (1992), it was thought that almost no information would be extracted from planetary detections, other than the planet/star mass ratio q. Stated conversely, the mass, distance, orbital period, eccentricity, radius, composition, etc. would all remain completely unknown. Rather, the attitude of microlensers at the beginning was similar to those charting another seemingly hopeless venture: "We choose to do these things not because they are easy but because they are hard."

3.2 Point Lens Microlensing

Consider a system of an observer, a point-like body of mass M (lens) at distance D_L, and a more distant source at D_S, with the angle between the source and lens being θ_S (i.e., the lens is at the center of the coordinate system). As derived above and shown in Fig. 3.2, the light will be deflected by an angle $\alpha = 4GM/bc^2$, and so arrive at

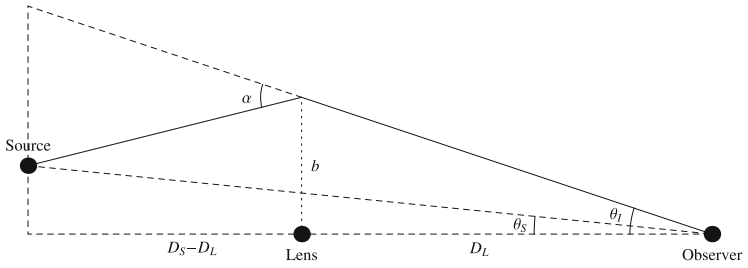

Fig. 3.2 Gravitational lensing geometry in the general case. The image of the source is seen at an angle θ_I instead of θ_S. In realistic astrophysical cases, all angles appearing in this figure are very small (order milliarcsecond for a stellar lens)

the observer at an angle $\theta_I \neq \theta_S$. The small angle approximation implies equality to the two lengths $(\theta_I - \theta_S)D_S = \alpha(D_S - D_L)$. Combining this with the definition $b = \theta_I D_L$ and Eq. (3.1) yields

$$\theta_I(\theta_I - \theta_S) = \frac{4GM\pi_{\rm rel}}{c^2{\rm AU}} \equiv \theta_{\rm E}^2. \tag{3.4}$$

This quadratic equation is easily solved

$$u_\pm = \frac{u \pm \sqrt{u^2+4}}{2}; \qquad u \equiv \frac{\theta_S}{\theta_{\rm E}} \qquad u_\pm \equiv \frac{\theta_{I,\pm}}{\theta_{\rm E}}. \tag{3.5}$$

Note that $u_- < 0$, meaning that this image is on the opposite side of the lens from the source.

The magnifications $A_\pm$ of these images (in the limit of point sources) are given by the derivatives of the image motion with respect to the source motion. Along the lens-source direction these are $\partial u_\pm/\partial u$, while in the transverse direction they are $u_\pm/u$, and so

$$A_\pm = \pm\frac{u_\pm}{u}\frac{\partial u_\pm}{\partial u} = \frac{A \pm 1}{2} \tag{3.6}$$

where I have suppressed the mathematical negativity (i.e., parity) of the minor image by enforcing $A_- > 0$, and where

$$A = \frac{u^2+2}{u\sqrt{u^2+4}} = (1 - Q^{-2})^{-1/2}; \qquad Q \equiv 1 + \frac{u^2}{2}, \tag{3.7}$$

That is, the sum of the magnifications is $A_+ + A_- = A$ and the difference is $A_+ - A_- = 1$. The first expression is the conventional way this magnification is expressed. In this form, it is evident that $A \to 1/u$ for $u \ll 1$. The second form is also interesting, however. It tells us that $A \to 1 + 2/(u^2+2)^2$ for $u \gg 1$. Combined, these two expressions tell us that microlensing is extremely localized: inside the Einstein radius the magnification rises quickly, while outside it approaches unity extremely rapidly. At the boundary, $A(1) = 3/\sqrt{5} \simeq 1.34$. Note that the second form of A is also useful for inverting the equation, in which Q is dual to A, i.e., $Q = (1 - A^{-2})^{-1/2}$, with $u = \sqrt{2(Q-1)}$.

3.2.1 Finite Source Effects

Of course, stars are not actually point sources and this is quantified by the parameter

$$\rho = \frac{\theta_*}{\theta_{\rm E}} \tag{3.8}$$

where θ_* is the source angular radius. In principle, one can evaluate the magnification of a finite source by integrating the point-lens magnification over its surface. However, it is more instructive to consider the case of a finite source that is perfectly aligned with the lens. Then its boundary is a single value of $u = \rho$ and this maps onto the two image circles, one inside and one outside the Einstein radius, $u_\pm = (\sqrt{\rho^2 + 4} \pm \rho)/2$. Since by Liouville's theorem, the surface brightness is conserved, then (for a uniform source) the magnification is

$$A(\rho) = \frac{\pi(u_+^2 - u_-^2)}{\pi\rho^2} = \sqrt{1 + \frac{4}{\rho^2}} \tag{3.9}$$

Hence, for $\rho \ll 1$ (the usual case in microlensing) $A \to 2/\rho$, while for $\rho \gg 1$, $A \to 1 + 2/\rho^2$. For the usual case, $\rho \ll 1$, for which the magnification takes the limiting form $A \to 1/u$, it is straightforward to integrate over a perfectly aligned source, even taking account of its limb-darkened surface-brightness

$$S(\theta) \propto 1 - \Gamma\left(1 - \frac{3}{2}\sqrt{1 - \frac{\theta^2}{\theta_*^2}}\right), \tag{3.10}$$

where Γ is the linear limb-darkening parameter. Then

$$A = \frac{\int d\theta\, \theta S(\theta)/u}{\int d\theta\, \theta S(\theta)} = \frac{2}{\rho}\left[1 + \left(\frac{3\pi}{8} - 1\right)\Gamma\right] \qquad (u \to 0). \tag{3.11}$$

In the opposite limit $u \gg \rho$, one finds by Taylor expanding to quadrupole order that

$$A = \frac{1}{u}\left[1 + \frac{\rho^2}{8u^2}\left(1 - \frac{\Gamma}{5}\right)\right] \qquad (u \gg \rho). \tag{3.12}$$

In particular, Eq. (3.12) implies that an extended source that is approaching a point lens but has not yet transited will rise *more quickly* than a point source, which may be mistaken for effects due to a planet or a binary.

3.2.2 Microlensing Events

This brings us to the question of "microlensing events." Einstein seems to have realized only that stars could be magnified, but not that these magnifications would be changing on short timescales. Liebes (1964) seems to be the first to have clearly stated this, although the resulting point-lens lightcurves are now generally called "Paczyński" curves, following his 1986 paper that made the first practical proposal for microlensing observations (Paczyński 1986). These curves (flux $F(t)$ as a function of time) are defined by three geometric parameters (or four, if there are

significant finite-source effects), plus two flux parameters

$$F(t) = f_s A(\mathbf{u}(t; t_0, u_0, t_\mathrm{E}), \rho) + f_b; \quad \mathbf{u}(t; t_0, u_0, t_\mathrm{E}) = (\tau(t), \beta) = \left(\frac{t - t_0}{t_\mathrm{E}}, u_0 \right). \tag{3.13}$$

Here f_s is the source flux, f_b is the blended flux in the aperture that remains unmagnified during the event, t_0 is the time of maximum, u_0 is the impact parameter (in units of θ_E), t_E is the Einstein crossing time,

$$t_\mathrm{E} = \frac{\theta_\mathrm{E}}{\mu_\mathrm{geo}} = \frac{\sqrt{\kappa M \pi_\mathrm{rel}}}{\mu_\mathrm{geo}}, \tag{3.14}$$

and μ_geo is the lens-source relative proper motion in the geocentric frame (i.e., at t_0).

Figure 3.3 illustrates the magnification as a function of time for microlensing events with several values of the impact parameter u_0 assuming a point-source. If the finite source effect is important, the magnification saturates as described in Sect. 3.2.1. This saturation effect is shown in Fig. 3.4, where an angular source size of $0.1\theta_E$ is assumed.

Equations (3.13) and (3.14) make manifest the fundamental problem of microlensing: most lightcurves are fit by just five parameters ($t_0, u_0, t_\mathrm{E}, f_s, f_b$), and the only one of these that tell us anything about the lens is t_E, which is a complex

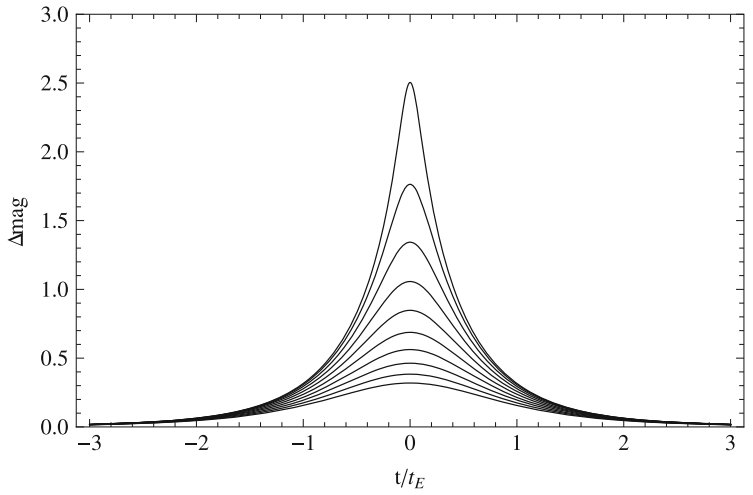

Fig. 3.3 Magnification as a function of time in microlensing events for a set of impact parameters $u_0 = 0.1 \ldots 1$ in steps of 0.1. The lower u_0, the higher the peak magnification. Here we are assuming a point-source [magnification given by Eq. (3.7)]

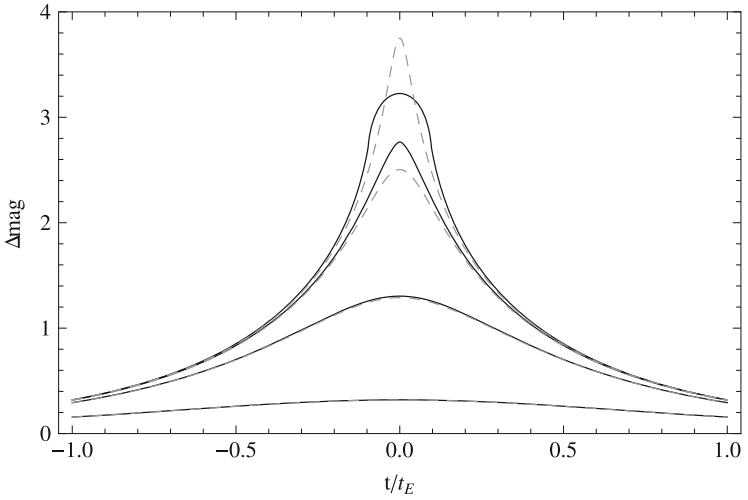

Fig. 3.4 Magnification as a function of time in microlensing events for an impact parameters $u_0 = 10^{-n}$ with $n \in \{-1.5, -1, -0.5, 0\}$. The angular source size is $0.1\theta_E$. Note that when the impact parameter is greater than the source radius, the magnification is higher than the corresponding Paczynski curve (*dashed*). When the impact parameter is smaller than the source radius (source passing right behind the lens), the magnification saturates

combination of three quantities $(M, \pi_{\rm rel}, \mu_{\rm geo})$, two of which are themselves combinations of lens and source properties.

3.2.3 Observations of Microlensing Events

As previously noted, searches for microlensing events have been primarily directed toward the bulge of our Galaxy, since these lines of sight have the highest star density, which raise the low probability of an individual microlensing events quadratically due to the abundance of both lenses and sources. In the last decade, two main collaborations (OGLE observing from Las Campanas, Chile, and MOA from Mt. John Observatory, New Zealand) have alerted nearly all microlensing events by surveying hundreds of fields in the bulge every night from February to early November. These surveys have been aided by several follow-up collaborations (PLANET, MicroFUN, Robonet, MiNDSTEp) exploiting dedicated telescopes scattered at all longitudes in order to ensure full round-the-clock coverage of the most interesting events.

Since 2006 MOA has carried out a high-cadence survey of its fields, with the aim of catching short-duration anomalies in their own data even when no anomaly alert had yet been issued. OGLE started high-cadence survey of its central fields in 2010. The two main surveys have been aided by a third one in Israel, named Wise,

which is disfavored by the northern latitude but still provides coverage of the most important fields. Even with high-cadence surveys running, follow-up observations are still welcome, particularly in the auxiliary low-cadence fields (which are several times larger than the high-cadence fields), but also in order to provide independent confirmations of anomalies, as well as obtaining coverage whenever the survey telescopes are unable to observe for bad weather conditions or technical reasons. The number of planets yearly discovered by survey telescopes alone has now outnumbered those for which follow-up telescopes have played any role, although these survey-only planets typically tend to be less completely characterized.

3.2.4 *Measuring the Einstein Radius, $\theta_{\rm E}$*

Section 3.2.2 has left us with the severe degeneracy between all physical quantities enclosed in the Einstein time t_E. The equations presented there also give the first hint as to how the problem will be solved. If ρ is also measured then $\theta_{\rm E}$ can be determined provided one can measure θ_*, since $\theta_{\rm E} = \theta_*/\rho$. Moreover, θ_* usually can be measured quite well because f_s can be determined from Eq. (3.13). Hence, if we momentarily assume that the source has the same color as the centroid of the Galactic-Bulge red clump (and note that if, as is almost always the case, the source is in the Bulge and so suffers the same extinction as the clump), then $\theta_* = \theta_{*,\rm clump}(f_s/f_{\rm clump})^{1/2}$, where the angular size of clump stars is known from their measured color (Bensby et al. 2013) and magnitude (Nataf et al. 2013) and empirically measured color/surface-brightness relations (Kervella et al. 2004; Bessell and Brett 1988). Then, since the color of the source is usually known from having monitored the event in two or more bands, the difference in color (so surface brightness) is also known, permitting a well-determined correction to the above formula using the same relations (Yoo et al. 2004).

Hence, whenever ρ can be measured, one also determines the product $\theta_{\rm E}^2/\kappa = M\pi_{\rm rel}$ as well as the geocentric proper motion $\mu_{\rm geo}$. Now, unfortunately, this turns out to be quite rare, since typically $\theta_* \sim 0.6\,\mu$as whereas $\theta_{\rm E} \sim 0.3$ mas, so that the probability of finite-source effects is just $p \simeq \rho = \theta_*/\theta_{\rm E} \sim 1/500$. Just the same, as we will see below, this turns out to be quite important.

3.2.5 *The Microlens Parallax, $\boldsymbol{\pi}_{\rm E}$*

Yet another very important parameter, which also can be measured for only a minority of point-lens events, is the "microlens parallax,"

$$\boldsymbol{\pi}_{\rm E} \equiv \frac{\pi_{\rm rel}}{\theta_{\rm E}}\frac{\boldsymbol{\mu}}{\mu} \quad \Rightarrow \quad \pi_{\rm E} = \sqrt{\frac{\pi_{\rm rel}}{\kappa M}}. \tag{3.15}$$

The reason that this combination of parameters is measurable in principle is that if the observer is displaced by 1 AU (in the observer plane) and in a direction $\hat{\mathbf{n}}$, then the lens-source separation vector $\boldsymbol{\theta} = \mathbf{u}\theta_{\rm E}$ will be changed by an angle $\Delta\boldsymbol{\theta} = -\pi_{\rm rel}\hat{\mathbf{n}}$. This corresponds to a displacement in the Einstein ring of

$$\Delta\mathbf{u} = \frac{\Delta\boldsymbol{\theta}}{\theta_{\rm E}} = -\frac{\pi_{\rm rel}}{\theta_{\rm E}}\hat{\mathbf{n}} = -\pi_{\rm E}\hat{\mathbf{n}}. \tag{3.16}$$

The time evolution of this displacement is then determined by the parameter combinations $\pi_{\rm E}$ and direction of motion $(\boldsymbol{\mu}/\mu)$, which is to say, by $\boldsymbol{\pi}_{\rm E} \equiv (\boldsymbol{\mu}/\mu)\pi_{\rm E}$ (Gould and Horne 2013).

The good news is that if both $\pi_{\rm E}$ and $\theta_{\rm E}$ can be measured, then one can derive

$$M = \frac{\theta_{\rm E}}{\kappa\pi_{\rm E}}; \qquad D_L = \frac{\rm AU}{\pi_{\rm E}\theta_{\rm E} + \pi_s}, \tag{3.17}$$

where the source parallax π_s is usually known quite well. The bad news is that it is quite rare for either $\pi_{\rm E}$ or $\theta_{\rm E}$ to be measured for point-lens events, and extremely rare for both to be measured in the same point-lens event. Indeed, out of roughly 20,000 microlensing events discovered to date, there are only two published point-lens events with such measurements, and Gould and Yee (2013) showed that even this number was unexpectedly large.

Why are parallax measurements rare? There are basically two methods of making these measurements from Earth. The first is to observe the event from the accelerated frame of Earth (which one cannot help doing). The larger is $\pi_{\rm E}$, the greater is the impact of this acceleration on the structure of the lightcurve. Unfortunately, typical microlensing events have $t_{\rm E} \sim 20$ days, during which time Earth's accelerated motion is quite well approximated by uniform motion in a straight line. Hence, there is usually no perceptible effect unless $\pi_{\rm E}$ is extremely large. However, this statement does not fully capture the difficulties. When there is an effect, it is usually highly concentrated in the component of $\boldsymbol{\pi}_{\rm E}$ that is parallel to Earth's acceleration (i.e., the direction of the Sun) because this component ($\pi_{\rm E,\parallel}$) leads to an asymmetry in the lightcurve, i.e., falling faster than it rose. Since microlensing is intrinsically symmetric, this effect easily stands out. However, the other component $\pi_{\rm E,\perp}$ induces a symmetric distortion, which is very difficult to disentangle from other symmetric effects. Mathematically, $\pi_{\rm E,\parallel}$ and $\pi_{\rm E,\perp}$ enter the lens equation at third and fourth order in time, respectively (Smith et al. 2003; Gould 2004). Hence, while events yield "information" on $\boldsymbol{\pi}_{\rm E}$, they do not usually yield measurements, and because this 1-D information on $\pi_{\rm E,\parallel}$ is generally of little use, it is rarely even noted in publications.

3.2.6 *Space-Based Microlens Parallax*

The second method is to observe the event simultaneously from two positions separated by $\mathbf{D}_\perp$ (projected on the sky). Normally, the second observer must be in solar orbit, so we typically designate the two as "Earth" and "satellite." The two observers see different t_0 and u_0, from which one derives

$$\boldsymbol{\pi}_{\rm E} = \frac{\rm AU}{D_\perp}(\Delta\tau, \Delta\beta); \qquad \Delta\tau = \frac{t_{0,\rm sat} - t_{0,\oplus}}{t_{\rm E}}; \quad \Delta\beta = u_{0,\rm sat} - u_{0,\oplus}. \tag{3.18}$$

If this is applied to ground-based observations, then $\mathrm{AU}/D_\perp \sim 1/20,000$. Hence, useful results will be obtained only if $\pi_{\rm E}/\max[u_0, \rho] \gtrsim 200$, so that the different lightcurves can be distinguished with $\sim 200/20,000 = 1\,\%$ photometry. That is, very high-magnification events are required, as was indeed the case for the two actual measurements (Gould et al. 2000; Yee et al. 2009).

This is the driver to go to space. However, in this case there is a fourfold degeneracy due to the fact that while t_0 and $|u_0|$ can basically be read off the lightcurve, the sign of u_0 cannot (Refsdal 1966; Gould 1994). Hence,

$$\Delta\beta = \Delta\beta_{\pm\pm} = \pm|u_{0,\rm sat}| \pm |u_{0,\oplus}|. \tag{3.19}$$

Figure 3.5 illustrates the four possible configurations for the source trajectory with respect to the lens as seen from Earth and space. Note, however, that

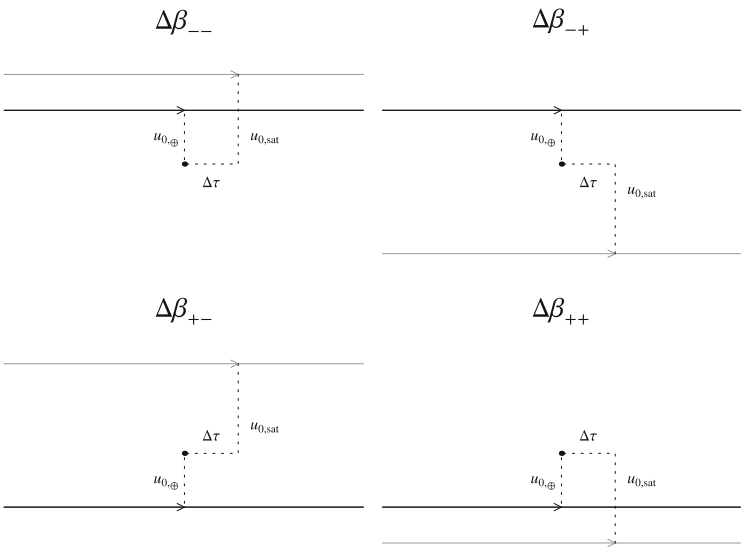

Fig. 3.5 The four possible configurations for the source trajectory as seen from Earth and a satellite. If the orbital motion of the Earth and the satellite is neglected, these four trajectories give rise to the same microlensing light curves. By introducing some curvature, the orbital motion may help break this degeneracy

$\Delta\beta_{++} = -\Delta\beta_{--}$ and $\Delta\beta_{+-} = -\Delta\beta_{-+}$, implying that $\pi_{\mathrm{E},++} = \pi_{\mathrm{E},--}$ and $\pi_{\mathrm{E},+-} = \pi_{\mathrm{E},-+}$. Hence, there is really only a twofold degeneracy in π_{E} and so in the inferred estimates of M and D_L. The remaining degeneracy concerns only the direction of motion, which is generally of less interest.

Calchi Novati et al. (2015a) and Yee et al. (2015a) showed that this twofold degeneracy could be broken in the great majority of cases, mostly by the so-called Rich argument but sometimes due to higher-order effects in the ground-based data. Among the 22 events that they examined, for only two were the parallaxes significantly ambiguous. The Rich argument states that if $(t_0, |u_0|)_\oplus$ is close to $(t_0, |u_0|)_{\mathrm{sat}}$, then the probability that $u_{0,\oplus}$ and $u_{0,\mathrm{sat}}$ have opposite signs $(\Delta\beta_{\pm\mp})$ is roughly $(\pi_{\mathrm{E},++}/\pi_{\mathrm{E},+-})^2$. Calchi Novati et al. (2015b) published a catalog of 170 lightcurves observed by the *Spitzer* satellite at a projected separation from Earth of $D_\perp \sim 1.3$ AU, which will provide the basis for a much richer study.

3.2.7 *Brief Summary of Point-Lens Microlensing*

In brief, point lens lightcurves were originally thought to provide almost no information about individual events. While this remains true for the great majority of microlensing events actually observed, it is in principle possible to get additional information from higher-order effects in ground-based microlensing and/or space-based parallaxes. Such additional information has so far proved to have relatively few scientific implications for point lenses, but lays the basis for understanding how these effects operate in the case of binary and planetary lenses, for which they play a much bigger role.

3.3 Binary Lens Microlensing Basics

Since a single lens is described by three geometric parameters $(t_0, u_0, t_{\mathrm{E}})$, it follows immediately that a binary lens is described by six. In principle, these might be chosen to be $(t_{0,1}, u_{0,1}, t_{\mathrm{E},1}, t_{0,2}, u_{0,2}, t_{\mathrm{E},2})$, i.e., the times of closest approach, impact parameters, and Einstein timescales of the two lenses. Here the respective $u_{0,i}$ are normalized to $\theta_{\mathrm{E},i} \equiv \sqrt{\kappa \pi_{\mathrm{rel}} m_i}$, and similarly the Einstein timescales are $t_{\mathrm{E},i} = \theta_{\mathrm{E},i}/\mu_{\mathrm{geo}}$. While this parameterization has never (to my knowledge) been formally written down and is only occasionally used as a didactic device (in the case of very widely separated lenses), it is in fact extremely closely related to the standard form of the binary lens equation

$$\mathbf{u} - \mathbf{y} = -\sum_{i=1}^{n} \epsilon_i \frac{\mathbf{y} - \mathbf{y}_{m,i}}{|\mathbf{y} - \mathbf{y}_{m,i}|^2} \qquad \epsilon_i \equiv \frac{m_i}{M} \tag{3.20}$$

for the case $n = 2$. Here $\mathbf{y}$ is the image position and $\mathbf{y}_i$ is the lens position of mass m_i. All positions are normalized to the Einstein radius of the *total* mass M. This equation simply equates the vector offset of the source from the image $(\mathbf{u} - \mathbf{y})$ with the sum of deflections caused by the n masses, each according to Eq. (3.1).

Before continuing, I note as a check that when $n = 1$, this equation becomes $(\mathbf{y}-\mathbf{u})|\mathbf{y}-0|^2 = (\mathbf{y}-0)$, where we have adopted $\mathbf{y}_1 = 0$, i.e., the (single) lens is the origin. From the form of this equation, $\mathbf{y}$ must be parallel to $\mathbf{u}$. Hence $(y-u)y = 1$, which is the original single-lens equation.

The reason that $(t_{0,1}, u_{0,1}, t_{\rm E,1}, t_{0,2}, u_{0,2}, t_{\rm E,2})$ is almost never used (or even thought about) for binary lenses, while $(t_0, u_0, t_{\rm E})$ is always used for single lenses is that it corresponds neither to the morphology of the light curve nor to the physical parameters of greatest interest. Rather, one keeps the first three parameters $(t_0, u_0, t_{\rm E})$, where now $t_{\rm E}$ corresponds to the total mass (as in the single lens) and (t_0, u_0) are now referenced to some fiducial center (which could be the center of mass but could be some other definite location). Then the three additional parameters are (s, q, α), where s is the component separation in units of $\theta_{\rm E}$, $q = m_2/m_1$ is the mass ratio, and α is the angle of the source trajectory relative to the binary axis.

As a practical matter, for point sources, binary (and multiple) lenses are solved by first writing Eq. (3.20) in complex notation

$$\zeta = z - \sum_{i=1}^{n} \frac{\epsilon_i}{\bar{z} - \bar{z}_i} \tag{3.21}$$

where the real and imaginary components correspond to the first and second components in the real formulation. By writing $z = \zeta + \sum_i \epsilon_i/(\bar{z} - \bar{z}_i)$, conjugating this to $\bar{z} = \bar{\zeta} + \sum_i \epsilon_i/(z - z_i)$, and then substituting the second into the first, one finds that this reduces (for binaries) to a fifth-order polynomial in z, with coefficients that are functions of the ϵ_i and ζ. This is a bit messy, but the computations are extremely rapid. When the source cannot be treated as a point, the situation is more difficult, but I defer this issue until I have discussed caustics.

3.4 Binary Lens Caustics

Binary-lens light curves are many-fold richer than single lenses, primarily because binaries create "caustics," closed curves of formally infinite magnification. The simplest case to understand is a Chang–Refsdal caustic, which in its mathematical limit occurs when the intrinsically symmetric field of a point lens is perturbed by a background shear. Then the caustic is quadrilateral with one axis aligned by the shear.

3.4.1 Wide Binaries Have Two Quadrilateral Caustics

This situation is well approximated by a binary lens with $s \gg 1$, in which case each star induces a shear field on the other, and so each develops a quadrilateral caustic (see Fig. 3.6, top panel). These caustics have cusp-to-cusp axes of width $w = 2q/s^2$ if s is normalized to the Einstein radius of the mass associated with this caustic (and so the companion is really being thought of as the generator of an *external* shear). This makes sense: the shear goes as m/b^2 where b is the physical separation.

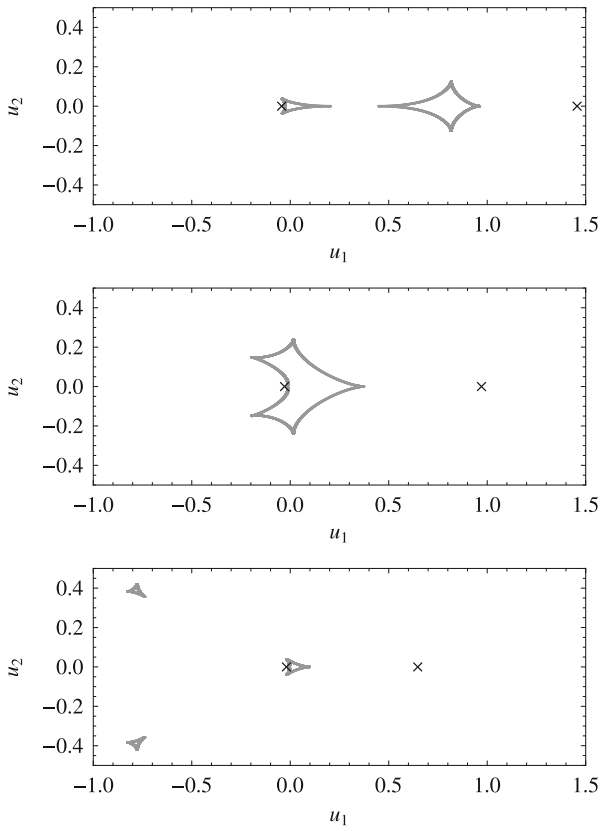

Fig. 3.6 Caustics of a binary lens with a mass ratio of $q = 0.03$. The origin is in the center of mass of the system, with the two masses (indicated by two *crosses*) lying along the u_1 axis, the secondary being on the *right* and the primary on the *left side*. In the *top panel*, we have a separation of $s = 1.5$ in units of θ_E: two 4-cusped caustics lie along the axis of the system, with the secondary one bigger as it is more heavily perturbed by the primary object. In the *middle panel* the separation is $s = 1$: we have a single large 6-cusped caustic. In the *bottom panel* the separation is $s = 2/3$: a central 4-cusped caustic stands in the center of mass and two small triangular caustics are far on the side opposite to the secondary. Note that this central caustic resembles the primary caustic of the opposite limit, illustrating the wide/close degeneracy $s \to 1/s$ discussed in the text

Let us now consider the case of a planetary lens $q \ll 1$, still with $s \gg 1$. Hence, $w_{\rm host} \rightarrow 2q/s^2$, since s is basically the same with or without including the planet mass. Then consider the same geometry from the planet's point of view $w_{\rm planet} = 2Q/s^2_{\rm planet}$ where $w = w_{\rm planet}\sqrt{q}$, $s = s_{\rm planet}\sqrt{q}$ (because both are normalized to $\theta_{\rm E,planet}$ rather than $\theta_{\rm E}$) and $Q = 1/q$. Hence $w = \sqrt{q}/s^2$.

That is, there are Chang–Refsdal caustics associated with both the planet and the host, but the planetary caustic is larger by a factor $\sqrt{q}$. The caustic associated with the host is normally called a "central caustic."

3.4.2 *Resonant Binaries Have One Six-Sided Caustic*

Regardless of the mass ratio, as the two masses are moved closer, both caustics get larger and in particular more extended along the binary axis. Eventually the inner cusps must merge. This occurs at (Erdl and Schneider 1993),

$$s^2 = \frac{(1+q^{1/3})^3}{1+q}, \tag{3.22}$$

at which point, by simple topology, there must be a single six-sided caustic, with two cusps along the binary axis and four off axis (Fig. 3.6, middle panel).

3.4.3 *Close Binaries: One Quadrilateral & Two Triangular Caustics*

In the opposite limit of extremely close binaries, the situation is less intuitive (Fig. 3.6, bottom panel). However, as first shown by Dominik (1999) (and by Griest and Safizadeh 1998 for the special case of planetary lenses), a quadrupole expansion of the lens equation for a very close binary asymptotically approaches the tidal (shear) expansion for a very wide binary. Hence the central caustic structures are essentially identical. Bozza (2000) and then An (2005) carried out this expansion to next order, showing that this degeneracy between close and wide binaries could be very deep.

If we now restrict consideration to planetary lenses, we can gain some intuition for the "rest" of the caustic structure of close binaries, which we will see consists of two triangular caustics.

Planetary lenses can be thought of as perturbations of single lenses. Recall that a single lens has images at $u_\pm = (u \pm (u^2+4)^{1/2})/2$, [i.e., Eq. (3.5)]. That is, it has one image outside the Einstein ring on the same side of the lens (soon to become "host") and one inside the Einstein ring on the opposite side. The planet betrays its presence by perturbing one of these two images. Inverting equation (3.5), the

unperturbed images land directly on the planet when $u = s - s^{-1}$. Hence, if it is a "close" planet, it must be perturbing the latter (inside) image. Now, according to Fermat's principle, images are formed at stationary points on the time delay surface. The outer image is at a minimum of the time delay surface. This stationary point is quite stable, so that when the planet is aligned to this image, it further magnifies the image. In practice, it is sitting inside a quadrilateral caustic. But the inner image is at a saddle point of the time delay surface and so is unstable. When perturbed, it is basically annihilated. This means $A \to A_+ = (A + 1)/2$. Hence, there cannot be a caustic along the binary axis at this point. Since the lens equation must be symmetric with respect to the binary axis, there must be two caustics that are on opposite sides of the binary axis. As the masses are moved further apart, one cusp of each of these caustics must merge with an off-axis cusp from the central caustic. Since the result of this merger is to form a six-sided caustic, the off-axis close binary caustics must each have three sides. For $q \ll 1$ the triangular caustics must be close to the axis simply because the planet's gravity is not strong enough to perturb an image that is far away. For $q = 1$, they must be perpendicular to the axis by symmetry. Hence, with increasing q, they move gradually further from the binary axis.

3.4.4 Magnification Behavior During Approach to a Caustic

What happens at a caustic? As a source crosses from outside a caustic toward the inside, suddenly two new images appear. How is this possible? As described above, the binary-lens equation can always be transformed into a fifth-order polynomial, which always has 5 complex solutions. These solutions may all be solutions of the lens equation but some may not because the polynomial was formed by conjugating the lens equation and substituting back into itself. For example, we can start with a simple equation $x-2 = 0$. If this is satisfied (namely $x = 2$), then it is also true that $x^2 = 4$. But this equation has two solutions $x = \pm 2$, only one of which solves the original equation. Therefore, after the polynomial is solved, one must check whether each of the 5 polynomial solutions solves the original lens equation.

The mapping of the caustics onto the image plane is called the "critical curve." That is, infinitely magnified point sources always appear on the critical curve. These come in pairs (of opposite parity), one just inside and the other just outside the critical curve. Hence, in principle there can be 5, 3, or 1 images. It seems obvious that there cannot be just 1 image from a binary lens, something which is true but not trivial. As the source approaches the side of a caustic from the inside, the images that are about to get annihilated always approach infinite magnification according to $A \propto \delta^{-1/2}$ where δ is the separation from the caustic. Formally, these new images must dominate the binary-lens' magnification at sufficiently small separation. However, if this "small separation" is smaller than the source size, the caustic crossing can be "weak," or even imperceptible.

3.5 Binary Lightcurve Computation

At first sight, it would appear that if one can calculate the magnification of a point source, then it would be a trivial matter to numerically integrate by carrying out many such calculations, for the case that the magnification changes significantly over the source. However, because the magnification diverges at the caustic, this approach must fail unless one determines the exact location of the caustic and organizes very careful numerical integration at the boundary. In practice, this approach is never adopted. Rather calculations are triaged into three classes, which I designate "point-source," "hexadecapole," and "finite-source."

3.5.1 Point-Source, Quadrupole, and Hexadecapole Approximations

If the source is sufficiently far from any caustic, then the magnification is changing slowly over the source, and hence the total magnification can be approximated as its value at the center of the source. Actually, the key criterion is not related to the gradient of the magnification: the point-source approximation is exact in the limit of a pure gradient. Rather, quadratic variation on the scale of the source size is what first undermines point-source calculations. For this reason, a quadrupole order expansion of the magnification field is a sufficient approximation at the onset of the distortions engendered by the approach of a caustic. At next (i.e., even) order, the hexadecapole approximation can be applied (Pejcha and Heyrovský 2009). Gould (2008) gives explicit prescriptions for these two levels of approximation (including limb darkening) that require 5 and 13 point-lens evaluations, respectively. Combined, the point-lens, quadrupole, and hexadecapole approximations suffice for the overwhelming majority of points on essentially any binary microlensing light curve. Nevertheless, the overwhelming majority of computation time is spent on the remaining points for which the source straddles or lies very near a caustic.

3.5.2 Two Methods of Finite-Source Computation

There are broadly two classes of methods for evaluating the magnification in this case: "inverse ray shooting" and "contour integration." Depending on the application, the necessary computations may be completed in a few hours on a laptop, but could require 10^4 to 10^5 CPU hours. Hence, it is appropriate to pay some attention to computing resources.

Both methods rely on fundamentally simple concepts. Inverse ray shooting relies on the fact that while the path (actually multiple paths—one for each image) from the source to the observer is difficult to compute, the path from the observer to the

source plane via any given image position is trivial: one simply evaluates the vector deflection due to each mass (cf. Eq. 3.20). For this reason, this method can be used for an arbitrarily large number of lenses, and indeed it was invented for microlensing of quasars by the stars in an intervening galaxy (Kayser et al. 1986). By Liouville's Theorem, surface brightness is conserved, so one can consider a uniform density of rays over the image (lens) plane. Those that land on the source are "counted" while those that miss the source are not. The method has the advantage that limb darkening is easily accommodated simply by weighting each ray by the surface brightness at the location on the source that the ray happens to intersect.

The main disadvantage is that it is computationally intensive to shoot a dense web of rays, particularly for large (i.e., highly magnified) images. There are two ways to meet this challenge. The first is to shoot a large part of the image plane once, store the results in an efficient manner, and then apply these to many different models being tested (and of course, many points within each model) (Dong et al. 2006). This means that the method can efficiently explore models covering a range of $(t_0, u_0, t_E, \rho, \alpha)$. However, it cannot accommodate varying (s, q), so these parameters must be searched on a grid. This in itself is not a severe problem, but the method basically fails if one must consider orbital motion because in this case separate rays-shootings are required for each point (or at least every several points), which defeats the purpose of shooting a wide grid.

The other way that this approach can be carried out is to efficiently identify those regions of the image plan that must be shot in order to entirely cover a given source (Bennett 2010). However, no matter how efficient, the whole area covering each image must be shot densely in order to obtain a high precision estimate of the magnification, which is the fundamental computational limiting factor for this method.

A second approach is to map the source boundary onto the image plane and then evaluate the total area of the images using Stokes' Theorem (Dominik 1995; Gould and Gaucherel 1997; Dominik 1998). Although the individual computations are of order 10 times longer (since they require solution of a fifth order polynomial, rather than a simple algebraic equation), the method is vastly more efficient because the integration is 1-D rather than 2-D. There are several potential disadvantages. First, the method intrinsically assumes uniform source brightness. However, on one hand, this approximation is almost always adequate during the most difficult phase of modeling, i.e., searching a vast parameter space for the approximate minimum on the likelihood surface. And, on the other hand, the method can easily be adapted to include limb darkening by approximating the profile as, say, 10 annuli of different but uniform surface brightnesses. Second, for the most difficult situations, primarily when the source boundary passes close to a cusp, the task of connecting different fragments of the image boundaries to form a set of closed loops on the image plane can become confused. However, Bozza (2010) has developed an algorithm for dynamically determining the density of boundary evaluations, which effectively evades this problem. The final problem is that, to date, the polynomial expressions for multi-body lenses are only known for two-body (5th order) and three-body (tenth order) lenses. Hence, this method cannot yet be applied to four-body lenses

and higher. Happily, among the 20,000 lens systems discovered to date, none have required four-body solutions.

3.6 Higher-Order Effects in Binary Lenses

As stated above, the basic description of a binary lens requires six parameters, $(t_0, u_0, t_\mathrm{E}, s, q, \alpha)$. A seventh parameter, ρ is often required because a large fraction of recognized binaries exhibit caustic crossings or cusp approaches, for which the source size is crucial. Recall that, by contrast, for single lenses, it is extremely rare that ρ is required. This means that for typical binary lenses, $\theta_\mathrm{E} = \theta_*/\rho$ is measured, so the product $M\pi_\mathrm{rel} = \theta_\mathrm{E}^2/\kappa$ is known.

It is also the case that the microlens parallax $\boldsymbol{\pi}_\mathrm{E}$ is measured much more frequently for binary than single lenses. Long before this became apparent from the accumulation of analyzed events, An and Gould (2001) suggested that this would be so due to the more complicated structure of binary lightcurves, which pins down the timing of caustic crossings with extreme precision. Recall that it was the symmetric (in time) character of single-lens lightcurves that renders it so difficult to measure the $\pi_{\mathrm{E},\perp}$ component of $\boldsymbol{\pi}_\mathrm{E}$. Of course, if $\boldsymbol{\pi}_\mathrm{E}$ and θ_E can both be measured, then so can $M = \theta_\mathrm{E}/\kappa\pi_\mathrm{E}$ and $\pi_\mathrm{rel} = \theta_\mathrm{E}\pi_\mathrm{E}$.

3.6.1 Binary Orbital Motion

However, binary lenses are potentially richer yet. When microlensing planet searches were first proposed by Gould and Loeb, it was not even considered that orbital motion might be detectable. This is because the binary components are separated (in projection) by $r_\perp = s\theta_\mathrm{E}D_L$, which is typically of order 5 AU, implying periods of 10 or more years, while the binary events last only a few weeks. Moreover, at least for planetary events, the caustic structures that yield information on the instantaneous (s, α) are probed for only a day or so.

Hence, when orbital motion was first detected in microlensing event MACHO-97-BLG-41 (Albrow et al. 2000), it was thought to be due to an accidental and highly unlikely circumstance: the source happened to traverse a very small triangular caustic due to a close binary and then also traverse the relatively small central caustic 5 weeks later. From the details of the latter crossing, it was possible to completely reconstruct all the main binary parameters $(t_0, u_0, t_\mathrm{E}, s, q, \alpha)$, and hence to "predict" (i.e., postdict) the time when the source passed by the triangular caustic and at what impact parameter. This prediction turned out to be completely wrong. First, the predicted path missed the caustic completely, whereas the actual path went directly over the caustic. Second, the closest approach was predicted to be about a week earlier than it actually was. These two differences could be explained, respectively, by a change in the orientation of the binary on the sky (relative to the source

trajectory) and a change in the projected separation. These can then be incorporated into the fit by introducing two new parameters, which are often written $d\alpha/dt$ and ds/dt. In order to put these on a symmetric basis, they can also be expressed as a vector with units of inverse time

$$\boldsymbol{\gamma} \equiv (\gamma_{\parallel}, \gamma_{\perp}) = \left(\frac{1}{s_0} \frac{ds}{dt}, \frac{d\alpha}{dt} \right). \tag{3.23}$$

In fact, and despite the early pessimism, these degrees of orbital motion, i.e., the projected internal relative velocity in the plane of the sky (in units of $D_L \theta_{\rm E}$), are measured relatively frequently for binaries and sometimes even for planets. The underlying reason is that caustic crossings typically last about $\theta_*/\mu \sim 0.5\,{\rm mas}/4\,{\rm mas\,yr^{-1}} \sim 1\,{\rm h}$, and hence the time of crossing can be measured with a precision of 1 min or less with reasonably good data. Even in one day, the position of a caustic can change 10^{-3} Einstein radii, for a period of $P \sim 5\,{\rm yr}$, corresponding to an hour for a $t_{\rm E} = 30\,{\rm day}$ event, which is large compared to the precision of measurement. Of course, these rough estimates do not give us a reliable estimate of the precision of measurement in any individual case, particularly because the two dimensions of projected motion are tightly linked with each other (as well as other parameters) in their impact on the light curve. One indication of this linking (there will be several more discussed below) is that Bennett et al. (1999) presented a solution in which the displaced caustic in MACHO-97-BLG-41 was attributed to a circumbinary planet based on a completely independent data set. Eventually this conflict was resolved in favor of the orbiting binary by Jung et al. (2013a) by combining all available data, but this example is a clear warning that inferences from higher-order effects must be vetted carefully.

3.6.2 Ratio of Projected Kinetic to Potential Energy, β

One check that is available if $\theta_{\rm E}$ and $\pi_{\rm E}$ are also measured is the ratio of projected kinetic to potential energy of the binary

$$\beta \equiv \left(\frac{E_{\rm kin}}{E_{\rm pot}} \right)_{\perp} \equiv \frac{v_{\perp}^2 r_{\perp}}{2GM} = \frac{\kappa M_{\odot}}{8\pi^2} \frac{\pi_{\rm E}}{\theta_{\rm E}} \frac{\gamma^2\,({\rm yr})^2 s^3}{(\pi_{\rm E} + \pi_S/\theta_{\rm E})^3}. \tag{3.24}$$

Since $v_{\perp} \le v$ and $r_{\perp} \le r$, this ratio must be strictly less than unity or the binary would not be bound. In addition, if $\beta \ll 1$, it is a warning sign that the solution is suspicious (though not absolutely ruled out). This is because $v_{\perp}$ and $r_{\perp}$ each represent two of the three dimensions of their respective vectors, so one expects that typically $v/v_{\perp}$ and $r/r_{\perp}$ will each be of order only a few at most.

Now, if all the quantities going into Eq. (3.24) are measured, then there remain only two parameters to specify the complete orbital motion of the binary. That is, from q and $M = \theta_{\rm E}/\kappa\pi_{\rm E}$, one specifies the two component masses. From $\boldsymbol{\gamma}$,

$D_L = \mathrm{AU}/(\pi_\mathrm{E}\theta_\mathrm{E}+\pi_s)$ and θ_E, one specifies two components of the internal velocity vector, and from (s,α) one specifies two components of the physical separation. Hence, the binary orbital trajectory would be completely known if v_z and r_z could be measured, i.e., the displacement and velocity into the plane of the sky.

3.6.3 Complete Orbital Solutions

Is this possible? At first sight it would seem not because microlensing is, by its nature, sensitive only to instantaneous positions on the plane of the sky. But a moment's reflection reminds us that this is also true of astrometric measurements, which of course routinely yield 3-D binary orbits out of 2-D data, courtesy of Kepler's Laws. As noted above, the information that is in principle available can be orders of magnitude more precise than needed to measure $\boldsymbol{\gamma}$, particularly if different caustic crossing occurs weeks apart, as occurred for MACHO-97-BLG-41.

Incredibly, Shin et al. (2011; 2012) published such complete orbital solutions for three different microlens binaries. While it now appears that the lens stars are too faint for these solutions to be checked by RV measurements, Skowron et al. (2011) published strong constraints on the full orbit for another event with a much brighter lens. Yee et al. (2015b) did carry out RV observations of this lens and confirmed the microlens orbital solution.

3.7 Degeneracies in Binary Lenses

Because higher-order effects usually give rise to lightcurve deviations that are not easily detectable, much less interpretable by eye, there can be two or more physical conditions that give rise to very similar (in some cases, nearly identical) lightcurve features. Some of the degeneracies that apply to binary lenses are "inherited" from single lenses, while others derive from interactions of a variety of effects, some of which are specific to binaries.

3.7.1 Parallax Degeneracies in Binaries Analyzed at Their Roots

There are several degeneracies between different parallax solutions and several others that derive from interaction of parallax with other microlensing effects. The first degeneracy that is inherited from single lenses is so trivial that it is usually not even recognized; $u_0 \leftrightarrow -u_0$. That is, for single lenses observed from unaccelerated platforms, the lightcurve is identical whether the source passes the lens on its right

($u_0 > 0$) or left ($u_0 < 0$) (see Gould 2004, Fig. 4, for conventions, or Fig. 3.5 in this text). Then if a binary is approximated as having no transverse orbital motion ($\gamma_\perp = 0$) and the observer is approximated as being in rectilinear motion (as opposed to being on an accelerated platform), there is absolutely no way to distinguish between solutions with source directions α and $-\alpha$, provided that one passes the lens on its right ($u_0 > 0$) and the other on its left ($u_0 < 0$). That is,

$$(u_0, \alpha) \leftrightarrow -(u_0, \alpha). \tag{3.25}$$

And moreover, there is no scientific interest in doing so, since these represent indistinguishable physical systems.

Next, recall that for ground-based parallaxes, $\pi_{\mathrm{E},\parallel}$ is third order in time, whereas $\pi_{\mathrm{E},\perp}$ is fourth order in time. Given that most events are short compared to a year (or more specifically, compared to a yr/2π), this means that it very often happens for single lenses that $\pi_{\mathrm{E},\parallel}$ is well determined, while $\pi_{\mathrm{E},\perp}$ is poorly determined (or basically undetermined). Despite this, there are almost no cases in the literature that this is reported, and this for the simple reason that such 1-D parallaxes are almost never of scientific interest. That is, if $\pi_{\mathrm{E},\parallel}$ is well determined, but $\sigma(\pi_{\mathrm{E},\perp}) \gg |\pi_{\mathrm{E},\parallel}|$, then $\pi_\mathrm{E} = \sqrt{\pi_{\mathrm{E},\parallel}^2 + \pi_{\mathrm{E},\perp}^2}$ (which is needed to determine M and π_rel) can take on any value in the interval $|\pi_{\mathrm{E},\parallel} < \pi_\mathrm{E} \lesssim 2\sigma(\pi_{\mathrm{E},\perp})$. Hence, such a measurement is only of interest if $|\pi_{\mathrm{E},\parallel}|$ happens to be so large that the large uncertainty in $\pi_{\mathrm{E},\perp}$ does not undermine the physical interpretation. Gould (2004), Park et al. (2004), and Ghosh et al. (2004) published three examples for single lenses, in which much of the emphasis was exploring the phenomenon itself rather than the scientific interest of the measurement. The only example of which I am aware of such a measurement being published because of the scientific interest derived from large $|\pi_{\mathrm{E},\parallel}|$ for a single lens was Batista et al. (2009a). Most of the published examples have been for binary and planetary lenses, such as OGLE-2005-BLG-071 (Dong et al. 2009) and MOA-2009-BLG-266 (Muraki et al. 2011).

However, even if this 1-D degeneracy can be broken, there often remains, even for single lenses, a discrete remnant of this continuous degeneracy, which is called the "ecliptic degeneracy" because it is exact if the lensing event lies on the ecliptic. That is, if the observer's acceleration projected on the sky is restricted to one dimension, there is absolutely no way to distinguish between single lens solutions characterized by $\boldsymbol{\pi}_\mathrm{E} = (\pi_{\mathrm{E},\parallel}, \pi_{\mathrm{E},\perp})$ and $\boldsymbol{\pi}_\mathrm{E} = (\pi_{\mathrm{E},\parallel}, -\pi_{\mathrm{E},\perp})$. Even though fields near the ecliptic occupy a small fraction of the whole sky, they contain all of the area that is targeted for microlensing planet searches. In particular, the Galactic center lies just $\sim 6^\circ$ south of the ecliptic.

When this $\pi_{E,\perp}$ degeneracy is combined with the (u_0, α) degeneracy discussed above, one obtains a symmetry,

$$(u_0, \alpha, \pi_{\mathrm{E},\perp}) \leftrightarrow -(u_0, \alpha, \pi_{\mathrm{E},\perp}). \tag{3.26}$$

If we now consider the possibility that the binary axis is rotating (i.e., $\gamma_\perp \equiv d\alpha/dt \neq 0$), then

$$(u_0, \alpha, \pi_{\mathrm{E},\perp}, \gamma_\perp) \leftrightarrow -(u_0, \alpha, \pi_{\mathrm{E},\perp}, \gamma_\perp). \tag{3.27}$$

These degeneracies are discussed in detail by Skowron et al. (2011).

The discrete degeneracy described by Eq. (3.27) is further complicated by the fact that $\pi_{\mathrm{E},\perp}$, and $\gamma_\perp$ can easily be continuously degenerate with each other (Batista et al. 2009b; Skowron et al. 2011).

3.7.2 Space-Based Parallax Degeneracies For Binaries

Like single lenses, binaries are impacted by space-based parallax degeneracies, but these take on a significantly different form. Recall that in the fourfold single lens degeneracy, there were two pairs of degenerate solution that had the same magnitude of π_{E} (and so were equivalent for the most important application). This degeneracy remains for binaries, with a form that is very similar to the ecliptic degeneracy

$$(u_0, \alpha, \pi_{\mathrm{E},-+}) \leftrightarrow -(u_0, \alpha, \pi_{\mathrm{E},+-}), \qquad (u_0, \alpha, \pi_{\mathrm{E},++}) \leftrightarrow -(u_0, \alpha, \pi_{\mathrm{E},--}). \tag{3.28}$$

And of course this degeneracy also extends to rotating binaries [with finite $\gamma_\perp$, as in Eq. (3.27), Zhu et al. 2015a]. The only difference between this degeneracy and the ecliptic degeneracy is that the axis of the degeneracy is the Earth-satellite separation vector (projected on the sky) rather than the ecliptic. However, if the satellite orbits near the ecliptic (as for example, both *Spitzer* and *Kepler* do) and if the lens lies on or near the ecliptic, then these degeneracies are identical.

It was originally believed that the other twofold degeneracy, which affects point lenses (in the magnitude of π_{E}), would not impact binaries because binaries break the symmetry in point lenses that apparently gives rise to it. However, Zhu et al. (2015a) found that in the very first case of a caustic-crossing binary with space-based parallax, that this degeneracy persisted. This degeneracy appeared to be "accidental" in the sense that it would have been broken if the space data had covered the caustic entry as well as its exit. However, since space data in their present form are taken for limited intervals due to Sun-angle restrictions, and since some space data (such as *Spitzer*) are triggered from the ground by, e.g., caustic entrances, this type of degeneracy is likely to be fairly frequent.

Another, more profound, type of degeneracy that arises from a single caustic crossing was anticipated theoretically by Graff and Gould (2002), but not given much thought in practice until it was rediscovered by Shvartzvald et al. (2015). Since caustics are 1-D structures, it is generally not obvious from the lightcurve exactly where along the caustic it is being crossed, even though the caustic structure

itself may be very well-defined from the more complete ground-based light curve. If there are two caustic crossings, this degeneracy is in general easily resolved, but (as noted above), events with two space-based crossings are likely to be the exception rather than the rule.

3.7.3 Xallarap vs. Parallax

A completely different type of degeneracy affecting parallax measurements comes from xallarap, which is the effect on the light curve due to accelerated motion of the source about its own binary companion. A moment's reflection shows that xallarap can in principle perfectly mimic ground-based parallax effects provided that the orbit of the source exactly mimics the reflex motion of Earth. However, this also provides the key clue as to how to distinguish xallarap from parallax. One considers a range of, for example, circular source orbits. These have three degrees of freedom not possessed by Earth. First, they can have arbitrary period (rather than just a year). Second, they can have arbitrary inclination (rather than the ecliptic latitude of the source, and third, they can have arbitrary phase (rather than the ecliptic longitude). If a free fit to these three xallarap parameters returns the Earth values, then the chance that the light curve distortions are due to xallarap is very slight. In practice, because events are short relative to a year, such 3-parameter fits are relatively unconstrained. However, if the xallarap period is fixed at $P = 1$ yr, and the fit for the other two parameters returns the ecliptic coordinates of the event, then the chance of xallarap is small.

Poindexter et al. (2005) studied 22 single-lens events with apparent parallax signals $\Delta\chi^2 > 100$ and found that five of these had significant evidence for xallarap. This is in rough agreement with naive estimates. That is, roughly 10 % of all G dwarfs (the progenitors of essentially all microlens sources) have binary companions with $P \lesssim 1$ yr, and so should have xallarap effects (whether detectable or not). This is of course a factor $\sim$10 smaller than the number of events that are in principle affected by parallax (i.e., 100 %). But just as parallax is easier to detect if $\pi_{\rm E} = \sqrt{\pi_{\rm rel}/\kappa M} \propto \sqrt{1/D_L - 1/D_S}$ is big, xallarap is easier to detect if $\sqrt{1/D_{LS} - 1/D_S}$ is big. Here $D_{LS} = D_S - D_L$. To more directly compare these, we should compare the arguments $D_{LS}/D_L D_S$ vs $D_L/D_{LS}D_S$. Since there are more lenses that are close to the source than to the observer, this comparison significantly favors xallarap. On the other hand, while Earth is orbiting the Sun, microlensed sources are typically orbiting much lower mass companions, so the amplitude of their motion is correspondingly smaller. Overall, it is quite plausible that of order 10 % of all events with parallax-like signals are actually due to xallarap. In practice, xallarap is rarely checked for in binary and planetary events with apparent parallax signatures, so this historical evidence and these order-of-magnitude estimates should be regarded as a caution. For completeness I note that xallarap does not affect space-based parallaxes.

Finally, I have already remarked on the close-wide degeneracy, which for planetary lenses can be essentially exact. This degeneracy is so well known and so well understood that it is virtually always checked for, and essentially all papers report on the results of this check.

3.8 Major Discoveries of Planetary Microlensing

Fewer than 1 % of all exoplanets and strong exoplanet candidates were found by microlensing. Nevertheless, because microlensing discovery space is unique, this relative handful of detections contains quite a few major discoveries.

3.8.1 Cold Neptunes Are Common

The first major discovery was that "Cool Neptunes Are Common," which was the title of the Gould et al. (2006) paper announcing the discovery of the second cold Neptune. It is instructive to understand the original basis of this claim and how it has been corroborated.

The first of these two cold Neptunes was OGLE-2005-BLG-390 (Beaulieu et al. 2006), whose lightcurve is shown in Fig. 3.7. Gould and Loeb (1992) had argued that planetary microlensing events could be analyzed basically by eye, and while this is not true of all events, it is true of this one. One simplifying factor is that the source is extremely bright and so very likely unblended. In fact, detailed modeling

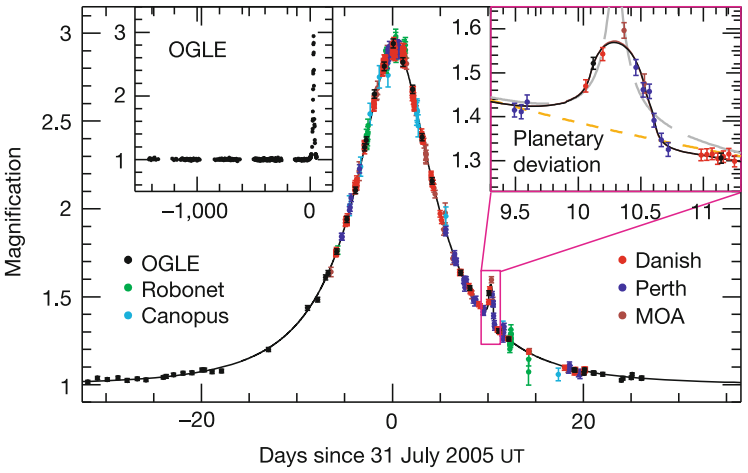

Fig. 3.7 Light curve and model for the planetary event OGLE-2005-BLG-390. Reprinted by permission from Macmillan Publishers Ltd: Nature 439, 437, copyright (2006)

shows that this is the case. Hence, the fact that the underlying event rises by $\Delta I \simeq 2.5 \log(3/\sqrt{5}) = 0.32$ mag about $\Delta t = 10$ days before peak implies that it was just entering the Einstein ring, i.e., $u \simeq 1$ at that point. Since the maximum magnification is $A_{\rm max} \simeq 3$, this implies $u_0 \sim 0.35$. Hence, $t_{\rm E} = \Delta t \sqrt{1 - 0.35^2} \simeq 9.4$ days.

The next point to know is that the planetary deviation, which occurs about 10 days after peak, has a half-width of $t_p = 0.3$ days and height (above the curve traced by the main event) of $A_p = 0.2$. The reason that this amplitude is so low is that the source radius is much bigger than the Einstein ring of the planet, $\theta_* \gg \theta_{{\rm E},p}$, implying that only a small part of the source is effectively magnified. Now, if the planet were an isolated mass (which obviously it is not), then according to Eq. (3.9),

$$A_p = \frac{2}{\rho_p^2} = 2\left(\frac{\theta_{{\rm E},p}}{\theta_*}\right)^2 \tag{3.29}$$

Remarkably, however, Gould and Gaucherel (1997) proved that this formula does actually apply to planetary caustics (just as it would to free-floating planets), provided that the caustic is associated with the major image (i.e., "wide binary" case). [For large sources superposed on minor image caustics, $A_p \simeq 0$.] Hence, Eq. (3.29) is in fact valid.

The duration of the planetary anomaly provides another simple relation: from $\mu = \theta_{\rm E}/t_{\rm E}$ and $\mu = \theta_*/t_p$, we have,

$$\frac{t_p}{t_{\rm E}} = \frac{\theta_*}{\theta_{\rm E}}. \tag{3.30}$$

Finally, combining these, we obtain the mass ratio,

$$q = \frac{m_p}{M} = \frac{\theta_{{\rm E},p}^2}{\theta_{\rm E}^2} = \frac{\theta_{{\rm E},p}^2}{\theta_*^2}\frac{\theta_*^2}{\theta_{\rm E}^2} = \frac{A_p}{2}\frac{t_p^2}{t_{\rm E}^2} \simeq 1.0 \times 10^{-4}. \tag{3.31}$$

This by-eye estimate compares very favorably to the value $q = 0.8 \times 10^{-4}$ based on detailed modeling. The latter should be compared to the value for Neptune/Sun, i.e., $q = 0.5 \times 10^{-4}$.

However, there still remains the question of why we call this a "cold Neptune"? The first point is that since the perturbation occurs 10 days after peak, i.e., $u = 1$, this implies a normalized projected separation $s = (u + \sqrt{u^2 + 4})/2 = 1.6$ (Eq. (3.5)). This is at the edge of the so-called lensing zone, where most planets are found, simply because this is the region of significant magnification of the images due to the host.

As noted above, we were able to measure the ratio of the source size to the Einstein radius $\rho = \theta_*/\theta_{\rm E} = t_p/t_{\rm E} \simeq 0.03$. This has proven to be the case for most planetary events simply because the planet is usually recognized from an interaction of the source with a caustic due to the planet, whose profile depends on ρ. Because

the source color (so surface brightness) and magnitude are measured during the event, they yield θ_*, which in this case is found to be $\theta_* = 5\,\mu\text{as}$. Hence, $\theta_E = 0.17\,\text{mas}$. By the definition of $\theta_E = \sqrt{\kappa M \pi_{rel}}$, this implies

$$\frac{M}{M_\odot}\,\frac{\pi_{rel}}{\text{mas}} = 0.0034\,. \tag{3.32}$$

Hence, there are two choices. If the host is a brown dwarf $M < 0.08\,M_\odot$, then the Neptune is certainly cold. If not, then $\pi_{rel} < 0.04\,\text{mas}$, i.e., $D_L > 6\,\text{kpc}$. At the lower limit, $(M, D_L) = (0.08\,M_\odot, 6\,\text{kpc})$, the Neptune would still be cold. In general, its projected separation is given by

$$r_\perp = s\theta_E D_L = 2.2\,\text{AU}\frac{D_L}{8\,\text{kpc}}. \tag{3.33}$$

Hence, in this limiting case, it would lie 0.6 AU from a very late M dwarf. Only a much more massive host could warm the Neptune at these separations. At much higher masses, π_{rel} is extremely small, according to Eq. (3.32), and therefore the lens lies close to the source, i.e., in the bulge at $D_L \simeq 8\,\text{kpc}$. Hence, according to Eq. (3.33), it must have mass $M > 0.8\,M_\odot$ in order for the Neptune to lie inside the snow line (typically taken to be $r_{snow} = 2.7(M/M_\odot)$).

Is there anything that rules out such "large" masses? No. However, there is an argument that this is improbable. From Eq. (3.32), $M > 0.8\,M_\odot$ implies $\pi_{rel} < 4\,\mu\text{as}$, and so $D_{LS} = \pi_{rel} D_S D_L < 250\,\text{pc}$. Hence, while such solutions are not ruled out, they occupy a very small part of phase space and hence are very unlikely. In any case, the $r_p > 2\,\text{AU}$.

The other detection from Gould et al.'s (2006) "sample of two" cold Neptunes, OGLE-2005-BLG-169, lies at the opposite extreme of ease of interpretability. First, this is a high magnification event in which (as in the great majority of such cases) the planetary anomaly occurs at peak. Hence, the features of this anomaly are entangled with those defining the underlying event due to the host star. When Gould and Loeb (1992) suggested that planetary events could be interpreted by inspection, they explicitly did not consider high-magnification events (which were suppressed by the formalism they used). To my knowledge, no high-mag event has ever been interpreted by eye at the level just described for OGLE-2005-BLG-390. A second complication in this case, however, was that only half the peak was covered. This is basically due to the "organized chaos" by which high-mag events were observed at that time, which still largely prevails. In particular, its high-mag nature was derived from OGLE observations on the night of the peak, but for technical reasons, intensive observations could not be organized from Chile. Instead, intensive observations were initiated ad hoc from Arizona several hours later. As with quite a few other high-mag planet detections, visual inspection of the light curve does not reveal any planetary deviation at all. Rather, it is only the strong deviations from a point-lens fit that provides evidence for a planet. Even so, it was only the fact that these deviations show an abrupt change of slope that convinced the authors

to publish this result. Such slope changes are a generic feature caused by caustic crossings but are essentially never observed in the intrinsic variability of stars, nor in systematic effects due to, e.g., rapidly changing transparency or clouds. Moreover, in the months following the detection of this anomaly, it was not possible to unambiguously assign planet parameters to the event. Many different solutions were found by many different workers in the field until this event triggered some of the first systematic solution-search algorithms that resolved most of these ambiguities. At the time of publication (2006), there remained two classes of solutions, but these had essentially the same physical implications.

Gould et al. (2006) then argued that each of these two detections could lead to an estimate of the frequency of cold Neptunes, one from high-mag events and the other from planetary-caustic events. The best estimates of the frequency were high in both cases. Then they argued that while each such estimate had extremely large errors (due to "statistics of one"), the combined estimate is much more robust.

Given the number of caveats, uncertainties, and unconventional statistical approaches of the original paper, it is of interest to see whether and how these conclusions were tested and/or confirmed. One important avenue of confirmation was the detection of the host star in Keck and *Hubble Space Telescope (HST)* data almost a decade later. Of course, independent of whether there was any planetary signature in the light curve, there certainly was an underlying event, and therefore it was virtually inevitable that this star would eventually be detected as it separated from the source. Hence, this detection of the lens does not in itself prove that it is a "host" of a planet. However, as I have emphasized several times, caustic crossings allow measurement of the self-crossing time $t_* \equiv \rho t_{\rm E}$. When combined with a measurement of θ_* (from the source color and magnitude), this enables a measurement of the proper motion $\mu = \theta_*/t_*$. Hence, the magnitude of the lens-source separation was predicted specifically as a result of the planetary (i.e., caustic-crossing) model. This prediction was somewhat uncertain due to the above-mentioned ambiguities, but the whole range of possible proper motions was greater than twice the value for typical events. This high predicted proper motion was confirmed by two independent measurements using high-resolution imaging from Keck (Batista et al. 2015) and *HST* (Bennett et al. 2015).

More fundamentally, the conclusion that cold Neptunes are common is confirmed by the continuing stream of their detections. Perhaps the best summary of this result is given by Fig. 10 from Fukui et al. (2015) and Fig. 1 from Sumi et al. (2015), which shows a scatter plot of detections in mass and projected separation. As one moves from Jupiter to Neptune mass planets, the density of detections per unit log mass remains roughly constant. Since detection efficiency scales roughly with the size of the planetary Einstein radius, $\theta_{{\rm E},p} \propto m_p^{1/2}$, this implies that cold Neptunes are intrinsically about four times more common than cold Jupiters.

3.8.2 *Free-Floating Planets Are Common*

Another major discovery from microlensing is that free-floating planets (FFPs) are extremely common, i.e., about twice as common as stars (Sumi et al. 2011). This result is extremely puzzling because it is quite difficult to understand how planets can exist in free space, not directly associated with stars, except if they were somehow expelled from their former hosts. And if they were expelled, then there should be something left behind that expelled them, which in the great majority of cases should be heavier than the bodies that were expelled. However, since these FFPs are estimated by Sumi et al. to be of order $M \sim M_{\rm jup}$, this would imply that most stars should typically have super-Jupiter companions. This is certainly not the case, since such massive planets are easily detectable due to their large caustics, and are quite rarely detected in microlensing events, i.e., certainly much less frequently than cold Neptunes or even cold Jupiters.

Another possible explanation for this result is that these planets are not actually free, but just so far from their hosts that the latter leave no trace in the microlensing event. How far is "far"? This depends on the geometry of the event and the quality of the data. From Eq. (3.7), for $u \gg 1, A \to 1+2/(u^2+2)^2$. Hence, for $u_0 = 3$ or 4, the peak magnification of the primary event is just $\sim$1.015 or $\sim$1.006, respectively. Thus, even at these separations the "bump" from the event due to the host would be undetectable except for the very small minority of bright-source events. At larger separations, such "primary events" would be completely invisible regardless of source brightness. Of course, while by chance some events with $s = 5$ will have $u_0 \sim 5$, others will be aligned differently, and therefore the source will eventually (before or after the planetary anomaly) pass close to the host. Poleski et al. (2014) found such an event, although the planet was of roughly Neptune mass, not a super-Jupiter.

However, hosts need not betray themselves directly. True FFPs will give rise to true point-lens events, i.e., with Paczyński curves, whereas bound planets are a form of "wide binaries" and therefore have caustic structures. If the data are good enough, then the deviations from a point-lens curve can be detected. Bennett et al. (2012) found a dramatic example of a bound planet with no direct trace of the host in the light curve. Of course, the larger is s, the better the data have to be to detect such things. Sumi et al. (2011) looked for both types of signatures for each of their FFP candidates. While the minimum distance varied from case to case, typical values of s must be quite large, $s \gtrsim 5$ to "avoid" detecting signatures of the host so consistently. Thus, even if these FFPs turn out to be bound, it remains quite puzzling how so many Jupiter-mass planets form so far from their hosts (2 per star), when they are detected much less frequently at closer separation $0.5 < s < 2$ where microlensing is most sensitive to them.

Thus we come to the final possible explanation for these FFP events, i.e., that they are not FFPs at all but rather some other microlensing effect or even not microlensing at all. We should therefore examine the evidence for FFPs in this skeptical light.

The basic evidence for an FFP population is a "bump" in the Einstein-timescale distribution of events at short timescales ($t_{\rm E} \lesssim 1$ day) found by the MOA collaboration over 1.5 seasons. This invites three questions. First, could the bump be due to microlensing events generated by stars (or brown dwarfs), rather than FFPs? Second, could it actually be due to FFPs, but just a large statistical fluctuation from an intrinsically small signal? Third, could it be just a set of short timescale variations due to something other than microlensing, in particular, cataclysmic variables (CVs)?

The answer to the first questions is "no." There is a rather deep theorem (Mao and Paczyński 1996) that such a bump at short timescales can only be due to a separate population of low mass objects. I begin by writing down the Einstein timescale directly in terms of physical variables and then consider the limit of small $\pi_{\rm rel}$

$$t_{\rm E} = \frac{\sqrt{\kappa M \pi_{\rm rel}}}{\mu} \rightarrow K \frac{M^{1/2} D_{LS}^{1/2}}{v_{\rm rel}}. \tag{3.34}$$

where K is a constant and $v_{\rm rel} = D_L \mu$. Hence, there are exactly three ways to get a short microlensing event: low mass, small distances from lens to source, and fast relative lens-source velocity. We can immediately rule out the last as an explanation for the FFPs. FFP candidates have timescales of 1 day, compared to $t_{\rm E} \sim 20$ days for typical stellar events with $M \sim 0.3\, M_\odot$. Hence if this were to be the explanation, the stars would have to be moving about 20 times faster than the $v_{\rm rel} \sim 150\,{\rm km\,s^{-1}}$ that is typical of normal microlensing events. This is several times higher than the escape velocity of the Galaxy. Of course, faster-than-average $v_{\rm rel}$ could help explain the short timescale of a particular event, but the dominant explanation for each individual short event must be something else.

The first possibility (lower mass) is the hypothesis we are testing. Hence, the only other possibility is small D_{LS}. By an argument analogous to the one just given, these distances would have to be $\sim 20^2 = 400$ times smaller than the $D_{LS} \sim 1$ kpc that is typical of bulge lensing events. Of course, such small distances are possible, but they are rare (as I will show explicitly, below). But the main point is that there cannot be any *structure* imposed on the timescale distribution by such small D_{LS} events because the bulge does not have strong density structures on these length scales.

In fact, the smoothness of this density distribution leads directly to a simple power-law for the event timescale distribution at small $t_{\rm E}$. To demonstrate this I start with the generic event rate equation in the limit of small D_{LS}

$$\frac{d^3\Gamma}{dM\, dD_{LS}\, dv_{\rm rel}} = K' \sqrt{M D_{LS}}\, v_{\rm rel} f(v_{\rm rel}) g(M), \tag{3.35}$$

where $f(v_{\rm rel})$ is the distribution function of relative velocities, $g(M)$ is the mass function, and K' is another constant. Note that by not including a weighting function for D_{LS}, I am implicitly assuming that the density of lenses is uniform

over sufficiently small D_{LS}. This can be rewritten as,

$$\frac{d^3\Gamma}{dM\,dt_{\rm E}\,dv_{\rm rel}} = \frac{K'}{K^2}\sqrt{MD_{LS}}v_{\rm rel}^3 f(v_{\rm rel})\frac{g(M)}{M}t_{\rm E} = \frac{K'}{K^3}v_{\rm rel}^4 f(v_{\rm rel})\frac{g(M)}{M}t_{\rm E}^2. \tag{3.36}$$

After two integrations, this yields,

$$\frac{d\Gamma}{d\ln t_{\rm E}} = \frac{K'}{K^3}\langle v_{\rm rel}^4\rangle\langle M^{-1}\rangle t_{\rm E}^3. \tag{3.37}$$

Equation (3.37) tells us two things. First, and most important, the short timescale distribution due to a compact (but otherwise arbitrary) mass distribution of lenses is a pure power law $d\Gamma/d\ln t_{\rm E} \propto t_{\rm E}^3$, and therefore cannot have any "bumps." Hence, if there are bumps, this must be due to structure in the mass distribution, i.e., a separate population at low mass. (In principle, a bump could be imposed on the observed distribution if there were a bump on 1-day timescales in the detection efficiency, since the observed rate is the product of the underlying rate and the detection efficiency. However, there is no such detection-efficiency bump.) Second, it tells us that the coefficient is, as one would expect, heavily weighted by the high-end tail of the velocity distribution and somewhat by the low-end tail of the mass distribution.

Hence, we move to the second question: could the bump be due to a large statistical fluctuation. Since the bump is comprised of about 10 excess of events, it might at first appear as though such a fluctuation is plausible. The problem is that the number of events expected at the timescale of the shortest one that is detected is only about 0.01, and at the second shortest, just slightly larger. Thus, the argument against this possibility is essentially the same as the argument in favor of cold Neptunes being common. That is, if there were only one detection at low (e.g., 1 %) probability, it could be chalked up to a 1 % random occurrence. While these are quite rare (1 %) as results from properly posed experiments, they are extremely common if one is just trolling about nature looking for weird things. However, with two such detections, this possibility is reduced by another factor $\sim$100, and so one must take the result much more seriously.

This argument then interpenetrates with the third question: could these be intrinsic variables, in particular CVs? Of course, Sumi et al. (2011) were very familiar with the problem of CVs masquerading as microlensing events, and the MOA group does occasionally issue microlensing alerts for events that eventually turn out to be CVs. However, first, this very experience led them to develop criteria that efficiently removed CVs. Second, and more important, one of the very short events is high-mag and is certainly microlensing at 99.9999 % confidence. And the other has a high quality lightcurve that is extremely likely to be microlensing (although not with quite the same confidence). Hence, even if some of the other events turn out to be CVs (or some other type of previously unrecognized variable), the basic conclusion, which rests on the shortest events, remains valid.

Still, it remains the case that the basic result is so puzzling that it really must be checked by additional data. The MOA collaboration itself has subsequently acquired roughly five times more data than were used in the original data. The OGLE collaboration has acquired roughly three times more data, and of a dramatically higher quality. Hence, it will be interesting to see the analysis of these much larger data sets.

3.8.3 Solar-Like Systems May Be Common

The chaotic nature of microlensing searches that characterized the early days would seem to preclude systematic statistical analysis. For example, I happened to be awakened at 3 a.m. local time by the "ping" of my email that turned out to be a message from Andrzej Udalski saying that OGLE-2005-BLG-169 was at extremely high magnification based on a single point that he had decided to take while he was service observing at the OGLE telescope to support a Chilean program (which receive 10 % of all time from telescopes in Chile). He said he could not carry out intensive observations due to his commitment to service observations. I realized that the event could be observed at Kitt Peak (Arizona) for a few hours and called the MDM observer there, who happened to be OSU grad student Deokkeun An. Deokkeun happened to have a high-speed photometer mounted on the telescope and also happened to be basically done with his own program. Hence, he was able to get more than 1000 observations of this target, which is what led to the robust detection of this "cold Neptune."

How can such a concatenation of contingent possibilities be modeled to produce an "objective experiment"?

Obviously it cannot be. Nevertheless, Gould et al. (2010) realized that it was unnecessary to do so. However the events were selected (whether modelable or not), they could be concatenated in an "objective sample" provided that the observations did not depend on the known presence of absence of planets. Because high-mag events are in general observed based on their potential sensitivity to planets, rather than the recognized presence of planets, they potentially fit the bill. Gould et al. demonstrated (using the cumulative distribution of peak magnifications $A_{\rm max}$) that this was actually the case for events with $A_{\rm max} > 200$, of which there were 13 events in the period 2004–2008. These contained a total of six ice-giant and gas-giant planets, and so led to the first statistical estimate of such planets beyond the snow line.

However, what I want to focus on here is not the general rate but the fact that two of these six planets were in the same system: OGLE-2006-BLG-109 (Gaudi et al. 2008; Bennett et al. 2010). See Figure 3.8. This system was almost a complete doppelganger for the Solar system (in the first approximation that the Solar System consists of the Sun, Jupiter, and Saturn—which is 99.99 % correct by mass!). The two planets have almost exactly the same mass ratio to each other and to their host as Jupiter, Saturn, and the Sun. Moreover, they have almost the same projected-

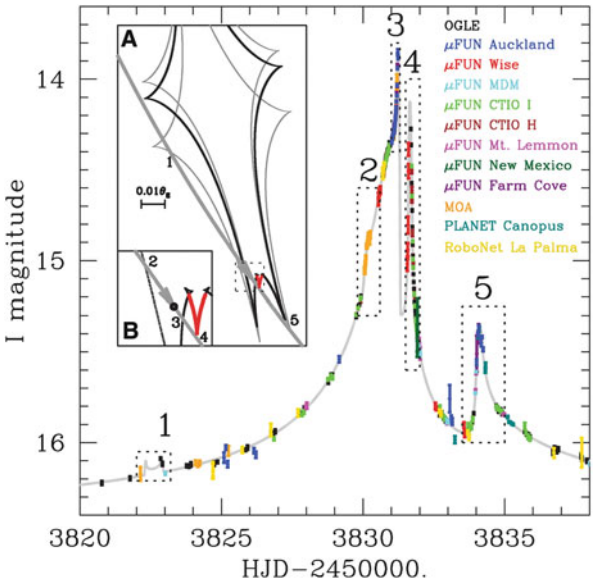

Fig. 3.8 Light curve and model for the event OGLE-2006-BLG-109. Two planets similar to Jupiter and Saturn are causing the perturbations in scaled-down version of our Solar system. From Science, 319, 927 (2008). Reprinted with permission from AAAS

separation ratio as the ratio of Jupiter-to-Saturn semi-major axis, And, while the actual values of these separations are smaller, they are smaller directly in proportion to the smaller mass of their host (as one would expect if the formation process were governed by the snow-line, which is believed to scale directly with stellar mass).

Based on this one detection, Gould et al. (2010) concluded that such solar-like systems were about 1/6 less common than in the solar system (i.e., occurred in 1/6 of all stars). With just a single detection, the error bar on this measurement is quite large, as I have just emphasized above when describing the two other exciting results listed above that were initially derived from two (rather than one) detections.

One may hope that with the advent of new microlensing surveys that are far more capable of detecting planets than the chaotic approach underlying the Gould et al. (2010) result, this result could be confirmed or contradicted based on a much larger sample. In fact, progress on this front is likely to be slow.

High-mag events are especially sensitive to multi-planet systems for a very simple reason first recognized by Gaudi et al. (1998): high-mag events detect planets primarily because they probe the central caustic induced by the planet on the magnification field of the host star. Hence, each planet can give rise to a perturbation in the same part of the light curve. Since high-mag events are densely monitored over peak for exactly this reason, there is a good chance to detect multiple planets. By contrast, while the chance of detecting individual planets via their planetary caustics is much larger for randomly sampled microlensing light curves (simply because the caustics are bigger), the chance of detecting two planets in this fashion is lower by the square. These two effects (rarity of high-mag events/central caustics vs. squared probability of low-mag events/planetary caustics) roughly cancel (see Zhu et al. 2014), but there is another factor at work. The higher-cadence surveys

that have been inaugurated in the last 5 years, which are capable of finding planets without follow-up observations, really only cover about 10–15 square degrees at high cadence. By contrast, the high-mag sample of Gould et al. (2010) was drawn from roughly 100 square degrees, from which high-mag candidates were identified for intensive follow-up based on low-cadence data. However, as high-cadence surveys have taken hold, which should enable the detection of several dozen planets per year, interest in the very labor intensive work required for high-mag follow-up has waned. Thus, a more precise measurement of the frequency of solar-like systems will likely have to wait several more years.

3.8.4 Terrestrial Planets in Low-Mass Binaries Are Common

In early 2013, one of the ∼2000 microlensing events that OGLE would find that year showed a slight dip (just 0.25 mag) while it was still quite faint, $I > 17.5$. Because of the high quality of OGLE data, there could be no doubt that this dip was real. To microlensers, this short (<1 day) dip could mean only one thing: a planet inside the Einstein ring, with the source headed directly toward the host. Why? As discussed in Sect. 3.4, the magnification pattern of a single lens has two images, with the one on the opposite side of the lens being at a saddle point on the time-delay surface and therefore easily (mostly) destroyed if perturbed by a planet at the position of this image. If the image had been totally annihilated (and the event were completely unblended), then the magnification would drop by a factor $\delta A/A = A_-/A = (1 - A^{-1})/2$. In fact, the annihilation is not perfect and moreover, in this case the event was blended, so this formula cannot be directly applied to the 0.25 mag drop to make a quantitative statement about the geometry. However, qualitatively, this is still the only way to make a short dip in a microlensing event.

How do we know that the source was headed directly for the host? Recall from Sect. 3.4 that the dip in the magnification pattern is flanked by two triangular caustics with one edge of each caustic facing the other. Extending away from the two cusps of each of these edges is a ridge of excess magnification. Therefore, if the source had passed over the planetary position at significant angle relative to the planet-host axis, it would have experienced bumps (small or large) before and/or after the dip. However, no such bumps were seen in either OGLE data, nor in MOA data, which although of lower quality significantly aided in the initial interpretation of this bump.

This realization led the Microlensing Follow Up Network (μFUN) to focus efforts to obtain dense continuous observations of the event as it approached its peak in the hopes of probing the central caustic and thereby learning more about the planet. These dense observations revealed something quite surprising: a huge caustic that could not possibly be due to the planet. Rather, there must be a third body in the system that was of comparable mass to the host.

Of course, this caustic (whose full crossing lasted about a day) would have been noticed even without μFUN's dense monitoring. However, detailed modeling

of these dense observations by Cheongho Han revealed something else that was extremely surprising. When the data from the "dip" were completely removed (i.e., several days on each side), the planet was completely recovered (i.e., both its mass and position) just from the distortions it induced on the central caustic generated by the binary. Hence, the planet could have been detected and characterized even if the source had missed the tiny planetary anomaly (provided that the dense observations of the binary-caustic feature had been taken).

Such central caustics are subject to the standard degeneracy between wide and close binaries, so that initially it was not known whether this was a circumbinary planet (close solution) or a circumsecondary planet (wide solution). Fortuitously, however, the wide solution "predicted" that the source should have passed relatively close to the companion about a year earlier. Nominally, such a passage should have produced a bump that OGLE could have alerted, but in fact the amplitude of this bump was only a few percent and so well below OGLE's 0.06 mag threshold. In addition, the individual data points had scatter that was much larger than the inferred bump. However, the bump is clearly visible in binned data (see Fig. 3.9, taken from Gould et al. 2014).

Finally, for this case, it proved possible to measure both the mass and distance of the system. First, $\theta_{\rm E}$ was easily measured from the densely covered caustic crossings. Second, the microlens parallax was also easily measurable. The parallel component ($\pi_{\rm E,\parallel} \simeq \pi_{\rm E,East} \sim 0.7$) was very large and so would have been measurable from the asymmetry in the light curve even if this were a point-lens event (see Fig. 3 of Gould et al. 2014). However, because this event had multiple structures, the normally more difficult $\pi_{\rm E,\perp}$ was also easily measured (An and Gould 2001).

The net result is a very well-characterized system, with a roughly $2\,M_\oplus$ planet orbiting at ~ 1 AU from a very late M dwarf ($M \sim 0.13\,M_\odot$) that is itself orbiting a slightly heavier companions at ~ 15 AU. The whole system lies just 1 kpc from Earth.

Gould et al. (2014) then performed what at the time was only an illustrative calculation. They asked how many such systems (terrestrial planet orbiting one component of a low-mass binary) should they have detected if every single low-mass star were in one. The answer was: about one. Of course, as they noted, statistical inference from a single detection is extremely weak, but nevertheless this calculation formed an important benchmark for the next discovery.

This discovery came from an event that occurred just a few months later but proved more difficult to analyze, OGLE-2013-BLG-0723. As with OGLE-2013-BLG-0341, interest in this event was piqued by a very short "blip" on a rising, otherwise normal-seeming point-lens event. However, contrary to OGLE-2013-BLG-0341, this blip was initially uninterpretable. The event seemed to peak normally about 40 days later, but then suddenly erupted into a huge caustic crossing 10 days further on (Fig. 1, Udalski et al. 2015a).

Although the "blip" took the form of a short spike, and therefore is naively consistent with a major-image caustic (wide-planet), in fact detailed modeling shows that this geometry does not work. Rather, the spike is caused by passage

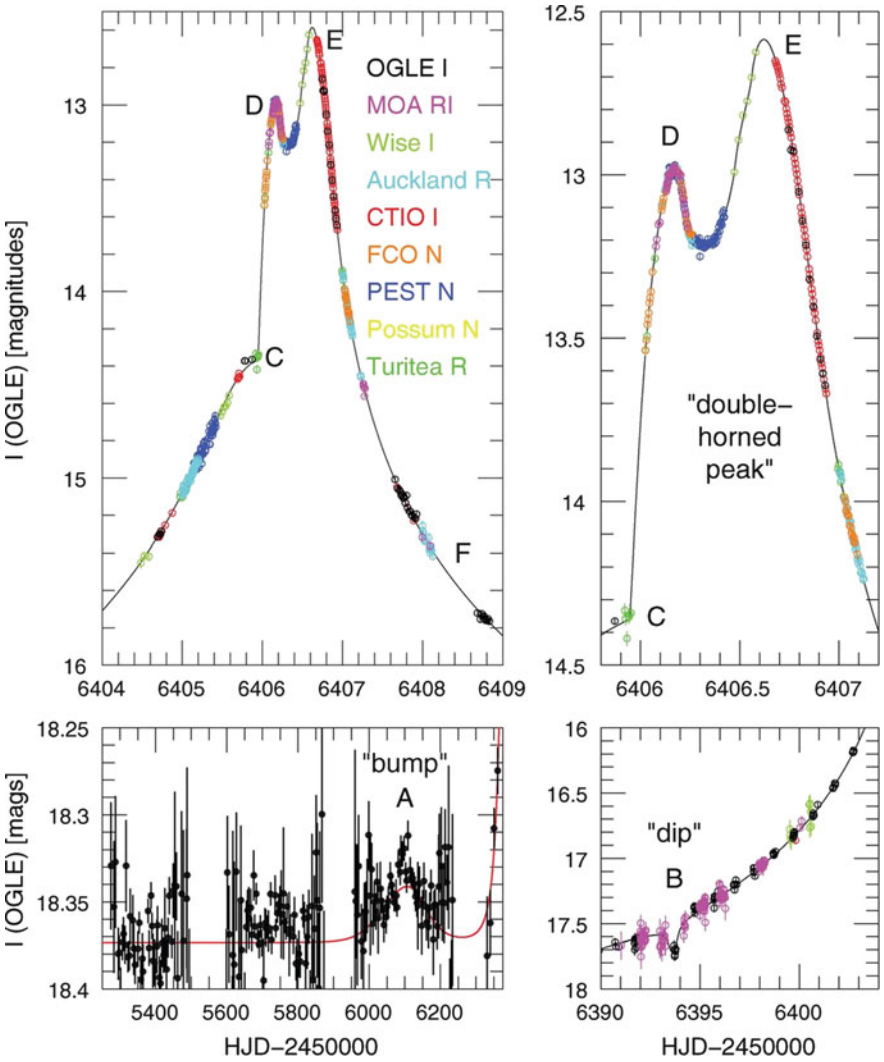

Fig. 3.9 Light curve and model for the event OGLE-2013-BLG-0341. The double horned perturbation is caused by a binary lens, but its shape is distorted by a third small body, also detected through the small perturbations along the rising part of the light curve. From Science, 345, 46 (2014). Reprinted with permission from AAAS

of the source by the outlying cusp of one of the triangular caustics of a close-planet. This is not at all obvious from the light curve, in particular because the location of these triangular caustics relative to the position of the host is displaced by the presence of the binary companion. The difficulty of solving this event is a reminder that not all microlensing solutions can yet be derived by computer searches: some still require human intuition.

As in the case of OGLE-2013-BLG-0341 there is a degeneracy between the close binary and wide binary solutions, so that initially both circumbinary and circumsecondary solutions appear viable. And, as in the case of OGLE-2013-BLG-0341, this degeneracy is resolved in favor of the wide binary solution, but by a very different argument. The wide binary solution is consistent with a static binary, which in turn is consistent with the small level of orbital motion expected from a well-separated low-mass binary, while the close binary solution requires significant orbital motion, which is also what is expected. Thus, at first sight, the orbital motions of each solution seems equally consistent with its own hypothesis. In fact, however, this apparent symmetry derives from a deep asymmetry buried in the lens equation.

In Sect. 3.7, I discussed the close–wide degeneracy mostly at first order, but noted that An (2005) analyzed this degeneracy to second order. This further analysis was motivated by the fact that the caustics for the two solutions (wide and close) seemed "rotated" with respect to each other (see e.g., Fig. 8 of Afonso et al. 2000), although the models themselves were each static. As discussed by Udalski et al. (2015a), it is therefore natural that allowing for orbital motion (including actively rotating caustics) provides an additional degree of freedom to accommodate this degeneracy. This degree of freedom is required in the case of OGLE-2015-BLG-0723 because the static close binary solution is strongly ruled out.

Hence, we can imagine two cases, one in which the static wide binary model is correct and is mimicked by some close binary model with an arbitrary degree of orbital motion to match the observed lightcurve, and the other in which a close binary model with finite orbital motion is correct and is mimicked by some wide binary model with an arbitrary degree of orbital motion to match the observed lightcurve. However, in the latter case, one would hardly expect the "fine-tuned" mimicking model to be consistent with a static model (which is what is predicted by the physics of a wide model). Formally the probability of such fine tuning is $\exp(-\Delta\chi^2/2) \sim 10^{-300}$. Another argument against the close solution (which is formally much weaker, and hence secondary) is that the projected energy ratio of the close solution is $\beta = 0.023$, which is far below the expected range. See discussion following Eq. (3.24). Such an arbitrary β is just what one would expect from an ad hoc solution that mimics the true solution but would be exceptional for a real binary.

As with OGLE-2013-BLG-0341, OGLE-2013-BLG-0723 has easily measured finite source effects that enable measurement of $\theta_{\rm E}$ and a large $\pi_{\rm E}$ whose measurement is assisted by the presence of multiple features (An and Gould 2001). Therefore, the three lens masses and the system distance are all well measured. As for OGLE-2013-BLG-0341, both binary components are low mass, the planet is terrestrial ($0.7\,M_\odot$ in this case) and separated from its host by less than 1 AU. Note, however, that in this case, the host is a brown dwarf, $M_{\rm host} \sim 0.03\,M_\odot$.

As noted by Udalski et al. (2015a), the geometric probabilities of discovering this system are similar to that of OGLE-2013-BLG-0341 but the source star is somewhat fainter, which roughly doubles the expected number of detections (assuming all lenses were part of such systems). That is, two are expected and two detected (under this universalist assumption). Hence, whereas not much could be concluded about the frequency of these systems from the single detection of OGLE-2013-BLG-0341,

the addition of a second system does argue strongly that this type of system is very common.

This startling nature of this conclusion, together with the even more fundamental conclusion derived in the next section from the same discovery, makes it remarkable that as of this date (October 2015), not a single author had yet cited this paper after being on arXiv for 3 months.

3.8.5 Deep Unity of Star-Planet and Planet-Moon Formation

The main point made by Udalski et al. (2015a) about their discovery was that it implied a deep unity of star-planet and planet-moon formation by providing a missing link between the two. That is, when we compare the planet-host mass ratio of the basically ice-and-rock planet Uranus and of the ice-and-rock moon Ganymede, we find very similar values: $(4.4, 5.7) \times 10^{-5}$. And when we compare their semi-major axes (normalized to host mass), we also find very similar values $(19, 13)\,\mathrm{AU}\,M_\odot^{-1}$. A plausible explanation is that the "snow line" scales with host mass over the 3 orders of magnitude between the masses of host Sun and host Jupiter, and that at a factor few beyond the snow line the ratio of companion to host mass is a constant, i.e., is set by a scale free process.

The OGLE-2013-BLG-0723 system lends support to this conjecture. Comparing the three systems with host masses $\log(M/M_\odot) = (0, -1.5, -3)$, we have $(M_{\rm comp}/M_{\rm host}) = (4.4, 6.6, 5.7) \times 10^{-5}$ and $(a/M_{\rm host}) = (19, 11, 13)\,\mathrm{AU}\,M_\odot^{-1}$. Again, I remark that not a single author has commented on this startling result. To quote Fox News: "I report, you decide."

3.9 Non-Planetary Microlensing

While the main focus of these notes is planetary microlensing, it is important to realize that the same microlensing observations, and mostly the same types of microlensing analysis, can lead to important results about objects other than planets.

The most easily accessible such results concern binaries. This is because binaries can have large caustics. These first of all have large cross sections, leading to a high level of detection (relative to their underlying frequency). Of course caustic crossings also imply the opportunity for measurement of $\theta_{\rm E}$, which is generally not available for single-lens events. And second, the sharp features caused by these caustics enhance the probability of parallax measurements. We have already seen these factors at work in Sect. 3.8.4, where they led to the detailed characterization of two extremely interesting binary+planet systems. Such parallax measurements are actually essential for binary science because this is the only way to distinguish binaries whose components are in the mass range of ordinary stars (and so, which are much better studied by other techniques) from those that are not, such as brown dwarfs, neutron stars, and black holes.

One major set of results is the discovery of a new class of tight, low-mass, brown-dwarf binaries (Choi et al. 2013). These obey the same binding minimum energy relation found for brown-dwarf binaries at higher masses (which seems to be separated by a factor 10 from a similar relation for stellar binaries). Han et al. (2013), Jung et al. (2013b), and Park et al. (2015) subsequently made closely related discoveries. For the case of brown-dwarf binaries (or binaries composed of brown dwarfs and very low-mass stars) the measurability of the (usually difficult) parallax is automatically enhanced by the low mass, since $\pi_{\rm E} = \sqrt{\pi_{\rm rel}/\kappa M}$, but in all of the above cases $D_L \lesssim 3\,{\rm kpc}$, which means it was further enhanced by large $\pi_{\rm rel}$. In fact, all but one of these systems were $D_L < 2\,{\rm kpc}$. Because the majority of lenses are believed to lie in the Galactic bulge, and the great majority of the rest at $D_L > 2\,{\rm kpc}$, the relatively high number of interesting systems at short distances probably means that we are seeing only the tip of the iceberg due to the heavy bias of being able to characterize nearby systems. I will return to this question when I discuss future prospects.

At the other end of the spectrum are binaries with high-mass components, i.e., black-hole (BH) or neutron-star (NS) remnants. By the same formula $\pi_{\rm E} = \sqrt{\pi_{\rm rel}/\kappa M}$ these generally have small parallaxes, which naively may lead to the conclusion that they are difficult to measure. In fact the situation is mixed. The large mass leads to large Einstein radius $\theta_{\rm E} = \sqrt{\kappa M \pi_{\rm rel}}$, particularly for nearby lenses, and hence generically longer timescales, $t_{\rm E} = \theta_{\rm E}/\mu$. Recall that the most difficult aspect of measuring parallax from the ground is $\pi_{{\rm E},\perp}$ because it is fourth-order in time. Hence, historically, BHs have been regarded as the most favorable candidates for microlens parallax measurements. For example, Poindexter et al. (2005) evaluated three previously discovered BH candidates and found that they had, respectively, low, medium, and high likelihood to actually be BHs.

The "problem" with these candidates is that they are all isolated lenses and therefore do not have caustic crossings and so do not have measurements of $\theta_{\rm E}$. Thus, they remain "candidates" rather than detections. This would be the great value of BHs in binaries: they could yield caustic crossings.

Actually Dong et al. (2007) did measure the parallax of a candidate BH binary using *Spitzer* as a "parallax satellite." Unfortunately, this did not happen to be a caustic-crossing binary. Hence, there was no $\theta_{\rm E}$ measurement and so no confirmation of a BH.

The one major idea for measuring $\theta_{\rm E}$ for point-lens events (particularly BH events with their potentially large $\theta_{\rm E}$) is astrometric microlensing (Walker 1995; Hog et al. 1995; Miyamoto and Yoshii 1995). Although the two images of a point-lens event are separated by just $(u_+ - u_-)\theta_{\rm E} = \sqrt{1 + u^2/4}\theta_{\rm E}$, and so are not generally separately resolvable, the centroid of the images is displaced from the source by almost the same order, and measuring this displacement does not require resolving the images separately. This displacement is given by

$$\Delta\theta = \left(\frac{A_+ u_+ + A_- u_-}{A} - u\right)\theta_{\rm E} = \frac{u}{u^2+2}\theta_{\rm E}. \tag{3.38}$$

This displacement reaches its maximum $\Delta\theta = \theta_E/\sqrt{8}$ at $u = \sqrt{2}$, i.e., more than 1/3 of an Einstein radius. Several groups are attempting to apply this method using either *Hubble Space Telescope (HST)* or Keck adaptive optics (AO) observations, but there are no published results as of yet.

3.10 Future of Microlensing Planet Searches

Microlens planet searches are currently in a highly dynamic state. In 2007 and 2011, the MOA and OGLE collaborations had their first full years of their upgraded systems, with cameras covering, respectively, 2.2 and 1.4 square degrees, with, respectively, 100 and 280 Mpxls. This has enabled regular detection of planets in survey-only mode, either from each survey separately or by combining them. Combination is especially helpful since the locations of these telescopes are respectively in New Zealand and Chile.

3.10.1 KMTNet

The Korea Microlensing Telescope Network (KMTNet) inaugurated all three of its new telescopes in 2015 and will have its first full year of operations in 2016. Each camera (located in Chile, South Africa, and Australia) covers 4 square degrees with 340 Mpxls. This will greatly augment the power of survey-only detection. For example, KMTNet covers 16 square degrees with a 10 min cadence 24 h per day (weather permitting) for about 2 months per year, and has nearly continuous coverage for another 3 months per year. In addition, it covers a much larger area several times per day from each site. The high-cadence data can automatically cover caustic crossings (which typically last about 1 h) without the need for follow-up alerts, and are sensitive to FFPs down to Neptune (and perhaps Earth) masses. The entire setup permits a purely objective statistical survey, since no follow-up observations are required either to detect or characterize planets. This is a major new capability of microlensing planet searches.

3.10.2 Space-Based Parallaxes

While KMTNet has been in planning stages for 7 years, another new microlensing capability has erupted rather suddenly: space-based microlens parallaxes. The background story of this development demonstrates the circuitous route by which major initiatives sometimes emerge. In 2007, I was invited to a "Warm *Spitzer*" meeting to present ideas for *Spitzer* microlens parallaxes. I was excited by this

general possibility, even though my detailed investigation seemed to show that *Spitzer* capability was limited. The main problem was that *Spitzer* can only view a given portion of the sky over a range of Sun-angles $82.5° < \theta_\odot < 120°$. For targets near the ecliptic (like the prime microlensing fields), this implies two roughly 38-day observing windows (only one of which is simultaneous with the Earth viewing window). Since at the time I believed that the satellite must capture (or nearly capture) the peak of the event to reliably measure $u_{0,\rm sat}$ and (with still greater difficulty) measure $t_{\rm E,sat}$ in order to break the fourfold degeneracy (e.g., Gould 1995; Gaudi and Gould 1997), the 38-day window seemed quite short. Moreover, at this time, the fact that the *Spitzer* band at 3.6 μm was well outside the range of ground-based bands appeared to present severe problems for the interpretation (e.g., Gould 1995, 1999).

In any case, the target-of-opportunity (ToO) requirements to support such observations were judged too difficult for the reduced operations of Warm *Spitzer*, and proposals of this type were excluded from consideration. Hence, the actual path to space-based microlens parallaxes proved quite different.

3.10.3 Kepler Microlensing: (MP)^3

In 2013, a second *Kepler* reaction wheel broke, making continued observations in its standard mode impossible. Even before Ball Aerospace engineers figured out how to stabilize the spacecraft by pointing it near the plane of its orbit, Gould and Horne (2013) produced a plan to dedicate the partially crippled *Kepler* to microlens parallaxes.

Our initial conception was that *Kepler* would observe microlensing events in a manner similar to all previous ideas for microlens parallax satellites, and nominally similar to its standard mode of transit observations, i.e., choosing target stars from ongoing microlensing events discovered from the ground that lay in its large ($>100\,{\rm deg}^2$) field. However, as we worked to make this idea a practical reality, we soon discovered that *Kepler* could not communicate with the ground often enough to trigger on individual events. Hence was born a completely new idea of "Multiplexing for Massive Production of Microlens Parallaxes" (MP)^3. Instead of targeting individual events, a large contiguous area would be monitored. Of course, since *Kepler* cannot download its entire detector, this contiguous field is much smaller than its full field, probably only about $4\,{\rm deg}^2$. However, by chance, because the densest microlensing fields and the *Kepler* orbit are both near the ecliptic, the *Kepler*-accessible area is one of the most productive. NASA eventually approved this strategy as "Campaign 9" of the K2 mission. Observations are slated for April–June 2016.

The unique feature of the K2C9 is that it permits exploration of FFPs. Consider a typical Jupiter-mass planet in the Galactic Disk or Bulge, with $\pi_{\rm rel} = 0.12\,{\rm mas}$ or $\pi_{\rm rel} = 0.02\,{\rm mas}$, respectively. These will have microlens parallaxes $\pi_{\rm E} \sim 4$ and $\pi_{\rm E} \sim 1.6$, respectively. Hence, they will look quite different (or perhaps not give

rise to any perceptible signal at all) to a satellite at projected separation of 0.1–0.6 AU, which is the range of *Kepler*'s $D_\perp$ during K2C9. See Eq. (3.18). Similarly, FFP events discovered by K2 will look very different (or imperceptible) from Earth. Probably the main information about FFPs from this mission will come from the latter class of events because the ground-based observations will be at much higher resolution and cadence, and so are better able to detect the very weak signals from events that are first recognized from their strong K2 signal. However, even non-detections will be important since they will place a lower limit on $\pi_{\rm E}$. This will first of all prove that these are FFPs, since stellar events whose short $t_{\rm E}$ is explained by fortuitously small D_{LS} will have very small $\pi_{\rm E} \propto \sqrt{D_{LS}/M}$, and of course CVs will look the same to any observatory regardless of location. Moreover, even lower limits on $\pi_{\rm E}$ will provide crucial quantitative information.

Finally K2C9 will yield many other microlensing events as well, but since these are broadly similar in nature to those being discovered now by *Spitzer*, they are better discussed in that context.

3.10.4 Spitzer Microlensing

As mentioned above, the Gould and Horne (2013) paper led immediately to practical efforts to organize *Kepler* microlensing observations. Even before it appeared that these might be successful, they triggered the idea of trying again for a *Spitzer* microlensing campaign. I proposed (with Jennifer Yee and Sean Carey) to apply all of 2014 *Spitzer* Bulge window to microlensing observations. The organizational challenges engendered by triggering a massive number of ToOs remained essentially the same as in 2007, and of course it was unknown at this time whether such observations could be successful due to both recognized problems (see above) and perhaps unrecognized problems. The Director therefore approved a pilot program of 100 h of observations during the 38-day Bulge window to determine the feasibility of such a program.

An important technical issue was minimizing the strain on *Spitzer* operations of this (pared down but still huge) avalanche of ToOs. This was handled by uploading to *Spitzer* several weeks in advance a set of dummy observations at preset times and arbitrary location (but near the microlensing fields). Then each Monday at UT 15:00, the *Spitzer* microlens team would send a list of real observing targets to *Spitzer* operations, which would vet these in the usual way, and then upload them to the spacecraft for observations roughly Thursday through Wednesday. Hence, observations would begin from 3 to 10 days after they were first recognized as valuable targets. See Fig. 1 of Udalski et al. (2015b).

Clearly, this restriction makes FFP observations either extremely difficult or impossible. However, there is a wealth of other science that can be pursued. The most prominent objective is to measure the Galactic distribution of planets. That is, for planetary events, the combination of *Spitzer* parallaxes and finite source effects gives the lens distance, while for non-planetary events, the parallax alone gives good

statistical information on the distance (Calchi Novati et al. 2015a). By comparing the cumulative distribution of planetary detections (Udalski et al. 2015b; Street et al. 2015) to that of all events, one can therefore determine whether planets form more readily in the Galactic Disk or Bulge (Fig. 3 of Calchi Novati et al. 2015a). Actually one must compare the planet-sensitivities (Zhu et al. 2015b) rather than simply the count of the events.

Now, the process of choosing *Spitzer* targets can be just as chaotic as the one for choosing the high-mag events that led to the Gould et al. (2010) sample. In Sect. 3.8.3, I discussed how the very chaos of this process could lead to something that very well approximated a controlled experiment. A similar argument does actually apply to the 100-h 2014 *Spitzer* campaign because the sole aim of that campaign was to demonstrate the feasibility of *Spitzer* parallaxes, so the selection process was strictly carried out without reference to the presence or absence of planets. However, once detecting planets becomes a central objective (as did occur when we were awarded 832 h for the 2015 season), then it actually becomes quite difficult to "protect" the sample from biases either for or against observation by *Spitzer* due to the perceived presence or absence of planets. Yee et al. (2015c) carried out a systematic analysis of this problem and devised appropriate solutions. Based on (Calchi Novati et al. 2015a) successful derivation of flux constraints from $VI[3.6\,\mu\mathrm{m}]$ and $IH[3.6\,\mu\mathrm{m}]$ color-color diagrams, Yee et al. (2015c) also argued that it was possible to measure parallaxes even if *Spitzer* observations completely missed the peak, provided that at least some of the observations fell inside the Einstein ring (and of course that the microlensed source was bright enough for *Spitzer* photometry). This greatly expanded the number of events that could be observed by *Spitzer*.

The 2015 campaign showed that Yee et al. (2015c) were correct in this assessment. As a result, the *Spitzer* microlensing observations are returning a broad range of additional science, including the first strong candidate for a wide binary with massive remnant component (Shvartzvald et al. 2015), and the second isolated brown dwarf with a direct mass measurement (Zhu et al. 2015c). See Calchi Novati et al. (2015b) for *Spitzer* photometry of all events.

3.10.5 WFIRST Microlensing

WFIRST (Wide Field InfraRed Space Telescope) is a proposed mission that while not yet selected, is being heavily funded by NASA in anticipation that it will be selected. It was partly inspired by my own "White Paper" addressed to the 2010 Decadal Committee, entitled "Wide Field Imager in Space for Dark Energy and Planets" (Gould 2009). As I advocated, *WFIRST* will use the same wide field infrared imager to carry out cosmological and planetary microlensing studies, and will in addition carry out other wide field surveys (Spergel et al. 2013a,b). At present, the precise parameters of the microlensing survey are still under discussion,

but it will likely cover roughly 3 deg^2 toward the Galactic Bulge, once per 15 min, for six 72-day campaigns centered at quadrature.

Such a survey will robustly probe down to Earth-mass planets, including Earth-mass FFPs (e.g., Bennett and Rhie 2002). However, it will do many other things as well. For example, it provides the best hope of astrometric microlens mass measurements of BHs (Gould et al. 2014), will make precise asteroseismic measurements of about one million stars, and yield several hundred million parallax measurements that are an order of magnitude more precise than GAIA (Gould et al. 2015) as well as a wealth of information about thousands of ultra-faint Kuiper Belt Objects, including orbits, binarity, and density (Gould 2015). In fact, the potential applications of the *WFIRST* microlensing data set have barely been scratched.

3.11 Conclusions

Microlensing planet searches have grown over the last quarter century from a simple suggestion to a massive observational and theoretical undertaking. Their potential for the future is immense, with respect to their direct application to understanding planets and their indirect applications to many other fields. For a complementary perspective on many of the topics covered here, see Gaudi (2012).

Acknowledgement This work was supported by NSF grant AST 1103471.

References

Afonso, C., Alard, C., Albert, J.N., et al.: Astrophys. J. **532**, 340 (2000)
Albrow, M.D., Beaulieu, J.-P., Caldwell, J.A.R., et al.: Astrophys. J. **534**, 894 (2000)
An, J.H.: Mon. Not. R. Astron. Soc. **356**, 409 (2005)
An, J.H., Gould, A.: Astrophys. J. Lett, **556**, L113 (2001)
Batista, V., Dong, S., Gould, A., et al.: Astron. Astrophys. **508**, 467 (2009a)
Batista, V., Gould, A., Dieters, S.M., et al.: Astron. Astrophys. **529A**, 102 (2009b)
Batista, V., Beaulieu, J.-P., Bennett, D.P., et al.: Astrophys. J. **608**, 170 (2015)
Beaulieu, J.-P., Bennett, D.P., Fouqué, P., et al.: Nature **439**, 437 (2006)
Bennett, D.P.: Astrophys. J. **716**, 1408 (2010)
Bennett, D.P., Rhie, S.H.: Astrophys. J. **574**, 985 (2002)
Bennett, D.P., Rhie, S.H., Becker, A.C., et al.: Nature **402**, 57 (1999)
Bennett, D.P., Rhie, S.H., Nikolaev, S., et al.: Astrophys. J. **713**, 837 (2010)
Bennett, D.P., Bhattacharya, A., Anderson, J., et al.: Astrophys. J. **608**, 169 (2015)
Bennett, D.P., Sumi, T, Bond, I.A., et al.: Astrophys. J. **757**, 119 (2012)
Bensby, T., Yee, J.C., Feltzing, S., et al.: Astron. Astrophys. **549A**, 147 (2013)
Bessell, M.S., Brett, J.M.: Publ. Astron. Soc. Pac. **100**, 1134 (1988)
Bozza, V.: Astron. Astrophys. **355**, 423 (2000)
Bozza, V.: Mon. Not. R. Astron. Soc. **408**, 2188 (2010)
Calchi Novati, S., Gould, A., Udalski, A., et al.: Astrophys. J. **804**, 20 (2015a)
Calchi Novati, S., Gould, A., Yee, J.C., et al.: Astrophys. J. **814**, 92 (2015b)

Choi, J.-Y., Han, C., Udalski, A., et al.: Astrophys. J. **768**, 129 (2013)
Dominik, M.: Astron. Astrophys. Suppl. **109**, 507 (1995)
Dominik, M.: Astron. Astrophys. **333**, L79 (1998)
Dominik, M.: Astron. Astrophys. **349**, 108 (1999)
Dong, S., DePoy, D.L., Gaudi, B.S., et al.: Astrophys. J. **642**, 842 (2006)
Dong, S., Udalski, A., Gould, A., et al.: Astrophys. J. **664**, 862 (2007)
Dong, S., Gould, A., Udalski, A., et al.: Astrophys. J. **695**, 970 (2009)
Einstein, A.: Science **84**, 506 (1936)
Erdl, H., Schneider, P.: Astron. Astrophys. **268**, 453 (1993)
Fukui, A., Gould, A., Sumi, T., et al.: Astrophys. J. **809**, 74 (2015)
Gaudi, B.S.: ARAstron. Astrophys. **50**, 411 (2012)
Gaudi, B.S., Gould, A.: Astrophys. J. **477**, 152 (1997)
Gaudi, B.S., Naber, R.M., Sackett, P.D.: Astrophys. J. Lett. **502**, L33 (1998)
Gaudi, B.S., Bennett, D.P., Udalski, A., et al.: Science **319**, 927 (2008)
Ghosh, H., DePoy, D.L., Gal-Yam, A., et al.: Astrophys. J. **615**, 450 (2004)
Gould, A.: Astrophys. J. Lett. **421**, L75 (1994)
Gould, A.: Astrophys. J. Lett. **441**, L21 (1995)
Gould, A.: Astrophys. J. **514**, 869 (1999)
Gould, A.: Astrophys. J. **606**, 319 (2004)
Gould, A.: Astrophys. J. **681**, 1593 (2008)
Gould, A.: Astro2010: the Astronomy and Astrophysics Decadal Survey, Science White Papers, No. 100 (2009). arXiv:0902.2211
Gould, A.: J. Korean Astron. Soc. **47**, 297 (2015)
Gould, A., Gaucherel, C.: Astrophys. J. **477**, 580 (1997)
Gould, A., Horne, K.: Astrophys. J. **779**, 28 (2013)
Gould, A., Loeb, A.: Astrophys. J. **396**, 104 (1992)
Gould, A., Yee, J.C.: Astrophys. J. **764**, 107 (2013)
Gould, A., Yee, J.C.: Astrophys. J. **784**, 64 (2014)
Gould, A., Udalski, A., Monard, B., et al.: Astrophys. J. **698**, L147 (2000)
Gould, A., Udalski, A., An, D., et al.: Astrophys. J. **644**, L37 (2006)
Gould, A., Dong, S., Gaudi, B.S., et al.: Astrophys. J. **720**, 1073 (2010)
Gould, A., Udalski, A., Shin I.-G., et al.: Science **345**, 46 (2014)
Gould, A., Huber, D., Penny, M., Stello, D.: J. Korean Astron. Soc. **48**, 93 (2015)
Graff, D.S., Gould, A.: Astrophys. J. **580**, 253 (2002)
Griest, K., Safizadeh, N.: Astrophys. J. **500**, 37 (1998)
Han, C., Jung, Y.K., Udalski, A., et al.: Astrophys. J. **778**, 338 (2013)
Hog, E., Novikov, I.D., Polanarev, A.G.: Astron. Astrophys. **294**, 287 (1995)
Jung, Y.K., Han, C., Gould, A., Maoz, D.: Astrophys. J. **768**, L71 (2013a)
Jung, Y.K., Udalski, A., Sumi, T., et al.: Astrophys. J. **798**, 123 (2013b)
Kayser, N., Refsdal, S., Stabell, R.: Astron. Astrophys. **166**, 36 (1986)
Kervella, P., Thévenin, F., Di Folco, E., Ségransan, D.: Astron. Astrophys. **426**, 297 (2004)
Liebes, S.: Phys. Rev. **133**, 835 (1964)
Mao, S., Paczyński, B.: Astrophys. J. **374**, L37 (1991)
Mao, S., Paczyński, B.: Astrophys. J. **473**, 57 (1996)
Miyamoto, M., Yoshii, Y.: Astron. J. **110**, 1427 (1995)
Muraki, Y., Han, C., Bennett, D.P., et al.: Astrophys. J. **741**, 22 (2011)
Nataf, D.M., Gould, A., Fouqué, P., et al.: Astrophys. J. **769**, 88 (2013)
Paczyński, B.: Astrophys. J. **304**, 1 (1986)
Park B.-G., DePoy, D.L., Gaudi, B.S., et al.: Astrophys. J. **609**, 166 (2004)
Park, H., Udalski, A., Han, C., et al.: Astrophys. J. **805**, 117 (2015)
Pejcha, O., Heyrovský, D.: Astrophys. J. **690**, 1772 (2009)
Poindexter, S., Afonso, C., Bennett, D.P., Glicenstein, J.-F., Gould, A, Szymański, M.K., Udalski, A.: Astrophys. J. **633**, 914 (2005)
Poleski, R., Skowron, J, Udalski, A., et al.: Astrophys. J. **795**, 42 (2014)

Refsdal, S.: Mon. Not. R. Astron. Soc. **134**, 315 (1966)
Renn, J., Sauer, T., Stachel, J.: Science **275**, 184 (1997)
Shin, I.-G., Udalski, A., Han, C., et al.: ApJ **735**, 85 (2011)
Shin, I.-G., Han, C., Choi, J.-Y., et al.: ApJ **755**, 91 (2012)
Shvartzvald, Y., Udalski, A., Gould, A., et al.: Astrophys. J. **814**, 111 (2015)
Skowron, J., Udalski, A., Gould, A., et al.: Astrophys. J. **738**, 87 (2011)
Smith, M., Mao, S., Paczyński, B.: Mon. Not. R. Astron. Soc. **339**, 925 (2003)
Spergel, D., Gehrels, N., Breckinridge, J., et al.: Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA Final Report (2013a). arXiv:1305.5422
Spergel, D., Gehrels, N., Breckinridge, J., et al.: WFIRST-2.4: What Every Astronomer Should Know (2013b). arXiv:1305.5425
Street, R., Udalski, A., Calchi Novati, S., et al.: Astrophys. J. (2015, submitted). arXiv:1508.07027
Sumi, T., Kamiya, K, Bennett, D.P., et al.: Nature **473**, 349 (2011)
Sumi, T., Udalski, A., Bennett, D.P., et al.: (2015, submitted) arXiv:1512.00134
Udalski, A., Jung, Y.K., Han, C., et al.: Astrophys. J. **812**, 47 (2015a)
Udalski, A., Yee, J.C., Gould, A., et al.: Astrophys. J. **799**, 237 (2015b)
Walker, M.A.: Astrophys. J. **453**, 37 (1995)
Yee, J.C., Udalski, A., Sumi, T., et al.: Astrophys. J. **703**, 2082 (2009)
Yee, J.C., Udalski, A., Calchi Novati, S., et al.: Astrophys. J. **802**, 76 (2015a)
Yee, J.C., Johnson, J.A., Skowron, J., et al.: Astrophys. J. (2015b, submitted). arXiv:1506.01441
Yee, J.C., Gould, A., Beichman, C., et al.: Astrophys. J. **810**, 155 (2015c)
Yoo, J., DePoy, D.L., Gal-Yam, A., et al.: Astrophys. J. **603**, 139 (2004)
Zhu, W., Gould, A., Penny, M., Mao, S., Gendron, R.: Astrophys. J. **794**, 53 (2014)
Zhu, W., Udalski, A., Gould, A., et al.: Astrophys. J. **805**, 8 (2015a)
Zhu, W., Gould, A., Beichman, C., et al.: Astrophys. J. **814**, 129 (2015b)
Zhu, W., Calchi Novati, S., Gould, A., et al.: Astrophys. J. (2015c, submitted). arXiv:151004787

Part IV
The Direct Imaging Method

Chapter 4
Direct Imaging of Faint Companions

Riccardo Claudi

Abstract The exoplanets around stars in the solar neighborhood are expected to be bright enough for us to characterize them with direct imaging; however, they are much fainter than their parent stars, and separated by very small angles, so conventional imaging techniques are totally inadequate, and new methods are needed. The direct imaging of exoplanets is extremely challenging. Jupiter is 10^9 times fainter than our Sun in reflected visible light. A direct imaging instrument for exoplanets must suppress (1) the bright star image and diffraction pattern and (2) the stellar scattered light from imperfections in the telescope. The main goal of high-contrast imaging is primarily to discover and characterize extrasolar planetary systems. High-contrast observations, in optical and infrared astronomy, are defined as any observation requiring a technique to reveal a low mass companion that is so close to the primary, brighter by a factor of at least 10^5, that optical effects hinder or prevent the collection of photons directly from the target of observation. To overcome this, astronomers combined large telescopes (to reduce the impact of diffraction), adaptive optics (to correct for phase errors induced by atmospheric turbulence), and sophisticated image processing.

4.1 Introduction

The most challenging task for a planet hunter is to observe directly very faint companions (planets or brown dwarfs) orbiting a bright star, not only because of the small angular separation between planets and stars (e.g., <500 mas for a 5 au orbital radius at 10 pc), but also and mostly because the contrast between a planet and its host star ranges from 10^{-3} for hot giant planets in the infrared to 10^{-10} for Earth-like planets in the visible. Nevertheless, direct imaging is a very promising method as it provides a straightforward means to characterize planetary atmospheres with spectro-photometric measurements.

Direct imaging means the direct detection of the radiation from the planet and/or disk, picked out from under the glare of the parent star, allowing to take a snapshot

R. Claudi (✉)
INAF – Osservatorio Astronomico di Padova, Vicolo Osservatorio, 5, 35122 Padova, Italy
e-mail: riccardo.claudi@inaf.it

V. Bozza et al. (eds.), *Methods of Detecting Exoplanets*, Astrophysics and Space Science Library 428, DOI 10.1007/978-3-319-27458-4_4

of the whole system. In this way, it is possible to analyze the light reflected by the planet (visible) and also the intrinsic distribution of radiation emitted by the planet itself (infrared). In particular, direct imaging offers the possibility of determining the colors and spectra of a large number of planets, independently by their orbital inclination. In other words, if we are able to directly observe a planet, we can directly have the spectrum of its atmosphere using a conventional spectrometer or an integral field spectrometer. Thus it is possible to gather information about the physical and chemical characteristics of a planetary atmosphere without waiting for the transit (transmission spectroscopy) of the planet or its occultation (emission spectroscopy). The resolution will be low, because the planet is faint; however, since many molecular bands are intrinsically low-resolution features, it will be still possible to learn a lot about atmospheres. The study of multi-band photometric images and spectra in the visible and infrared, and their temporal variation, of planetary systems makes it possible to estimate not only the orbital and physical parameters of the planet, but also the structure and composition of its atmosphere, its surface properties and rotation rate, and the likelihood of life on that planet. Such information provides the means to distinguish between gas and ice giants and rocky planets under a variety of circumstances such as distance from the star, age, etc.

Kepler's third law says that a planet with semi-major axis a (au) and eccentricity e has orbital period P (yr) given by

$$P = \frac{a^{3/2}}{(\frac{M_\star}{M_\odot})^{1/2}}.$$

If the distance from star to observer is d (pc), then the maximum angular separation between the planet and the star is

$$\theta = \frac{a(1+e)}{d},$$

where usually θ is expressed in arcsec, a in au, and d in pc. This quantity is important because it gives us idea about the resolution that a telescope with a primary mirror of diameter D should have in order to resolve any faint companion. Independent sources in a perfect, diffraction-limited image can be discerned when they are separated by at least λ/D in angle (Oppenheimer and Hinkley 2009). This is also called the Nyquist criterion, which is commonly derived with a Fourier optics approximation (see Sect. 4.4). So to resolve a planet that orbits its host star at 5 au at a distance from the observer of 10 pc, the telescope should be able to have an angular resolution better than 500 mas. This resolution seems already achievable with a 4 m class telescope. However, if we would like to see more inward, close to the star, let us say 1 au, for a star slightly farer, 50 pc, for example, we need an angular resolution higher than 20 mas. Taking into account that the most of faint companions (e.g., extrasolar planets) are cool in a way that requires us to use infrared bands to observe them, the problem is one of the most severe. Obviously, to reach such a high-angular resolution is a necessary condition but is not sufficient. We have also to fight against

Table 4.1 Single-pupil systems

Instrument	Telescope	Wavelength (μm)	Ang. resol. (mas)	Coronagraph
ACS	HST	0.2–1.1	20–100	Lyot
STIS	HST	0.2–0.8	20–60	Lyot
NAOS–CONICA	VLT	1.1–3.5	30–90	Lyot/FQPM
VISIR	VLT	8.5–20	200–500	–
SINFONI–SPIFFI	VLT	1.1–2.45	28–62	–
SPHERE	VLT	0.95–2.32	24–62	Lyot/APLC/FQPM
PUEO	CFHT	0.75–2.5	4–140	Lyot
CIAO	SUBARU	1.1–2.5	30–70	Lyot
OSIRIS	Keck I	1.0–2.4	20–100	–
AO–NIRC2	Keck II	0.9–5.0	20–100	Lyot
ALTAIR–NIRI	Gemini N.	1.1–2.5	30–70	Lyot
GPI	Gemini S.	0.9–2.4	24–62	Lyot/APLC
PALM-3000 PHARO	Hale 200″	1.1–2.5	60–140	Lyot/FQPM
PALM-3000 Project1640	Hale 200″	1.06–1.76	43–71	APLC
AO–IRCAL	Shane 120″	1.1–2.5	100–150	–

Modified from Absil and Mawet (2010)

the glare of the star and other difficulties (see Sect. 4.4), but first things first. In order to have high-angular resolution, two main types of observing techniques can be distinguished: single-pupil observations with large telescopes working close to their diffraction limit and interferometric (or multiple pupil) observations coherently combining the light from a few individual telescopes with ground separations up to a few hundred meters. In the first case, typical angular resolutions of 50 mas can be reached (i.e., 5 au at 100 pc), while in the latter case, angular resolutions down to ∼1 mas are achievable.

The single-pupil technique uses a single telescope with a large primary mirror (see Table 4.1) both ground and space based. Most of them, for all the problems related to the direct imaging of faint companion technique, are equipped with adaptive optics modules (see Sect. 4.5) or coupled with coronagraphic systems (see Sect. 4.6) to eliminate the light from the star and, thus, achieve high-contrast imaging. In Table 4.1 most of the systems indicated are the so-called first generation high-contrast imagers (for example, NAOS–CONICA just to cite one). In recent years a new generation of imagers (SPHERE, Project1640, and GPI) has popped up and since 2014[1] is on duty. For a detailed description of these instruments see Sect. 4.8.

By coherently combining the light collected from a few individual telescopes, stellar interferometry achieves an angular resolution equal to $\lambda/2B$ with B the separation between the telescopes (see Fig. 4.1), referred to as "baseline" (see Absil

[1]The first light of Project1640 was taken in 2010 (Hinkley et al. 2011).

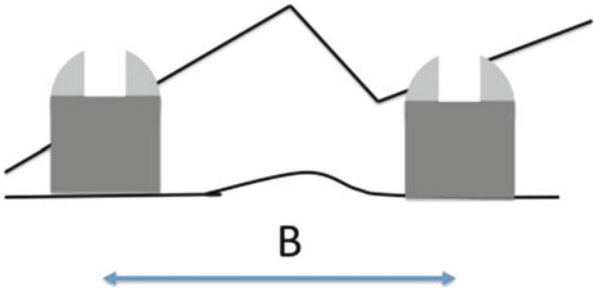

Fig. 4.1 The baseline *B* of multi-pupil systems

Table 4.2 Multiple pupil systems

Instrument	Interf.	Baseline (m)	Bands	Ang. res. (mas)	Spec. res.	Aperture
AMBER	VLTI	16–200	J,H,K	0.6–14	35–15,000	3
MIDI	VLTI	16–200	N	4–80	20–220	2
PIONIER	VLTI	16–200	H,K	1.5–45	15	4
V2	Keck I	85	H,K,L	2–5	25–1800	2
Nuller	Keck I	85	N	10–16	40	2
Mask	Keck	1–10	J to L	13–400	None	2
Classic	CHARA	34–330	H,K	0.5–7	None	2
FLUOR	CHARA	34–330	K	0.7–7	None	2
MIRC	CHARA	34–330	J,H	0.4–5	40–400	4
BLINC	MMT	4	N	250	None	2
LMIRCAM	LBTI	14–23	L,M	27–72	None	2
NOMIC	LBTI	14–23	N	72–200	None	2

Modified from Absil and Mawet (2010)

and Mawet 2010 and references therein). With baselines up to a few hundred meters in the world's leading facilities, interferometry currently reaches a resolving power equivalent to that of single-pupil telescopes much larger than even the most ambitious Extremely Large Telescope projects considered so far. With its angular resolution of typically 1 mas in the near infrared on hecto-metric baselines, stellar interferometry is the tool of choice to investigate the innermost parts of circumstellar disks in nearby star-forming regions. In particular, interferometry is currently the only suitable method to directly characterize the most important region of proto-planetary disks where dusty grains sublimate and where accretion/ejection processes originate. It also provides a sharp view of the region where terrestrial planets are supposed to be formed. For more evolved planetary systems, interferometry can be used to constrain the presence of circumstellar material in the inner few au, including large amounts of warm dust, (sub-)stellar companions, or even hot planetary mass companions. Most interferometric instruments are summarized in Table 4.2.

4.2 Discoveries and Biases

Up to now most of the positive detections of extrasolar planets (about 2000 objects, see Fig. 4.2) have been obtained exploiting indirect methods, i.e. searching for dynamical and photometric effects that an invisible companion causes on the status of its host star. The most efficient methods in this job are the radial velocities measurements and transit method. Radial velocity and transit searches have detected more than 1800 exoplanets and, in some favorable cases, enabled a spectroscopic characterization of the irradiated atmospheres of transiting giant planets (Vidal-Madjar et al. 2004; Wakeford and Sing 2015). These two techniques are biased towards planets in close orbits, typically smaller than 5 au . For this reason, the mass vs. semi-major axis parameter space in Fig. 4.2 is not homogeneously covered, and at the moment, just 87 planetary mass objects have been found at a separation larger than 5 au. Among these long-period objects, 53 have been found using the direct imaging technique, and 24 using the RV technique.

Exhaustive descriptions of these important indirect methods are given in the previous chapters of this book. However, it is interesting to briefly recall how they work, their limits, and observational biases.

The measurements of high precision radial velocities of stars exploit the Doppler effect due to the reflex motion of the star induced by the presence of an invisible body. This method allows us to have knowledge of the orbital properties of the system: the period P, the eccentricity e, the semi-major axis a, the argument of periastron ω, and the time of periastron T_0. The value of the mass of the companion obtained with this method is not the true value of the mass but it is known as

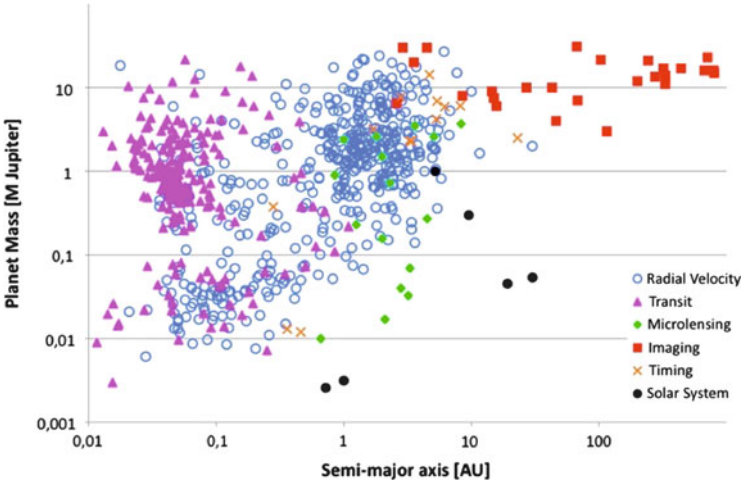

Fig. 4.2 Extrasolar planets discovered so far. The *different symbols* identify the different methods with which the planets have been discovered. Data taken from extrasolar encyclopedia (Schneider et al. 2013)

minimum mass ($M_p \sin i$) due to the degeneracy with the orbital inclination of the system. Furthermore it gives hints about the tidal history of the system (spin-orbit alignment through the Rossiter–McLaughlin effect), its architecture, and stability. The score of this method is more than 600 exoplanets with minimum masses ranging from giant planets (with masses of the order of Jupiter's mass and higher) to super earths ($1M_\oplus \leq M_p \leq 10M_\oplus$) (Mayor et al. 2014).

This method is very sensitive to giant planets that are close to their stars because in this case the radial velocity signal is stronger. Moreover there is also a time bias because planets with larger major axis have also larger periods (Jupiter is 5 au far away by the Sun and has a period of 12 years) inducing a very long time-based campaign to search for very low amplitude motion (see Fig. 4.3). We note that our knowledge of the period distribution is at present limited by the duration of radial velocity surveys and the sampling of long-period signals. This is also the region where significant overlaps with direct imaging and microlensing techniques are expected. Actually, concerning the latter, the already discovered objects have orbital parameters that are not well constrained because they only show long-term trends rather than a full orbital period.

Besides the instrumental noise, the radial velocity technique is also hampered by astrophysical sources of noise mainly due to the host star itself such as, for example, the stellar pulsations, surface granulation, and stellar jitter caused by star spots or other "instabilities" in the stellar atmosphere. At the 1 m/s level of precision, in fact, these physical phenomena in stellar photospheres give significant signals

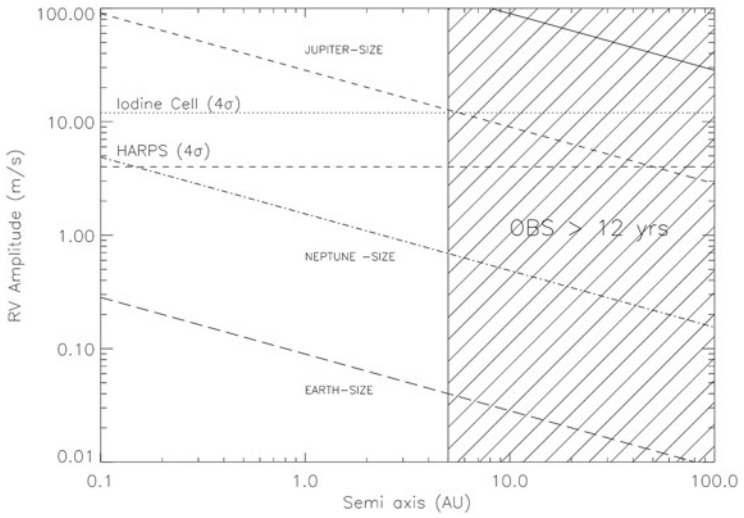

Fig. 4.3 In the figure indicated are the Doppler signals for four planets of different masses orbiting a solar-type star. Detectability threshold (4σ) for the two radial velocity measurement techniques (absorbing cell and simultaneous thorium) is also shown. The *shaded region* for a semi-major axis greater than 5 au shows a region where the method loses its sensitivity

that can hide planetary radial velocity signatures if not properly modeled. Solar-type stars have an outer convective envelope that exhibits variability on different timescales. In their paper (Pepe and Lovis 2008), the authors describe the several sources of noise that are necessary to take into account in high precision radial velocity measurements. In the midst of them they individuate four individual stellar noise contributions like granulation (G), meso-granulation (MG), super-granulation (SG), and active regions (AR) that are able to induce radial velocity variability at the meter-per-second level (see also Dumusque et al. 2011a,b). These contributions as well as their sum are shown in Fig. 4.4 where the gray part of the plot indicates the frequency region, in which all the planets discovered with radial velocity methods are confined.

From Fig. 4.4 it is possible to understand that the activity of the star may be one of the major noise sources over timescales of months and years, producing radial velocity jitter with amplitudes up to several $\mathrm{m\,s^{-1}}$. This could hide RV signals due to small planets also at moderate distances from the stars (e.g., Earth-mass planets in the habitable zone of solar-type stars for which the RV signal is about $10\,\mathrm{cm\,s^{-1}}$). Furthermore at high frequencies, the most important noise source is the stellar p-mode, density waves propagating inside the star and producing quasi-periodic oscillations of its surface. These pulsations can be observed directly and have lifetimes of days. These oscillations are very hard to model and so they cannot be easily subtracted from the signal. In summary, the radial velocity method is sensitive to larger and closer planets that orbits dwarf main-sequence stars. These biases could be recognized in the distribution of the planets in the semi-major axis—mass diagram shown in Fig. 4.2.

The transits technique is perhaps the most direct of indirect methods for discovering extrasolar planets. In fact the small fading of the stellar light during the transit is due to the opaque body of the planet that passes in between the observer and the star, allowing the observer to "see" the planet. It is harder to see the occultation (secondary eclipse), but by selecting the right wavelength region (e.g., NIR or MIR), it is possible to observe it also. From the observation of one or more transits it is possible to unveil several important pieces of information on the planetary companion. Besides the orbital parameters like the period P, the orbital radius a, the time of the contact points, and the inclination i of the orbit, it is possible to derive paramount information on the physical constitution of the planet. In fact, the fading of the light curve of the star is proportional to the square of the ratio between the radius of the planet and that of the star. This allows measurement of the radius of the planet and because the planet transits its star, the inclination of the orbit is about 90°. If the star is bright enough, it is possible to follow it up with a spectrograph and measure radial velocities in order to know the mass (the actual mass) of the planet. This fact gives in a chain the density and the surface gravity of the planet, giving hints on the inner constitution of the body of the planet (rocky or gaseous) and a first order of its possible constitution of the atmosphere (surface gravity and escape velocity). Results from Kepler spacecraft (Lissauer et al. 2014) show that transit search could reveal the architecture of extrasolar planets systems such as, for example, the crowded system Kepler-11 with its six planets (Lisseauer

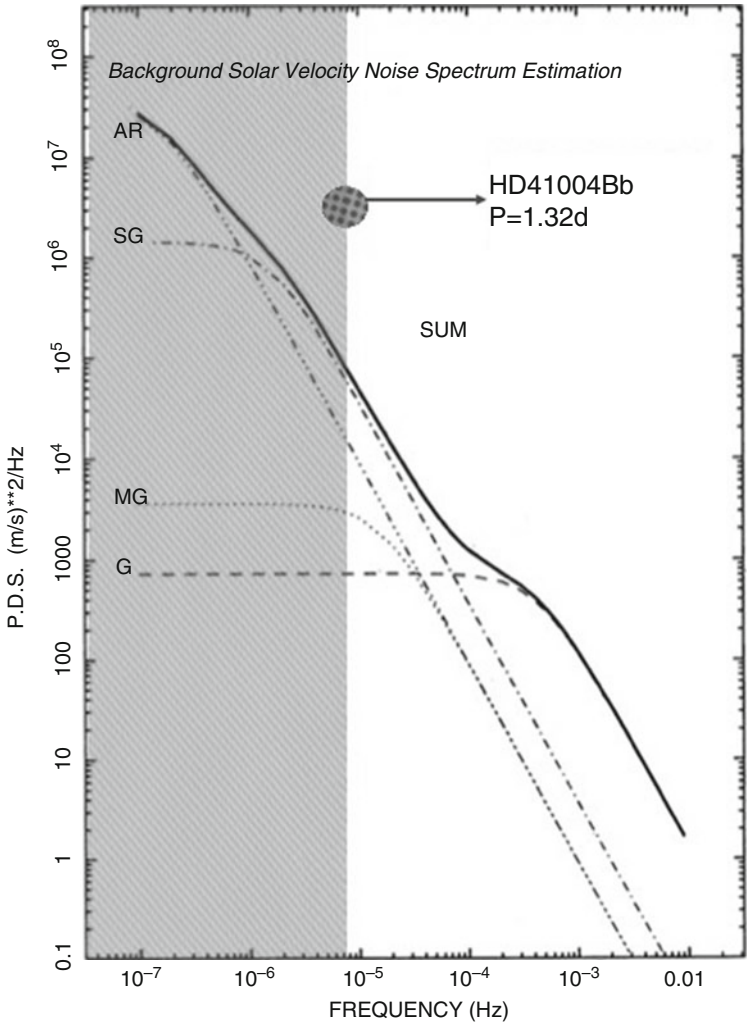

Fig. 4.4 The four contributions to the radial velocity background of the Sun described in the text and their sum are displayed in the frequency dominion (power spectrum). The position of the planet with the shorter period is highlighted. The plot is modified from Fig. 2 of Pepe and Lovis (2008)

et al. 2011) or the circumbinary system Kepler-16AB b (Doyle et al. 2011). The score of this method is about 1200 new planets with dimension down to Earth-sized planets like Kepler-20 e and f (Fressin et al. 2012).

During a transit, a small portion of the starlight is filtered through the upper atmosphere of the planet, where it is only partially absorbed. The absorption will be wavelength-dependent due to the scattering properties of atoms and molecules in the planetary atmosphere. At the wavelength of a strong atomic or molecular transition,

the atmosphere is more opaque, and the planet's effective silhouette is larger. This raises the prospect of measuring the transmission spectrum of the planet's upper atmosphere and thereby gaining knowledge of its composition (Seager and Sasselov 2000; Brown 2001).

Observations spanning occultations thereby reveal the relative brightness of the planetary disk, if the ratio of the radii is already known from transit observations. The planetary radiation arises from two sources: thermal radiation and reflected starlight. Because the planet is colder than the star, the thermal component emerges at longer wavelengths than the reflected component. With emission spectroscopy we have the emission spectrum of the planet averaged over the visible disk of the dayside. From that it is possible to derive information on the thermal structure of the atmosphere and on the Bond albedo of the planet (Deming et al. 2005; Charbonneau et al. 2005; Knutson et al. 2007). Occultation and transit spectroscopy thereby provide different and complementary information about the planetary atmosphere.

As for the RV method, transit search is also hampered by selection effect. First of all the probability to detect a transit depends on the orbital axis of the planet. The closer the planet, the higher the probability to observe it in transit. The photometric transit signal depends on the ratio between the surfaces of both the planet and the star disks, so it is deeper as this ratio is higher. This means that larger planets orbiting smaller stars are simpler to detect. Furthermore the activity of the host star and the presence of related photospheric structures like spots or faculae introduce an astrophysical noise into the light curve of the star. The amplitude of this source of noise depends on the star itself and its age. So generally the properties of the sample of a ground-based transit survey are shaped by selection effects that led to the discovery of large planets in short-period orbits around still stars. This is not true for high precision space based surveys as in the case of Kepler.

In this framework direct imaging could be of great help, as it is the only technique currently available to detect planets with semi-major axes greater than about 5 au in a reasonable amount of time. At wider orbits, the deep imaging technique with space telescopes (e.g., HST) or the combination of adaptive optics (AO) systems with very large ground-based telescopes (e.g., Palomar, CFHT, Keck, Gemini, Subaru, and VLT) is particularly well-suited. Direct imaging is then a complementary technique to explore the outer regions of exoplanetary systems. In particular, this technique allows us to study the regions beyond the snow line around young stars, offering the possibility to study young systems (direct imaging is not sensitive to activity), dynamical evolution of planetary systems, and the connection between recently formed giant planets and the circumstellar environment (Lafrenière et al. 2007a; Chauvin et al. 2010). Also, taking advantage of the intrinsic luminosity of young giant gaseous planets in the first phases of their evolution, we can infer their masses exploiting theoretical evolutionary models by, e.g., Chabrier et al. (2000), Baraffe et al. (2002), Baraffe et al. (2003), Fortney et al. (2005), and Burrows et al. (2006). However, the mass determination is subject to unconstrained physics and unknown initial conditions at very young ages (Marley et al. 2007; Spiegel and Burrows 2012) and large discrepancies on the derived mass are expected between the so-called hot-start and cold-start models (see Sect. 4.4).

Table 4.3 Different goals for the different methods

	Hot planets	< Snow line	> Snow line
Discovery: detection and statistics	Radial velocities and transits	Radial velocities space astronomy (GAIA) Microlenses ELT imaging	8 m imaging
Dynamical charac. and structure	Radial velocities and transits	Radial velocities space astronomy (GAIA) ELT imaging	Coupling 8 m imaging and GAIA
Atmospheric charac. and search for biosignatures	Transits duration Transm. and Occ. Spectroscopy	ELT imaging	8 m imaging and JWST ELT MIR

The use of extremely large telescope (ELT) and James Webb space telescope (JWST) is also indicated

Direct imaging provides insights on formation and migrations mechanisms for planetary systems. Moreover, this technique allows us to obtain photometric, spectroscopic, and astrometric measurements of the detected companions, and for this reason it is a fundamental technique to study the atmosphere of the known objects, their mass-luminosity function, and their orbits (see, e.g., Rameau et al. 2013a; Esposito et al. 2013; Currie et al. 2013). High-contrast imaging (see Sect. 4.4) is ideal for characterizing planets on wider orbits. Indeed, a 10 m telescope imaging at H band (1.6 μm) has a 32 mas diffraction limit. Such an instrument could resolve a planet on a 5 au orbit around a star at 150 pc, approximately the distance to the Orion star-forming region. Currently, the direct imaging biases are complementary to those of the main indirect methods: outer and young planets versus closer and cold planets. In a very general sense we can consider the space around the host star subdivided into three zones. The first, the closest, where hot planets with short orbital periods could exist, the other two defined as those inside and outside the snow line. This is a very rough definition but allows us to put some points about the different goals achievable by the different methods and summarize what we described in the previous paragraphs. The different methods could be used in order to achieve the different goals that are reported in a schematic form in Table 4.3. We have to take into account that this very general schematization could be very different for specific target groups such as, for example, the M stars.

4.3 Astrophysical Motivation

During the last 20 years, the advent of the Hubble space telescope (HST), the adaptive optics on 4–10 m class ground-based telescopes with the new high-contrast imagers, and long-baseline infrared stellar interferometry, has opened a new viewpoint on the formation and evolution of planetary systems. By spatially resolving the optically thick circumstellar disks of gas and dust where planets

are forming, these instruments have considerably improved our models of early circumstellar environments and have thereby provided new constraints on planet formation theories. Within the multiple observing techniques that have been used to constrain planetary formation and evolution, high-angular resolution techniques in the visible and NIR regime hold a very important place because they are among the few that have the potential of spatially resolving the various bodies and physical phenomena at stake in planetary systems. High-angular resolution techniques are also directly tracing the mechanisms that govern the early evolution of planetary embryos and the dispersal of optically thick material around young stars. Mature planetary systems are being studied with an unprecedented accuracy thanks to single-pupil imaging and interferometry, precisely locating dust populations and putting into light a whole new family of long-period giant extrasolar planets (Absil and Mawet 2010). The evolution of these techniques is pushing the limits deeper towards the region closer and closer to the star. This instrumental potentiality makes richer and more ambitious the list of scientific goals for direct imaging that now could span two bridges: from understanding of the physical mechanism of the formation of the planets in the protoplanetary disks to the architecture of planetary systems; from the study of the properties of planetary atmospheres to their composition. Our understanding of planet formation can only be complete when we reach a broader understanding of the physics of circumstellar disks. Indeed because disks are the birthplace of planets, a comprehensive understanding of how these systems form and evolve may shed more light on the planet formation process allowing us to understand which of the two leading theories of giant planet formation, the gravitational instability model (Boss 1997; Mayer et al. 2002), and the core accretion model (Mizuno 1980; Pollack et al. 1996; Laughlin et al. 2005; Alibert et al. 2005) describe the process in a better way. The planetary formation process and the disk evolution are also strongly tied to the architecture of the system and to the individual planetary atmosphere history and evolution. The direct imaging of a planetary system allows us to observe all components of the system in one shot and have information on the astrometry of the components by their position on the image and measure the flux output of the components. From the astrometry and the variation of the position in time it is possible to measure the orbital parameter and constrain the architecture of the system and the orientation of the planetary disk. The flux measurement gives us an idea about the spectrum of the object and on the polarization degree of light emitted by the planet. Moreover the observation of the object in different epochs allows us to recognize if there are yearly or seasonal variations of the flux and spectrum. In fact, after the discovery of planetary companions it is important to understand the physics of these bodies and how they are similar to or different from the planets of our Solar system. In other words, it is necessary to make a step towards characterization of each planet. This step could be done by observing the planets both in combined light and directly in broad photometric bands. The latter means to evaluate the color of the planets from which we are able to compare the results with a theoretical model in order to guess the atmospheric physical behavior. But the most natural way to observe exoplanet atmospheres is by taking an image of the exoplanet. This is very important

because contrary to the stars, which follow well defined physical relations that tie between them luminosity, surface temperature, radius, and age, planets are not such affable entities. Such a behavior is a powerful observational tool that permits us to derive, even if with some caveats, a wide range of stellar parameters from a few basic observables. Conversely, the knowledge of the mass may provide very little information about a planet, namely if it is a gas giant, an icy giant, or a rocky one, and in many circumstances it is hard to distinguish the last two categories from each other (Tinetti et al. 2013). For a planetary body, mass, radius, temperature, and chemical composition are often loosely correlated parameters, and cannot be disentangled from the initial conditions, history, and interaction with the host star. Planets can be very similar in mass and radius and yet be very different worlds. A spectroscopic analysis of the atmospheres is needed to reveal their physical and chemical identities.

The direct observation of planets can allow us to recognize the chemical species present in the visible or infrared region of the spectrum of giant planets and determine the pressure–temperature profile and the behavior of the condensation in the atmosphere. More challenging, also if possible in principle, is the observation of telluric planets. The atmosphere of telluric planets is strongly influenced by their cooling and impact history during evolution and could appear very different from each other. In this they are quite different from the atmosphere of the giant planets. For example, in the Solar system, the composition of the giant planet atmospheres is hydrogen dominated with CH_4, NH_3, and other hydrogenised species (Tinetti et al. 2013). The atmospheres of rocky planets (Mars and Venus) are instead predominantly made of CO, CO_2, and N_2, as hydrogen escaped due to their relatively low gravity field. H_2O is expected to be present, and is indeed observed, in the interior of the giant planets; its presence on terrestrial planets may be explained, at least partially, by an external origin, i.e. cometary and meteoritic impacts (e.g., Earth). The presence of chemically based life on a planet would change the composition of its atmosphere away from the biological steady state introducing an out of equilibrium chemical species as a global sign of life. This change would be recognizable even at astronomical distances in the spectrum of light reflected or emitted by a planet's atmosphere or surface. Biomarkers, one of the cornerstones of astrobiology, are defined as detectable species, or sets of species, whose presence at significant abundance strongly suggests a biological origin (Kaltenegger and Selsis 2008; Kaltenegger et al. 2010). In previous years some researchers stated that evidence for biology is simultaneous detection of O_2 or O_3, along with a reduced gas such as CH_4 or N_2O. This is a powerful diagnostic for a disequilibrium condition (see, for example, Kasting and Catling 2003; Turnbull et al. 2006; Lunine et al. 2008; Kaltenegger and Selsis 2008; Meadows and Seager 2010). These biomarkers have spectral features that are even detectable with a very modest spectral resolution of 30–40 in the NIR, opening this realm to the high-contrast imager that operates with this spectral resolution (see Sect. 4.8). Indeed there are many motivations that justify the use of direct imaging techniques. Many research efforts in this field have been filed and other efforts made to develop more and more

sophisticated techniques. The list of principle motivations can be summarized as follows:

- Exo-zodiacal powder properties
- Morphology of circumstellar disks
- Planets formation: core accretion/disk instability
- Architecture of planetary systems
- Planetary atmosphere composition
- Presence and characteristics of clouds
- Structure of planetary atmospheres (vertical distribution)
- Composition and structure of planetary surfaces (if present and visible)
- Temporal variation of atmospheric composition and structure
- Planetary rotation velocity
- Discovery of "weird" planets in planetary systems

4.4 Observational Issues

From the previous sections we can image what will be and what kind of results we would like to obtain from direct imaging observations. A clear example is given by the Gliese 229 B case (Oppenheimer and Hinkley 2009). This brown dwarf is well separated by its host star (>7 arcsec) with a relatively low contrast of about 10^4 in the optical and NIR. In that case (see Fig. 4.5) a basic stellar coronagraph and some standard spectrographs were sufficient to acquire data (Oppenheimer et al. 2001) and also spectroscopy of the low mass companion.

The first direct images of exoplanets were published in 2008 (Ducourant et al. 2008) for the planet orbiting the Brown Dwarf 2M1207A observed for the first time by Chauvin et al. (2004) with VLT/NACO. This detection came after 12 years from the discovery of the first exoplanet and after that more than 300 of them had been measured indirectly by radial velocity, transit, and microlensing techniques. This huge time lag occurred because direct imaging of exoplanets requires extraordinary efforts in order to overcome the barriers imposed by astrophysics (planet–star contrast), physics (diffraction), and engineering (scattering).

In the real world, in fact, the direct detection of extrasolar planets is challenging mainly for two reasons: (1) the large luminosity contrast with respect to the star, which is of the order of 10^6 for giant young planets with high intrinsic luminosity, and 10^8 to 10^9 for old planets seen in reflected and intrinsic light; and (2) the small separation between the star and the planet, of the order of a few tenths of arcsec for planets at a few au around stars at a distance up to 100 pc from the Sun. As a result, the light from the companion objects is completely overcome by the light of the host star (e.g., in Fig. 4.5, 1 au and 5 au orbits are indicated inside the glare of the star). Exceptions could be planets with a large albedo and would be very close to the star and, hence, would reflect a significant amount of starlight, but then it is too close to the star for direct detection. Other exceptions are young planets, which

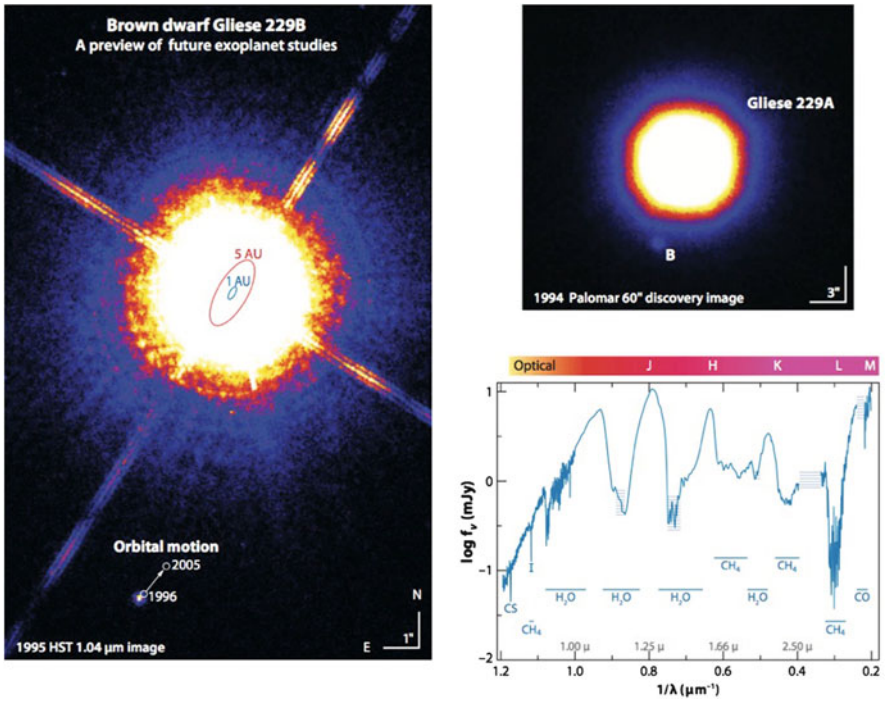

Fig. 4.5 HST and Palomar images of the Gliese 229 system. The system, an M-dwarf and the T-dwarf companion, has been discovered with coronagraphy. The numerous follow-up observations allow to obtain orbital information and spectroscopy over a broad wavelength region. The picture was taken by Oppenheimer and Hinkley (2009) and reprinted with permission from the authors

are self-luminous due to ongoing contraction and maybe accretion, so that they are only 2–4 orders of magnitude fainter (for 13 to 1M_J planets, respectively) than their (young) host stars (Burrows et al. 1997; Baraffe et al. 1998). Hence, direct imaging of planets is less difficult around young stars with ages up to a few hundred of Myr. The radiation emitted by a planet is compounded by two distinguished components: a reflected component and an intrinsic one (see Fig. 4.6). The first component depends on the brightness of the host star (its spectral type and luminosity class), the distance of the planet from the star (the semi-major axis), and the albedo of the planet. The second component is due instead only to the temperature T of the planet itself. The brightness of the star is a function of the wavelength and the effective temperature ($T_{\rm eff}$) of the star: $B(\lambda, T_{\rm eff})$. Considering a planet with a radius $R_{\rm p}$ that orbits the host star at an orbital radius a, the flux reflected by the planets will be

$$F_{\rm p,Vis} = A(\lambda, t)\phi(t)\frac{R_{\rm p}^2}{4a^2}B(\lambda, T_{\rm eff})R_\star^2,$$

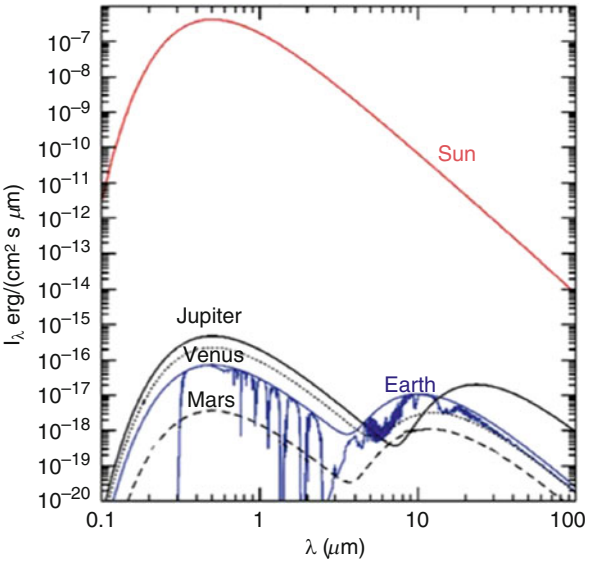

Fig. 4.6 Representation of Solar system components emission as observed from 10 pc. All components are represented by black bodies except the Earth for which the spectrum is shown. The figure was taken from Kaltenegger et al. (2010) and reprinted with permission from the authors

where $R_\star$ is the stellar radius, $\phi(t)$ is the phase angle, while $A(\lambda, t)$ is the albedo. The flux reflected by the disk of the planet is mainly in the visible region because of the irradiation characteristics of the host stars.

The last are fundamental quantities when the irradiation properties of a planet are under study. The phase angle ϕ is a function of the time and individuates the fraction of the planetary disk that is illuminated by the star. It is defined as the angle between the observer–planet direction and the planet–star direction, which varies from 0 to π rad. So $\phi = 0$ at superior conjunction with the planet at the opposite of the observer, $\phi = \pi/2$ at quadrature (maximum elongation for a circular orbit), and $\phi = \pi$ at inferior conjunction with the planet between the star and observer. The fraction of the planetary surface illuminated by the star is given by a phase law $f(\phi)$ at phase angle ϕ. For a Lambert sphere the phase law is (Traub and Oppenheimer 2010):

$$f(\phi) = [\sin(\phi) + (\pi - \phi)\cos(\phi)]/\pi.$$

For example, in an edge-on system $f(0) = 1$ at superior conjunction, $f(\pi/2) = 1/\pi$ at maximum elongation, and $f(\pi) = 0$ at inferior conjunction.

The fraction of the stellar light reflected by the planets depends on the planet itself or more precisely on the constitution and the status of its atmosphere. This fraction is called in a general sense albedo. The definition of albedo is not unique because it depends on what kind of reflection properties of matter we would like to

Table 4.4 The geometric and Bond albedo values for the planets of the Solar system (de Pater and Lissauer 2010)

Planet	a (au)	Geometric Albedo	Bond Albedo
Mercury	0.387	0.138	0.119
Venus	0.723	0.84	0.75
Earth	1.000	0.367	0.306
Mars	1.524	0.15	0.250
Jupiter	5.203	0.52	0.343
Saturn	9.543	0.47	0.342
Uranus	19.19	0.51	0.290
Neptune	30.07	0.41	0.310

take into consideration. But generally both geometrical albedo and Bond albedo are considered in describing the general reflectance properties of the planetary surface. The geometric albedo p of a planet is defined to be the ratio of planet brightness at phase angle $\phi = 0$ to the brightness of a perfectly diffusing disk with the same position and apparent size as the planet (Seager 2010). The geometric albedo will in general be wavelength dependent. The Bond albedo of a planet, $A_{\rm Bond}$, is defined to be the ratio of total light reflected to total light incident, where "total" here means bolometric, i.e., integrated over all wavelengths, and the whole planet. In Table 4.4 listed are the values of geometric and Bond albedo for the Solar system planets.

Considering the planet like a black body at a given temperature $T_{\rm p}$, the intrinsic emission component of the planets is given by the following:

$$F_{\rm p,IR} = \epsilon_{\rm IR} B(\lambda, T_{\rm p}) R_{\rm p}^2$$

where $\epsilon_{\rm IR}$ is the emission coefficient of the planet and is a function of the wavelength and of the chemical compound of planetary atmospheres and surface. The temperature $T_{\rm p}$ is the actual temperature of the planet. It could be due to the absorbed fraction of the host star radiation, which homogeneously heats the planet, in this case it is named equilibrium temperature ($T_{\rm eq}$), or to the sum of the absorption of radiation and the internal heat of the planet, if any. In this latter case it is the effective temperature of the planet ($T_{\rm eff}$). Furthermore, the contribution of the greenhouse effect due to the presence of the planetary atmosphere should be added to both the equilibrium and effective temperature. For example, greenhouse warming due to the Earth's atmosphere enhances the equilibrium temperature of the planet to about 30° (Kasting and Catling 2003).

To evaluate the equilibrium temperature of a planet at distance a from its star we have to take into account that the star's luminosity is given by

$$L = 4\pi r_{\rm s}^2 \sigma T_{\rm eff}^4,$$

where r_s is the radius and $T_{\rm eff}$ the effective temperature of the star. The flux from a star is diluted by a^2 when it reaches a planet at distance a. Of this, a fraction $(1-A_{\rm Bond})$ is absorbed by the planet. The resulting radiative equilibrium temperature $T_{\rm eq}$ of a planet is determined by setting the incident flux equal to the radiated flux, assuming that the heat from the incident radiation is uniformly distributed over a fraction f of its total surface area, and that it radiates with an emissivity of unity. We find

$$T_{\rm eq} = \left(\frac{1 - A_{\rm Bond}}{4f}\right)^{1/4} \left(\frac{R_{\rm p}}{a}\right)^{1/2} T_{\star}.$$

Here $f = 1$ for a rapid rotator and $f = 0.5$ for a tidally locked or slowly rotating planet with no transfer of heat from the hot to cold side. The value of $T_{\rm eq}$ refers to the fraction f of area over which the heat is spread; this simple formulation assumes that none of the incident heat is distributed to the $(1 - f)$ of the remaining area, which would therefore be very cold. Instead, to determine the $T_{\rm eff}$ value once the IR part of the planetary emission is observed, it should be fitted with a black body curve. Generally, the effective temperature $T_{\rm eff}$ is different by the $T_{\rm eq}$. For example, in the case of giant planets of Solar system the two values are different [in the case of Jupiter: $T_{\rm eq} = 110$ K and $T_{\rm eff} = 124.4$ K (de Pater and Lissauer 2010)]. The inner heat source of rocky planets is quite negligible and the two temperatures are equal.

The ratio between the emitted planetary flux and the stellar flux is defined as contrast. The contrast is a function of the wavelength, the properties of the planet, its age, and the apparent geometry of the planet–star system. It could be written in the following way:

$$C = \frac{(F_{\rm p,Vis} + F_{\rm p,IR})}{F_{\star}}.$$

The expected visible-wavelength contrast of typical Jupiter-like and Earth-like planets around nearby stars is shown in Fig. 4.6. To be more clear, the contrast for a planet like Jupiter (for an angular separation of 0.5 arcsec) is about 10^{-9} in the visible band and 10^{-6} in the NIR, while for Earth-like planets in the habitable zone of a G star the contrast will be of about 10^{-10} and 10^{-7} in the visible and NIR at separations of about 0.1 arcsec. Estimates based on theoretical models (Burrows et al. 1997; Baraffe et al. 2003) show that young planets could be also three orders of magnitude brighter than old ones[2] (see Fig. 4.7). This could be reflected on the contrast, hence on the observational strategy (see ahead) in the sense that it depends on the distance from the star for old planets (negligible intrinsic flux

[2]This is what is called a warm start. Not all agree with this vision of the evolution of the planets just after the formation. For example, (Marley et al. 2007) consider instead a cold start with planets that become brighter during the gravitational focus event when the gas is accreted from the disk.

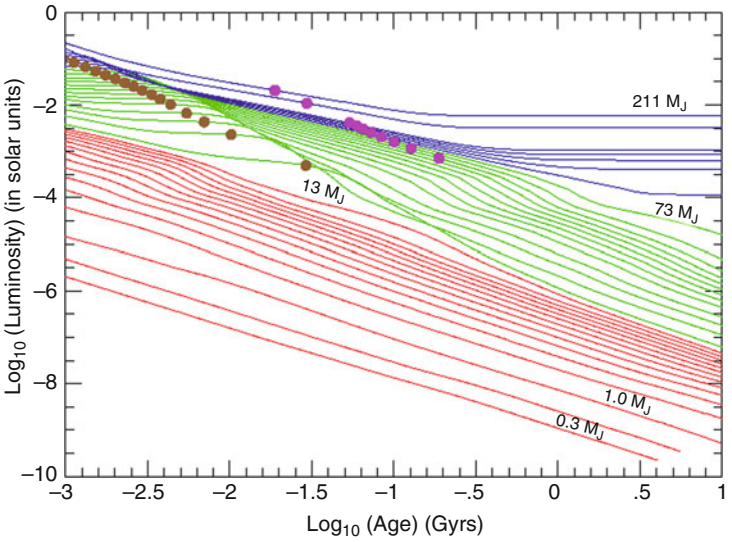

Fig. 4.7 Theoretical cooling models for stellar (*solid line*) and sub-stellar (*dashed line*) structures and giant planets (*dot dashed line*). The masses of the structures are labeled in Solar mass units. The figure was taken from Burrows et al. (2001) and reprinted with permission from the authors

component) while it is independent of the distance for young planets (high intrinsic flux component).

On the basis of these cooling models for planets, it is possible to see that detailed spectroscopic studies of even young (∼100 Myr) and hot (∼800 K) planets of about 1 M_J require a contrast of 10^{-8}. This depends upon the age and physical distance between the star and planet, but the point is that although one could distinguish one photon for every 10^6 from the star (see Fig. 4.7), probing the depths of the spectral features requires at least two more orders of magnitude in precision. The majority of planets are not hot and young, however, and in those cases the contrast is much worse making this kind of study really challenging. Study of planets similar to the Earth may require a contrast about 10^{-10} (Oppenheimer and Hinkley 2009; Traub and Oppenheimer 2010).

4.4.1 Image and Speckles Formation

We can consider the sketch outlined in Fig. 4.8 as a section of the entrance pupil of dimension D of a single telescope. The light coming from a star is then focalized on the focal plane of the telescope where it is possible to feed another instrument. The wavefront coming from a far away star could be considered as a plane wavefront [for a star at 10 pc, the deviation of this segment of the sphere is $\leq 10^{-17}$ for a 4-m telescope (Oppenheimer and Hinkley 2009)]. Once it has impinged on the

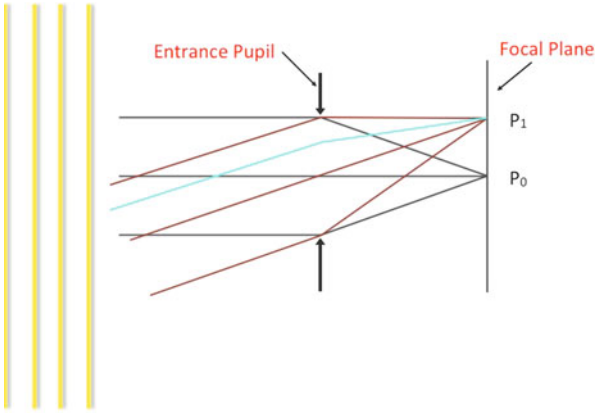

Fig. 4.8 The image formation on the focal plane of a single telescope with a circular aperture. The sketch shows the aperture size along one diameter

telescope aperture, each point inside the aperture becomes an emission center of spherical waves (Huygens principle) that are focalized by the mirror of the telescope onto a plane. The emission of secondary wavelets is only inside the aperture of the telescope. Considering Fig. 4.8, the wavelets that propagate in the direction orthogonal to the aperture are focalized in the P_0 point while those that move at a generic angle θ are focalized in the P_1. In the former case, because all secondary wavelets have an orthogonal direction to the aperture, there is no difference in their optical path and they arrive in P_0 all with no phase difference. Instead for the latter case, each wavelength moves with a different optical path depending on the position x_1 on the aperture (plane 1). The optical path difference between two generic secondary wavelets arriving in P_1 could be easily derived by geometrical consideration: $\Delta = x_1 \sin\theta_2$ where θ_2 is the generic angle seen by the focal plane (plane 2). The difference in optical path between two waves is translated in phase difference: $\delta\phi = (2\pi/\lambda)\Delta$. Substituting the value of the optical path difference in the last relation , the phase $\phi_1(x_1)$ is

$$\phi_1(x_1) = 2\pi x_1 \sin(\theta_2)/\lambda \cong 2\pi x_1\theta_2/\lambda\,.$$

On the focal plane we have to consider the sum of all the secondary wavelets that exit the pupil at angle θ_2 to a star image at a point in the image plane, i.e., plane 2. The amplitude of the electric field in this plane is denoted by $A_2(\theta_2)$. The amplitude $A_2(\theta_2)$ in the image plane is the algebraic sum of all wavelets across the pupil (Traub and Oppenheimer 2010)

$$A_2(\theta_2) = \sum \text{wavelets} = \int_D A_1(x_1)e^{i\phi_1(x_1)}dx_1/D,$$

where the integral is over the diameter D of the pupil, and $A_1(x_1)$ is the amplitude of the incoming wave in plane 1. Here the pupil is one-dimensional, but generally it is two-dimensional. The factor D as a divisor is necessary in order to normalize

the right-hand part of the equation, but we drop it for simplicity. Inserting the approximate expression for $\phi_1(x_1)$ and assuming $\theta_2 \ll 1$ we get

$$A_2(\theta) = \int_D A_1(x_1) e^{i2\pi\theta x_1/\lambda} dx_1. \tag{4.1}$$

In this case, where we retain only the linear approximation $\sin(\theta) = \theta$ in the pupil plane, we refer to Fraunhofer diffraction (Born and Wolf 1999, Chap. 8.3). The more exact, but more difficult to calculate, case is that of Fresnel diffraction, in which we retain higher-order terms. In a stellar case in which the light source could be considered at the infinite, the Fraunhofer diffraction gives exact results. For the case of a one-dimensional pupil, with an input amplitude $A_1(x_1) = 1$, we find

$$A_2(\theta) = \int_{-D/2}^{D/2} e^{i2\pi\theta x_1/\lambda} dx_1 = \frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda} D.$$

The measured intensity is $I = |A|^2$, or

$$I_2(\theta) = [\frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda}]^2 D^2.$$

The intensity pattern is thus the square of a $\mathrm{sinc}(X) = \sin(X)/X$ function, with a strong central peak at the point where the source star would have been imaged with geometrical optics (here $\theta_0 = 0$), and small secondary peaks. The first zero is the solution of $I_2(\theta_z) = 0$ and is given by $\theta_z = \lambda/D$. The corresponding results for a two-dimensional circular aperture of diameter D are roughly similar. The physical amplitude in the focal plane (Born and Wolf 1999, Chap. 8.5.2) is

$$A_2(\theta) = \sqrt{I_0} \frac{2J_1(\pi D\theta/\lambda)}{\pi D\theta/\lambda},$$

and the corresponding intensity is $I_2 = |A_2|^2$, giving

$$I_2(\theta) = I_0 \left[\frac{2J_1(\pi D\theta/\lambda)}{\pi D\theta/\lambda}\right]^2.$$

Here $J_1(X)$ is the Bessel function of first order. The first zero-intensity angle is $\theta_z = 1.22\lambda/D$ which is the famous result (Airy disk radius) for a clear circular aperture. The full-width at half-maximum (FWHM) of the intensity pattern is $\theta_z = 1.03\lambda/D \cong \lambda/D$ so this value is often referred to as the diameter of the diffraction-limited image of a point source. The relative intensities of the central and first four secondary maxima are $1.0, 0.017, 0.0042, 0.0016$, and 0.00078 at respective values of $0.0, 1.6, 2.7, 3.7, 4.7$ times λ/D. So, from a theoretical point of view, a diffraction-limited telescope provides images with an FWHM $\sim \lambda/D$, with λ the observing wavelength, and D the diameter of the telescope.

Unfortunately, once light reaches the Earth it passes through about 20 km of atmosphere and the resolution is degraded to λ/r_0, where r_0 is the Fried parameter (see Sect. 4.5) indicative of the seeing cell size, typically about 10 cm in the visible. The length r_0 is the diameter of a region over which the wavefront has a root mean square (RMS) variation of about 1 rad (technically, 1.015 rad, by definition in the Kolmogorov model of turbulence). The Fried parameter is strongly dependent on wavelength and on the observing site. In fact, the value of r_0 scales as $\lambda^{6/5}$, so it is larger in the infrared. The solution to the problem due to the turbulence of the atmosphere is twofold and the obvious one is a complete bypass of the Earth atmosphere by going into space (e.g., HST). The second solution is to design adaptive optics systems correcting the optical effect of the turbulence in real time, and thereby restoring the full or almost the full resolution of the telescope (see Sect. 4.5).

In imaging, the effects of wavefront errors manifest themselves as speckles. The image of a point source taken with a short exposure time with no wavefront correction is composed by speckles with short lifetimes (see Fig. 4.9). Speckles are interference figures coming from several coherent patches of diameter equal to r_0

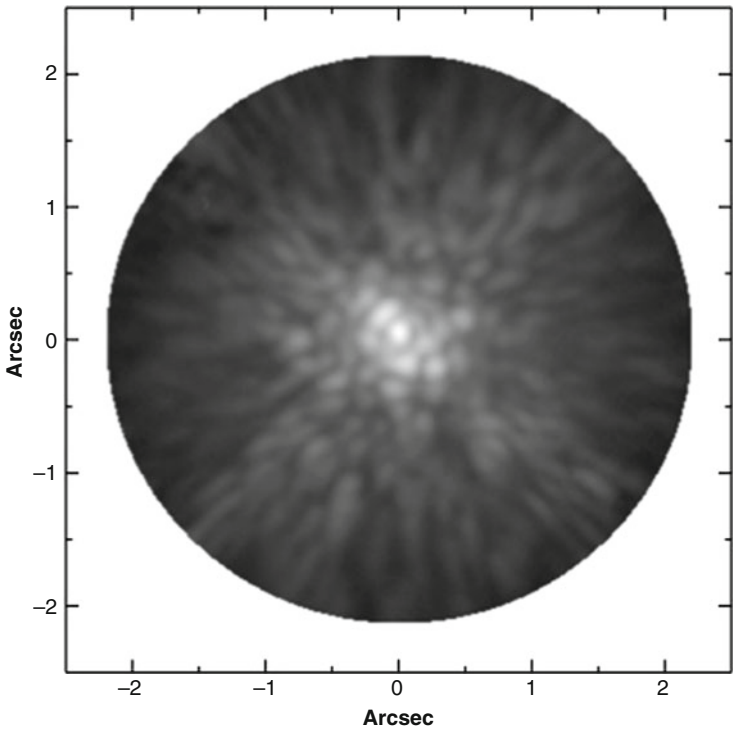

Fig. 4.9 Image taken with a coronagraph showing the presence of speckles. The image was taken by Oppenheimer and Hinkley (2009) and reprinted here with permission from the authors

(Racine et al. 1999) that are distributed just above the aperture of the telescope. Following the diffraction theory, one aperture of dimension r_0 should produce a point spread function (PSF) with an FWHM of $\sim \lambda/r_0$. Two apertures at a distance of D from each other constitute an interferometer that reproduces on the focal plane of the telescope an interference figure with fringes that run in an orthogonal direction to the conjunction of the two apertures. The width of this interference pattern is $\sim \lambda/D$. If we have a third aperture not aligned with the other two, we will have three different couples of aperture with the results of three different systems of interference fringes. If there is a constructive interference, the speckle is a bright one. Due to the random phase variation the pattern moves itself inside the dimension of the PSF causing the characteristic boiling of speckles.

In order to describe in a quantitative way the speckles formation, we can suppose that the wavefront incident on a telescope is advanced by a phase step $\phi/2$ on one half of the pupil, and delayed by $\phi/2$ on the other half. The electric field amplitude on the focal plane of the telescope using Eq. (4.1) is given by

$$A_2(\theta) = \int_0^{+D/2} e^{i\phi/2} e^{i2\pi x_1 \theta/\lambda} dx + \int_{-D/2}^{0} e^{-i\phi/2} e^{i2\pi x_1 \theta/\lambda} dx,$$

that results in

$$A_2(\theta) = \frac{\sin(\pi D\theta/2\lambda)}{\pi D\theta/2\lambda} \cos(\pi D\theta/2\lambda + \phi/2) D.$$

If the wavefront has $\phi = 0$, i.e., no phase jump, then we recover the standard diffraction result

$$A_2(\theta) = \frac{\sin(\pi D\theta/\lambda)}{\pi D\theta/\lambda} D,$$

which is a single peak at the origin. However, if the total phase step is $\phi = \pi$, i.e., a half-wavelength, then we get

$$A_2(\theta, \phi = \pi) = \frac{\sin^2(\pi D\theta/2\lambda)}{\pi D\theta/2\lambda} D,$$

which is a pair of peaks (speckles), each similar in width to the original single peak, and separated by about twice that width. This means that instead of having just one peak due to the star, we now have two adjacent images. Smaller phase differences will produce intermediate results, i.e., a pair of speckles but with unequal intensities and smaller separation. Considering more subdivisions of the pupil we can produce a system of speckles with the previously described characteristics. The random phase variation of incoming wavefront makes also the intensity of speckle to vary in a random way. The positions of the speckles are a function of the wavefront perturbations and the wavelength of light observed. In addition to

this, tiny amplitude perturbations in the incoming wavefront's electric field also translate into speckles in the image plane (with antisymmetric behavior), though these are generally much fainter than the ones due to phase errors. Speckles can originate by atmosphere (short-lived speckles) but also by optics of telescope and instruments (long-lived speckles). Real mirrors have surface shape errors that get smaller in proportion to the size of region being considered. Thus the surface errors are not a white noise process. In any case the result is a weakened star image surrounded by a highly variable, non-smooth background against which one seeks to find very faint objects. In the "frozen atmosphere" approximation, the patches of density fluctuations are carried along by the wind, so a typical timescale for wavefront change is $\tau_0 = 0.31 r_0/V$ where V is a typical wind speed in the overlying atmosphere; if $V = 10\,\mathrm{ms}^{-1}$, then τ_0 is about 3 ms. Such speckle lifetimes vary from site to site and with different adaptive optics systems and instrument configurations, but they tend to be on the order of a few milliseconds to ten seconds, with some lasting many minutes at $0.5 < \lambda < 2.5\,\mu\mathrm{m}$. A ground-based image of a star therefore is made up of approximately $(D/r_0)^2$ speckles, churning on a timescale of τ_0, and spread over an angular diameter on the sky of about λ/r_0 or 1 arcsec in the visible, independent of telescope diameter.

Generally, speckles behave in such a manner that they do not obey Poisson statistics, and they represent a noise source several orders of magnitude larger than the shot-noise behavior of the underlying perfect PSF (Racine et al. 1999). Speckles are also highly evanescent. They exhibit a correlated noise behavior (Soummer et al. 2007), and as such, one cannot simply let them average out into a smooth background against which one can pick out a much fainter source (as one does for objects fainter than the uniform sky background in many deep astronomical observations). Another aspect of this speckle noise is that it seems largely independent of the size of the telescope aperture. In summary, speckles are strictly related to the starlight. They are correlated, and blurring them out with long integration times or by using broad-band observations over large wavelength ranges results in no increase in sensitivity to objects fainter than the speckle background.

4.4.2 High-Contrast Imaging

The goal of high-contrast imaging is primarily to discover and characterize extrasolar planetary systems. High-contrast observations in optical and infrared astronomy are defined as any observation requiring a technique to reveal a celestial object of interest that is in such close angular proximity to another source brighter by a factor of at least 10^5, that optical effects hinder or prevent the collection of photons directly from the target of observation.

In 1919 the famous expedition of Sir Arthur Eddington to the island of Principe took pictures of several bright stars in the Hyades within a few arcseconds of the Sun's limb. This expedition and those observations confirmed the prediction of general relativity showing that the apparent position of those stars was distorted

by almost 2 arcsec due to the gravitational influence of the Sun (Dyson et al. 1920). These observations, together with Lyot's coronagraphic observations of the Sun's corona (Lyot 1939), could be considered the first high-contrast observations in the history of astronomy (Oppenheimer and Hinkley 2009).

Exoplanets are much fainter than their parent star and separated by very small angles, so conventional imaging techniques are totally inadequate, and new methods are needed. We have just seen the problematics that hamper the direct imaging of faint companions. We know that we can mitigate these problems. First of all it is important to use large telescopes to acquire higher angular resolution. Successively we have also to restore the diffraction figure of these telescopes equipping them with Adaptive Optics (AO) modules. But all this is not enough. Just to make a summary of the previous sections, the PSF of the telescope is the convolution of two main components:

- Flat but not infinite wavefront (limited pupil size). This generates a diffraction peak (Airy disk) that is usually expressed in units of λ/D.
- Perturbations with respect to flat wavefront. This generates speckles that are diffraction images offset with respect to the center of image.

So at first we need to eliminate the glare of the star hiding or eliminating the diffraction peak obtained through the AO by means a coronagraphic device (see Sect. 4.6). But the mere elimination of the diffraction peak does not assure that we will reach the contrast necessary to reveal the faint companion that is drowned in that variable, not smooth background, due to speckles. Eventually we need methods that are able to suppress or mitigate the speckle noise. With this aim several methods could be exploited (see Sect. 4.7) starting from observation strategies like in the case of angular differential imaging (ADI) or new concepts in the building of instrumentation (see Sect. 4.8) and data reduction.

4.5 Adaptive Optics

In 1953 the astronomer Horace W. Babcock had an idea: *If we had the means of continually measuring the deviation of rays from all parts of the mirror, and of amplifying and feeding back this information so as to correct locally the figure of the mirror in response to the schlieren pattern, we could expect to compensate both for the seeing and for any inherent imperfection of the optical figure* (Babcock 1953).

A time lap of 30 years was necessary until technology had matured to a level supporting a practical implementation of Babcock's concept. This concept is now called adaptive optics. The first astronomical AO instrument COME-ON was tested at the 1.52-m telescope[3] (Merkle et al. 1989; Rousset et al. 1990) of the Observatoire

[3]US military started to invest in AO in the 1970s and had commissioned the first adaptive optical system in 1982 (Davies and Kasper 2012).

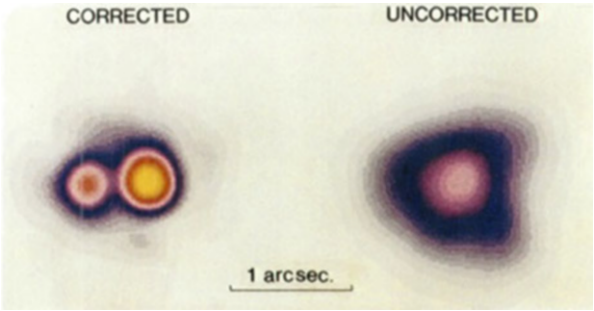

Fig. 4.10 First image taken with the adaptive optics prototype "COME-ON" system mounted at 1.52 m telescope of the Observatoire de Haute Provence. The binary star γ_2 And was observed in the K band (separation $0.5''$) by Rousset et al. (1990) and reprinted here with permission from the authors

de Haute-Provence (see Fig. 4.10) and later installed at ESO's 3.6-m telescope on La Silla in Chile (Rigaut et al. 1991). A summary of the history of AO was presented by Beckers (1993).

Before jumping into in the AO realm, it could be useful to describe in some details how the Earth's atmosphere comes into play and give some definitions. Ideally we can define the atmosphere like a slab of homogeneous gas that behaves like a laminar fluid. In this case the slab has a refraction index n_2 different from the empty space refraction index n_1 from which the wavefront is coming. As the incoming wavefront impinges on the Earth's atmosphere it is deviated at an angle given by Snell's law. In this case this is the only modification of the wavefront. As it is simple to imagine, the real atmosphere is really more complicated. At first because it is not homogeneous, then because it is a turbulent fluid and the air refraction index is actually a function of the local temperature T, pressure P, and of the wavelength under consideration (Cauchy's equation):

$$n - 1 = \frac{77 \times 10^{-6}}{T} \left(1 + 7.52 \times 10^{-3} \lambda^{-2}\right) \left(P + 4810 \frac{P}{T}\right).$$

There are small patches of atmosphere (bubbles of air) in which the refraction index is quite constant and the dimension of these patches is given by the Fried parameter. This parameter is the spatial scale inside which the wavefront statistically varies less than 1 rad of RMS phase aberration and it represents the average size of a turbulent cell. The Fried parameter is a function of the wavelength, of the refraction index of air, and the altitude (Beckers 1993):

$$r_0 \propto \lambda^{6/5} \left[\int_0^\infty C_n^2(h) dh\right]^{-3/5}.$$

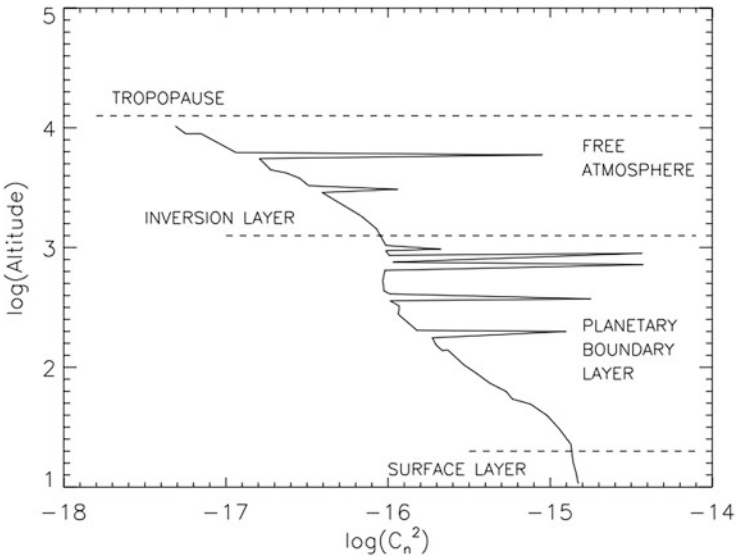

Fig. 4.11 The variation of the function $C_n^2(h)$ ($m^{-2/3}$) with the altitude (m). In the figure are also indicated the three layers in which the atmosphere is subdivided

The real atmosphere could be subdivided into function of the altitude in three different layers (see Fig. 4.11). The lower is the ground layer that is in between the surface of the Earth and 10 m of altitude. The ground layer is characterized by the exchange of heat between the surface and the air developing humidity and air stream that make this layer a turbulent one. Between 10 m from the surface up to about 1 km (the altitude of the inversion layer) there is the boundary layer that is characterized by the presence of ascending streams due to the difference of temperature that moves also air bubbles in a turbulent way. Over 1 km of altitude there is a free atmosphere that is characterized by turbulent motion of air bubbles, the dimensions of which are functions of temperature, wind, and the refraction index. All this structure is described in the $C_n^2(h)$ function that appears in the previous equation.

Other important parameters describing the status of the atmosphere are the isoplanatic angle and the coherence time. The isoplanatic angle, $\theta_0 \propto (\cos\gamma) r_0/h$ with h denoting the characteristic height of the turbulence, describes the angle out to which optical path variations deviate by less than one radian RMS phase aberration from each other. Given a certain correction direction, θ_0 provides the maximum angular radius from this direction at which reasonably good correction is achieved. θ_0 is typically on the order of a few arcseconds at visible wavelengths and strongly depends on the height distribution of the turbulent layers. The coherence time, $\tau_0 \propto r_0/v_{\mathrm{WIND}}$ with v_{WIND} denoting the average wind speed, describes the time interval up to which optical path variations deviate by less than one radian RMS phase aberration from each other. τ_0 therefore defines the required AO temporal correction bandwidth, which is typically a few milliseconds at visible wavelengths.

Due to the fragmentation in turbulence cells of the atmospheric layers each part of the incoming wavefront is perturbed when passing through these cells. The result is that it is no longer a flat wavefront but is deformed (high orders aberrations) and tilted. Both tip–tilt and high orders have the same final effect: a loss of angular resolution of the telescope, smearing and blurring of the diffraction image.

The primary purpose of AO in high-contrast imaging has been the production of diffraction—limited images—to which diffracted light suppression techniques can be applied. The first step in AO is achieved with a tip/tilt system, or fine guidance tracker, the lowest-order correction possible. This system corrects for large movements (up to a few arcseconds) of the stellar PSF due either to atmospheric variations, wind, or vibration in the telescope. Once the image of the star has been stabilized by the tip/tilt system, the remaining correction to the wavefront is achieved with a deformable mirror (see Fig. 4.12). The AO system tries to regulate the optical path variations (wavefront) by measuring the deviations using

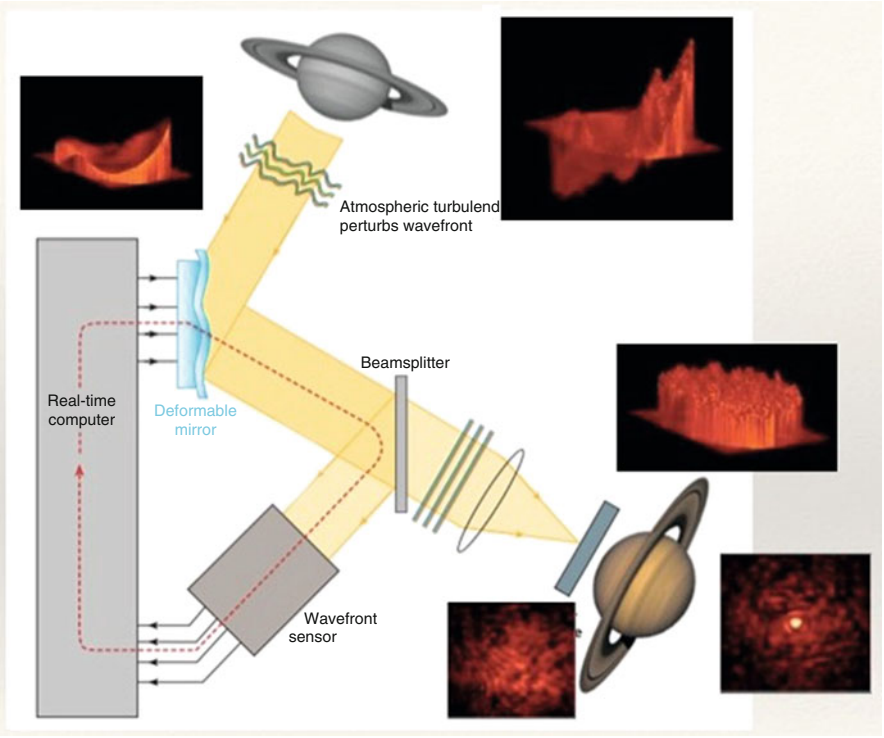

Fig. 4.12 The principle of working of an adaptive optics system. In clockwise direction it is possible to see also how the image appeared before correction, the appearance of the wavefront as measured by the wavefront sensor, the spectrum of strokes necessary to modify the shape of the deformable mirror, how the corrected wavefront appears in the focal plane, and finally how the image appears on the focal plane after the correction

a wavefront sensor (WFS), calculating an appropriate correction, and applying this correction to a deformable mirror (DM). This feedback loop is carried out several hundred times a second in order to comply with the temporal bandwidth requirement set by τ_0. The size of the resolution elements of the WFS (subapertures) and the DM (actuators) projected on the telescope entrance aperture should approximately match with r_0. The depicted setup uses WFS measurements of a single guide star to correct the wavefront in its direction. It is called single-conjugate adaptive optics (SCAO). SCAO suffers from image degradation over the field of view (FOV) set by θ_0. There is a wealth of other concepts involving multiple guide stars and/or DMs as well as more complex control strategies. Because AO wavefront sensing requires a light source above the atmosphere and near to the astronomical object, it is very often photon starved, and good sensitivity of the WFS is essential. In many cases, a visually bright enough natural guide star (NGS) is not available. In this case the solution is to create one's own guide star where needed.

The efficiency of an AO system is measured by the Strehl ratio. The Strehl ratio is the ratio of the peak intensity in a real image to that of a perfect image made with the same imaging system's fundamental parameters. Also approximated by $S \propto \exp(-\sigma^2)$ where σ is the root mean square of the wavefront error in radians, when $\sigma \ll 1$. Most AO systems, however, only operate at Strehl ratios of around 20–60 %, and the resultant data produced by, for example, a coronagraph or interferometer behind an AO system are generally limited by the remnant speckle noise. AO alone cannot solve the problem of high-contrast observations. AO can also be used to control speckle noise.

The main elements of an AO system are the WFS, a real-time computer, and a deformable mirror. A WFS is any device that allows us to measure the wavefront. It provides a signal with which the shape of the wavefront can be estimated with sufficient accuracy. It generally incorporates a phase-sensitive optical device or sensing scheme and a low-noise, high quantum efficiency photon detector. The main sources of uncertainty in a WFS are photon noise, chromaticity, aliasing, time delay, scintillation, and non-common-path errors. Photon noise is obviously fundamental; it can be minimized by using a bright star and large values of r_0 and θ_0 on the ground, or their equivalent in space (e.g., from polishing errors and thermal drift). Chromaticity arises when the WFS works at different wavelengths of those of science imager; it can be minimized by using the same wavelength band for both purposes. Aliasing arises when the WFS is sensitive to, but cannot distinguish between, a spatial-frequency mode of the wavefront and an odd harmonic. Time delay occurs because a detector must integrate a signal for a finite length of time before it can be read out, and in addition the servo system has a finite bandwidth, which effectively adds more time delay to the correction signal. The choice of integration time depends on photon rate, the desired signal to noise ratio, and detector read-out noise. Three different types of WFSs are currently used in AO: the Shack–Hartmann WFS, the Pyramid WFS, and the Curvature WFS (see Fig. 4.13). We briefly describe only the first two WFSs.

The Shack–Hartmann WFS is a simple system, used at many telescopes. In this system a beam splitter taps off part of the light, and a lens forms an image of the

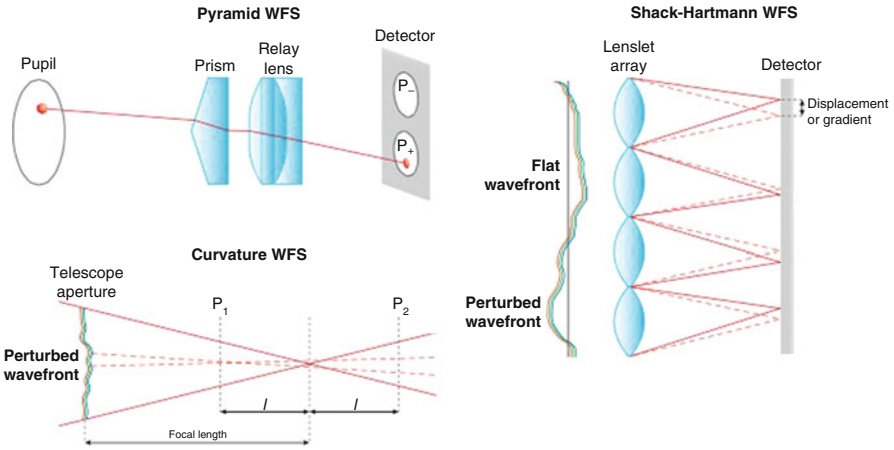

Fig. 4.13 Schemes that illustrate the working principle of the main wavefront sensor utilized in adaptive optics systems. The figure is taken from Davies and Kasper (2012) and reprinted with permission from the authors

pupil. This pupil plane is filled with a large number of small lenses. Each lenslet forms an image of the star onto a position-sensitive detector (a multiple quad cells or a CCD). A local tilt of the wavefront will produce a shift in the star image. So the positions of these spots represent the average wavefront slope or gradient over the subaperture and image position measurements can be converted to local wavefront slopes. This information can then be used to drive a DM, and a closed-loop control established.

The Pyramid WFS (Ragazzoni 1996) very much represents the Shack–Hartmann WFS when the pyramid (or prism in Fig. 4.13) is modulated. When an aberrated ray hits the prism on either side of its tip, it appears in only one of the multiple pupils. The intensity distributions in the multiple pupil images are therefore a measure of the sign of the ray's slope. If the prism modulates, the ray will appear in either of the pupil images depending on the modulus of the local slope. Thus, the intensity distribution integrated over a couple of modulations also measures wavefront slopes in the pupil.

The objective of the DM is to correct for the optical path differences introduced by the turbulent atmosphere. It usually consists of an array of actuators that are connected to a thin optical surface that deforms itself under the expansion of the actuators. The most important parameters for a DM are stroke, response time, spacing, and number of actuators. The largest AO DMs are the adaptive secondary or deformable secondary mirrors (DSMs). This kind of DMs is mounted on the 6.5-m multiple mirror telescope (MMT) and on the large binocular telescope (LBT) (Esposito et al. 2010). The more common DMs are medium-sized piezo DMs with an actuator spacing of several millimeters. They have less stroke than DSMs, but at about 10 μm peak-to-valley, it is still sufficient for 8–10 m class telescopes

(Davies and Kasper 2012). The DM of SPHERE belongs to this class of devices (Beuzit et al. 2008). Quite recently, micro-optical-electrical-mechanical systems (MOEMSs) have emerged as an alternative. They use electrostatic or voice-coil actuation mechanisms and are produced using standard semiconductor fabrication technologies. GPI mounts this kind of DMs (Macintosh et al. 2008).

AO systems need sufficiently bright ($V \sim 15$ mag) guide stars within θ_0 of the astronomical target in order to have performant analysis of the wavefront by the WFS. Actually, bright stars are not so common in the sky and this limits the sky coverage that can be attained using NGSs. To overcome this problem, the US Military first suggested using laser bacons to create artificial sources (Beckers 1993; Davies and Kasper 2012). Later, in an independent way Foy and Labeyrie (1985) proposed the same concept in the astronomical context. Laser guide stars (LGS) are based on two mechanisms. The first considers the Rayleigh scattering in the dense regions of atmosphere up to altitudes of about 30 km above the ground, while the second the resonance fluorescence of sodium atoms that are concentrated in a layer at about 90-km height. The first successful tests of a Rayleigh LGS were performed at the Starfire Optical Range 1.5-m telescope (Fugate 1992); the first astronomical LGS AO systems were installed at the Lick (Max et al. 1997) and Calar Alto (Eckart et al. 2000) observatories in the mid-1990s. Still it took another 10 years until the technology had matured enough to be installed at 8–10 m class telescopes such as Keck II (Wizinowich et al. 2006), VLT (Bonaccini Calia et al. 2006), Gemini North (Boccas et al. 2006), and Subaru (Hayano et al. 2010).

Also if the use of LGS improves the sky coverage with respect to NGSs, there are still limitations, notably the cone effect or focal anisoplanatism. Due to the finite distance between telescope and LGS, the backscattered beam does not sample the full aperture at the height of the turbulent layers. For a sodium LGS on an 8–10 m class telescope, the cone effect reduces the Strehl ratio obtainable by a factor of 0.85 in K-band and 0.6 in J-band (Davies and Kasper 2012). This moderate impact would become devastating for the next generation of ELTs, limiting the maximum achievable K-band Strehl to only 15 % and much less at shorter wavelengths. In order to overcome the cone effect, measurements from several LGSs can be combined to fully reconstruct the turbulence column in the direction of the astronomical target. The separation of the LGSs can be significantly larger than the isoplanatic angle θ_0, but their beams should still overlap at the highest turbulent layer in order to prevent there being unsampled turbulence. This technique is named laser tomography adaptive optics (LTAO). There are also other different kinds of AO systems that are deputed to probe different turbulent layers with different combinations of NGSs or LGS. The ground layer adaptive optics (GLAO) technique corrects only the modification of the incoming wavefront due to the ground layer that is very close to the telescope mirror. In order to do this, GLAO exploits different guide stars well separated so that the light coming from all the off-axis sources passes through the ground layer. The correction is done by combining all the wavefront of all the reference stars. The technique allows us to have a larger FOV (up to $\sim$10 min) bringing as advantages that the correction is

no longer dependent on the distance of the guide star and that there is a greater probability in having a better PSF reference.

The multi conjugate adaptive optics (MCAO) is based on the use of different reference stars or LGSs in different positions in a way that the different columns of atmospheric turbulence that they probe are partially superimposed at high altitude and fully superimposed at lower altitude. In this way it is possible to analyze two different turbulence layers at different quotes while using separate DMs, one conjugated to each layer, to correct them. MCAO was first discussed by Dicke (1975) and Beckers (1998), and the first systems were developed in the context of solar astronomy. Classical astrophysics had to wait for the MAD (MCAO demonstrator) at the VLT (Marchetti et al. 2008). The system used NGSs to deliver near-IR resolution down to 0.1 arcsec across a 120-arcsec field. Technically, the performance of an MCAO system is limited by a number of practical considerations. From a scientific perspective, arguably the most important issue is the FOV that can be corrected.

The last system that we discuss is the extreme adaptive optics (ExAO) concept that is used by most of the new generation of high-contrast imagers. The extreme AO (XAO) concept is not different from conventional SCAO, but the technical implementation is highly challenging because the main aim of this system is to provide exceptionally high performance on bright ($V < 10$ mag) stars trying to reach a Strehl ratio in excess of 90 % in the H band. The system is built in order to fully correct both atmospheric and instrumental perturbations. Image and pupil instability due to thermo-mechanical effects and non-common-path errors must be actively minimized using additional internal control loops. The first XAO system was PALM-3000 for the 5 m telescope at the Palomar Observatory with more than 3000 actuators on the DM (Dekany et al. 2011). Both GPI (Macintosh et al. 2008) and SPHERE (Beuzit et al. 2008), the newer high-contrast imagers, are using XAO (see Sect. 4.8).

4.6 Coronagraphy

Originally invented in 1930 by Lyot (1939) to study the Sun, a coronagraph is a telescope designed to block light coming from the solar disk, in order to see the extremely faint emission from the corona. It is constituted by an occulting disk in the focal plane of a telescope or out in front of the entrance aperture that blocks the image of the solar disk, and various other features, to reduce stray light (see Fig. 4.14).

With time the technique has been applied to the stellar observation. Coronagraphy is an important technique for high-contrast imaging where a very faint object is to be observed in the halo of the glare of the host star. We have seen in the previous sections that the diffraction of a point source due to the side of the telescope aperture generates an Airy pattern. For a ground-based telescope this pattern is obtained after correction of the incoming wavefront thanks to the AO. But in any case this pattern

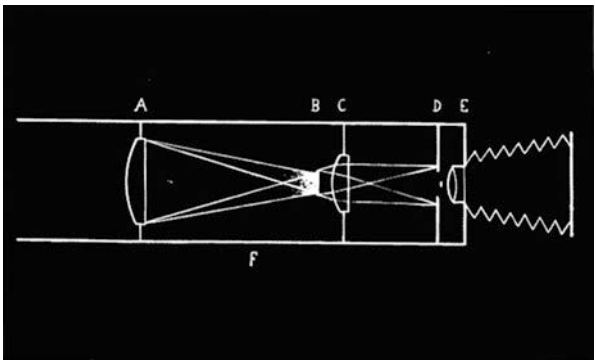

Fig. 4.14 The original scheme of the Lyot's coronagraph (Lyot 1939)

is many orders of magnitude brighter than any exoplanet or faint companion. The coronagraph suppresses the diffraction peak.

Before describing the principle of coronagraphs, it is necessary to introduce a way to describe in general the propagation of light through the optics involved in the instrumentation. We know that it is possible to describe it by means of the geometric optics or by physical optics. The former describes ideal diffraction-limited optical situations in the limit of zero wavelength, so no diffraction phenomena are seen. In other words, light rays are all you need to describe what you see. The physical optics describes ideal diffraction-limited optical situations (coronagraphs, interferometers, gratings, lenses, prisms, radio telescopes, eyes, etc.) considering that all photons start from the same atom, and follow the same many-fold path to the detectors, with the same amplitudes, phase shifts, and polarizations, and we will see a diffraction-controlled interference pattern at the detectors. Essentially, waves are needed to describe what you see. A conceptually elegant way to view the operation of an imaging system, including a coronagraph, is to take advantage of the Fourier-transform relation between the conjugate planes of an optical system and the fact that the Fourier transform of a product of functions is the convolution of the individual Fourier transforms; moreover, the Fourier transform of a convolution of two functions is the product of the individual Fourier transforms. In other words, $\mathrm{FT}(f \times g) = \mathrm{FT}(f) \cdot \mathrm{FT}(g)$, and also $\mathrm{FT}(f \cdot g) = \mathrm{FT}(f) \times \mathrm{FT}(g)$. Considering an optical system as a succession of conjugate planes, for example, a coronagraph applied in the focal plane of a telescope (see Fig. 4.15), we see that an ideal lens (or focusing mirror) acts on the amplitude in the pupil plane (telescope aperture, input pupil: plane 1), with a Fourier-transform operation, to generate the amplitude in the image plane (focal plane: plane 2). A second lens, after the image plane, would convert the image-plane amplitude, with a second Fourier transform, to the plane where the initial pupil is re-imaged (conjugate pupil plane, plane 3). A third lens after the re-imaged pupil would create a re-imaged image plane (conjugated focal plane, plane 4), via a third FT.

At each stage we can modify the amplitude with masks, stops, polarization shifts, and phase changes. These all go into the net transmitted amplitude, before the next FT operation. Referring to plane 1, we see that the input amplitude is $A_1(x_1)$,

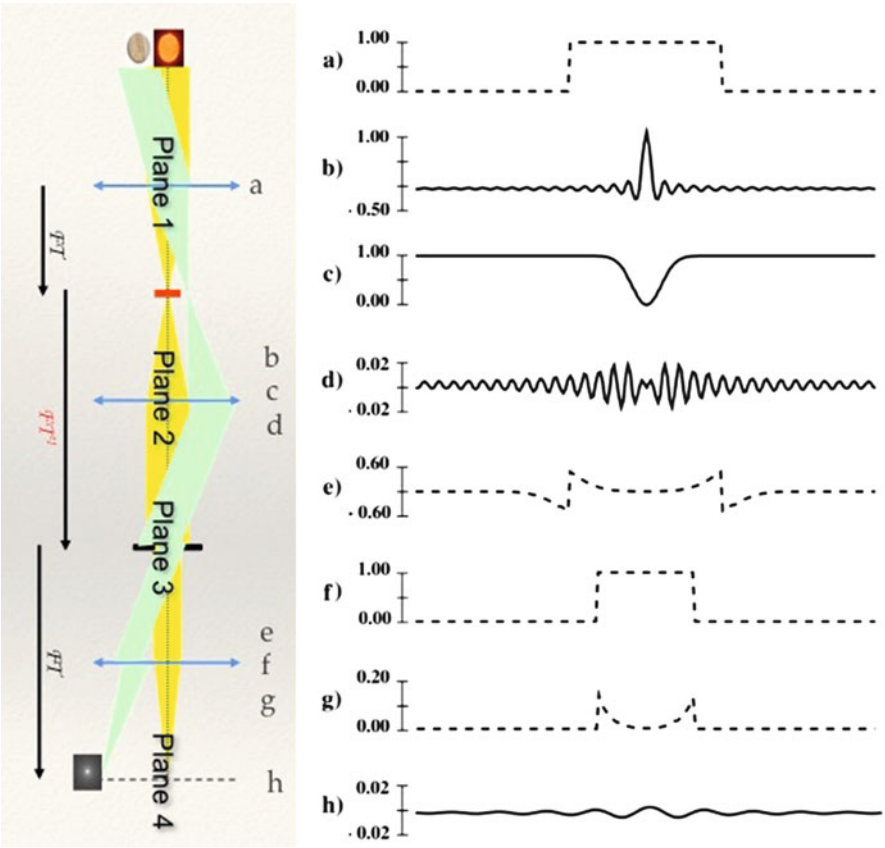

Fig. 4.15 Coronagraphy principle and Fourier optics. The optical scheme of a coronagraph with positions and electric field or stop profiles of (**a**) primary pupil for on-axis source, (**b**) image before image-plane stop, (**c**) image-plane stop, (**d**) image after image-plane stop, (**e**) pupil before Lyot stop, (**f**) Lyot stop, (**g**) pupil after Lyot stop, and (**h**) final on-axis image. Modified from Sivaramakrishnan et al. (2001)

the mask is $M_1(x_1)$, and the output is M_1A_1. At plane 2, the input amplitude is $\mathrm{FT}[M_1A_1](x_2) = \mathrm{FT}(M_1) \times [\mathrm{FT}(A_1)](x_2)]$. The mask is $M_2(x_2)$ and the output is $M_2 \cdot [\mathrm{FT}(M_1) \times \mathrm{FT}(A_1)](x_2)$. At plane 3 the input is the FT^{-1} of the plane 2 output. We multiply the mask M_3 times that function, and apply the convolution rules again. This gives the output from plane 3 as $M_3 \cdot [\mathrm{FT}^{-1}(M_2) \times (M_1 \cdot A_1)]$. At plane 4 the input is the FT of the plane 3 output. Substituting and simplifying we get the field at plane 4 to be $\mathrm{FT}(M_3) \times [M_2 \cdot \mathrm{FT}(M_1A_1)]$.

The value of this description will become clear when we look at individual coronagraph designs. The band-limited mask design (Kuchner and Traub 2002) will show clearly how this picture, and in particular the expression for the output of plane 3, can bypass difficult integrals to give a clear physical picture. Summarising

we have that the electric field incident on plane 1 is $A_1(x_1) = e^{2i\pi x_1\theta_0/\lambda}$ for a point source located at angle θ_0. We may describe in a mathematical way the input pupil of a system by a mask function $M_1(x)$ that has a value 1 where it is transparent and 0 otherwise. This kind of function is known as a rectangular function (stage a in Fig. 4.15):

$$\begin{aligned} \text{rect}(x, D) &= 1 \quad -D/2 \leq x \leq D/2 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

A lens in plane 1 produces an electric field $A_2(\theta_2)$ incident on plane 2 (stage b in Fig. 4.15)

$$A_2(\theta_2) = \int_{-\infty}^{+\infty} M_1(x_1)A_1(x_1)e^{2i\pi\theta_2 x_1/\lambda}dx_1.$$

Likewise, a lens in plane 2 produces an electric field $A_3(x_3)$ (stage d in Fig. 4.15) incident on plane 3

$$A_3(x_3) = \int_{-\infty}^{+\infty} M_2(\theta_2)A_2(\theta_2)e^{2i\pi x_3\theta_2/\lambda}d\theta_2,$$

where $M_2(\theta_2)$ (stage c in Fig. 4.15) is a mask on the output side of plane 2. Finally, a lens in plane 3 produces an electric field $A_4(\theta_4)$ (stage h in Fig. 4.15) incident on plane 4

$$A_4(\theta_4) = \int_{-\infty}^{+\infty} M_3(x_3)A_3(x_3)e^{2i\pi\theta_4 x_3/\lambda}dx_3.$$

Where $M_3(x_3)$ (stage f in Fig. 4.15) is a mask on the output side of plane 3.

The coronagraph is shown schematically in the left-hand part of Fig. 4.15. It uses two masks (the occulting mask c and Lyot stop f) to achieve the suppression of starlight. In the first stage, an image of the target star is formed at the center of a circular, opaque focal plane mask. The starlight is largely absorbed by this mask, but also diffracts around it. The beam after the focal plane mask is then brought back out of focus and an image of the telescope pupil is formed. In this plane, much of the residual starlight has been concentrated into a bright outer and inner ring around the conjugate location of the secondary telescope mirror (if there is one). This concentration of the starlight is critical to understanding a coronagraph and is due to the diffraction caused by the focal plane mask. The light of any object that is not significantly diffracted by the focal plane mask will be distributed in this pupil image evenly. In this manner, the coronagraph has effectively separated the light of the primary star, by using diffraction, away from that of a faint object next to the star. It can thus be filtered out further without greatly affecting the light from the fainter object. Indeed, the Lyot mask is placed in this pupil image. It downsizes

the telescope aperture slightly, while slightly increasing the size of the secondary obscuration, simply in order to block the bright concentrated rings of the central star's light. Finally, optics form an image after this Lyot mask, where the overall intensity of the central star has been reduced by more than 99 %, while a neighboring object will only be affected at the few percent level.

There are many concepts for coronagraphs, most of which have been invented specifically for exoplanet observations. This is a very exciting field, with new ideas coming along at a fast pace, very little of which can be found in any optics textbook.

Guyon et al. (2006) break all the coronagraphs down into four broad categories: interferometric coronagraphs, pupil-apodization coronagraphs, amplitude-based Lyot coronagraphs, and phase-based Lyot coronagraphs, listing 18 separate coronagraphs within these categories. Each coronagraph is characterized by several parameters.

The main parameter is the inner working angle (IWA). Although this is not normally a sharp step function, as discussed in Guyon et al. (2006), the flux from the planet is rapidly attenuated and/or the flux from the star rapidly increases as the IWA is approached, yielding a sharply falling detection sensitivity. Thus, the IWA is often treated as a hard lower limit in modeling of coronagraphic missions. Currently, the most mature coronagraph concepts assume an IWA of $\sim 4\lambda/D$; some newer, but less-proven, concepts suggest that an IWA of $\sim 2\lambda/D$ can be achieved.

A second parameter is the throughput of the planet or the fraction of planet's light that survive to the suppression due to the coronagraph. Most of the system has a throughput of about 80 %.

A third key parameter is the sensitivity of the coronagraph to low-order wavefront errors. Although all coronagraphs are sensitive to mid-spatial-frequency aberrations, some designs have additional sensitivity to low-order aberrations such as tilt (image position), focus, astigmatism, etc. These aberrations arise easily due to drift and misalignment, and their effects can quickly obscure or mimic a planetary signature.

A fourth parameter is the chromaticity or the ability of the coronagraph to suppress the light on a large range of wavelength. Low chromaticity is better.

We can proceed giving the description of some kinds of coronagraph considering only the most popular example of amplitude- and phase-based coronagraph. The interested reader could find a deeper description of all types of coronagraph in Guyon et al. (2006) and Sivaramakrishnan et al. (2001).

The most mature family of coronagraphs are those evolved from Lyot's original solar coronagraph, comprising a mask at an image of the star, and another mask at a later image of the entrance aperture (pupil). The most important variant of amplitude-based Lyot coronagraph concept is the band-limited coronagraph (BLC), devised by Kuchner and Traub (2002). This approach uses a focal plane mask with carefully tailored transmission profile, to almost perfectly confine the residual diffracted light to a finite outer region of the pupil. Typically, the BLC has an IWA of $2 - 4.5\lambda/D$, a throughput of 20–40 % and, for an IWA $> 4\lambda/D$, it has good robustness against low-order wavefront errors. Another Lyot variant is the Apodized-Pupil Lyot Coronagraph (APLC), which instead modifies the entrance pupil with a tapered

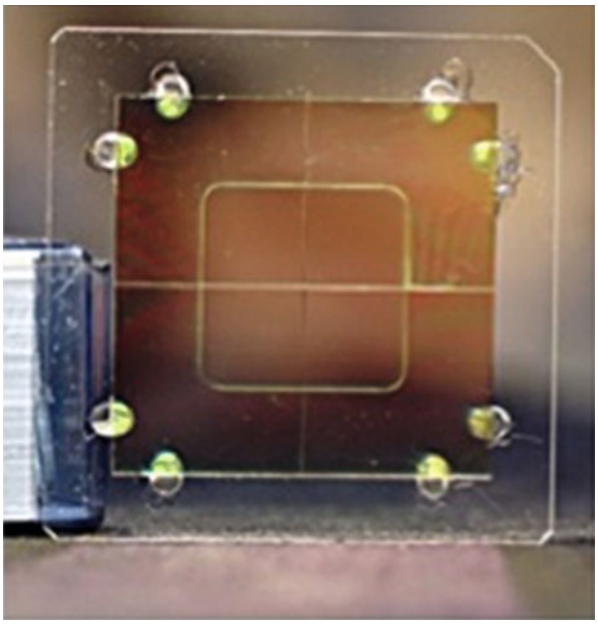

Fig. 4.16 The 4QPM phase mask of SPHERE (Boccaletti et al. 2004)

transmission mask.[4] This coronagraph class is capable of similar performance to the BLC in monochromatic light but is challenging to manufacture for broadband use. Manufacture of the necessary mask is a significant technical challenge. This coronagraph type is easily adapted to conventional telescopes with on-axis secondary mirrors, and has been selected by the major next-generation ground-based coronagraph projects (e.g., SPHERE and GPI).

An alternate approach to the Lyot coronagraph is to induce a phase shift in part of the starlight in the focal plane, creating destructive interference for on-axis starlight within the telescope pupil and effectively blocking the star (e.g., Boccaletti et al. 2004). In theory this technique could remove all of the starlight, while having no effect on anything else in the FOV. The best-known example of the phase-based Lyot coronagraph concept is the 4-quadrant phase-mask (4QPM) coronagraph. The phase mask that is positioned in the focal plane is four contiguous quadrants of transparent material onto which the star is focused at the symmetry point, with adjacent quadrants differing in optical thickness by a half-wavelength (see Fig. 4.16). The transmitted beam will be nulled out on its axis, but an object imaged mainly in one of the quadrants will be transmitted.

[4] Apodization is the technique that allows reduction of the sharp discontinuity in the transmitted wavefront at the edge of the pupil reducing the amplitude of the secondary maxima of the diffraction figure.

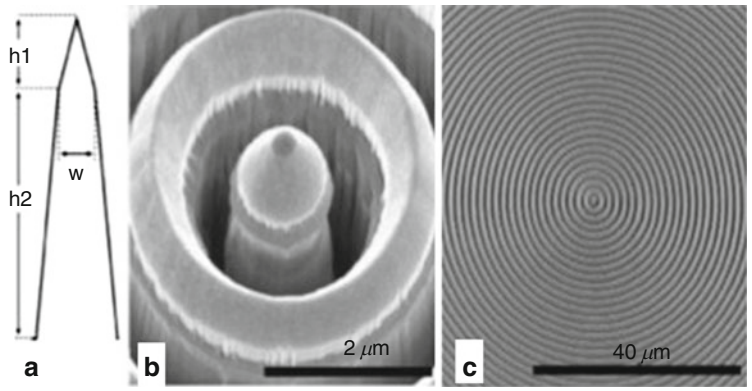

Fig. 4.17 The annular groove phase mask (AGPM) mounted on NACO at VLT. The figure was taken from Mawet et al. (2013) and reprinted with permission from the authors

Such coronagraphs often have near-ideal theoretical performance with high throughput and an IWA of $1\lambda/D$. However, they are also extremely sensitive to position in the focal plane, with performance degrading rapidly for even small tip/tilt errors or stars that are partially resolved. The 4QPM remains popular in various low-performance European mission concepts and ground-based instruments like SPHERE (contrast levels of 10^{-6} to 10^{-8}).

A recently proposed variant of the 4QPM is the Optical Vortex Coronagraph (see Fig. 4.17), which uses a spiral-staircase phase mask generating a longitudinal phase delay by operating on both polarizations. The center of the optical vortex is a phase singularity in an optical field, which generates a point of zero intensity, resulting from a phase screw dislocation of the form $e^{il_P\phi}$, where l_P, the number of half-wavelengths per turn, is called the topological charge, and ϕ is the azimuthal coordinate (Traub and Oppenheimer 2010).

When centered on the diffraction pattern of a star seen by a telescope, optical vortices affect the subsequent propagation to the downstream Lyot stop by redirecting the on-axis starlight outside the pupil. The annular groove phase mask (AGPM) has been mounted on NACO at VLT (Mawet et al. 2005, 2013) obtaining good results. The advantages of the AGPM coronagraph over classical Lyot coronagraphs or phase/amplitude apodizers are small IWA down to $0.9\lambda/D$ (e.g., $0''.09$ in the L band at the VLT, slightly smaller than the diffraction limit); clear 360° off-axis FOV/discovery space; outer working angle set only by the instrument and/or mechanical/optical constraints; achromatic over the entire working waveband (here L band); high throughput (here $\sim$88 %); and optical/operational simplicity.

A different type of starlight suppression is as old as the human hand blocking the bright star before it reaches the telescope. In 1960, Lyman Spitzer proposed combining a telescope and starshade in space for discovery of planets (Spitzer 1960). The size of the shade and the inter-spacecraft separation were enormous

and thus impractical, but over many years of refinements in starshade design have reduced the required starshade dimensions and improved the level of suppression. The most recent work has focused on optimizing the shapes of serrated-edge binary masks. Petal shapes have been found (e.g., Vanderbei et al. 2007) that permit operation at IWA < 100 marcsec at wavelengths from 0.5 to 1.1μm, using a shade with a nominal diameter of 40 m at a separation from the telescope as small as 40,000 km. It can observe the planet in the entire passband from 0.5 to 1.1 μm in a single integration, whereas internal coronagraphs typically are limited to 20 % passband slices at a time.

The appeal of external coronagraphs lies with their potential to circumvent many of the light suppression problems faced by internal coronagraphs by instead blocking the stellar light with a free-flying occulter located between 20,000 and 70,000 km (set, respectively, by Fresnel vs. Fraunhofer diffraction) from the telescope. This would allow the use of a generic diffraction-limited visible-light telescope.

The main drawback of the external occulter approach lies in its operational complexity relative to a single spacecraft, in fact it is necessary to control two vehicles (the telescope and the occulter) and the pointing to a target requires aligning the two spacecrafts. When many targets must be searched in a survey mode to find planets, travel time will limit the planet observing cadence; but this is offset by high telescope throughput and alternate observing strategies. Also, during travel time, the telescope conducts other astronomical observations. Clearly, a telescope in low-Earth orbit will not work, but a drift-away or L2 orbit would be feasible, assuming that the positioning control can be accomplished. The telescope-facing side of the occulter should look dark compared to the planet, so it must face away from the Sun. Thus we require that the angle between the occulter and the Sun be less than about 90°.

There are several challenges to working with coronagraphic data. Most notably, the precise position of the star behind the coronagraphic mask is difficult to measure. The point, after all, is to get rid of the star. For example, the relative position between the host star and any companions is required to establish physical association and to study orbital motion. A valuable solution to this problem was put forth by Marois et al. (2006a) and Sivaramakrishnan and Oppenheimer (2006), in which a periodic grid or a sinusoidal ripple on the wavefront is inserted at the telescope pupil (see Fig. 4.18). This causes four fiducial images of the occulted star to appear at known locations relative to the star outside the coronagraphic mask. The intersection of the two lines specified by spots on opposite sides of the star determines the star position with an accuracy that can be chosen for a given system or scientific goal. This technique is essentially the intentional introduction of permanent, well-understood speckles into the image that precisely locate the star. In addition, these calibrator spots also have a known brightness, based on the design of the grid, allowing accurate relative photometry between an occulted and an unocculted object within the coronagraphic image.

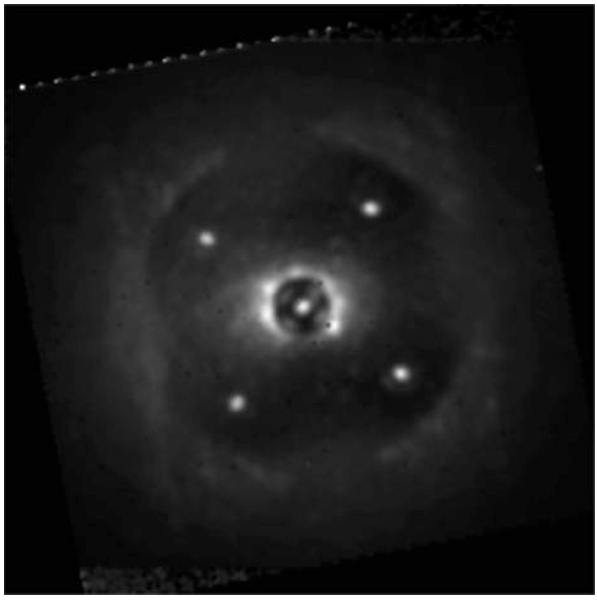

Fig. 4.18 Satellite spots (or waffles) created by sinusoidal ripple on the wavefront by the AO system of SPHERE

4.7 Speckle Suppression

The observation of stars with the aim to find a close companion is performed with telescopes equipped with AO modules with the add of coronagraph in order to eliminate the diffraction peak reconstructed with AO module. Again we know that this is not sufficient. In fact the image taken in this way is affected by image artifacts caused by quasi-static optical aberrations within the telescope, AO system, coronagraph, or camera that is not correctly calibrated or corrected. These aberrations produce slowly evolving speckle patterns. Several techniques have been developed to overcome these speckle patterns. If the wavefront sensing occurs in a plane that is different from the science focal plane, and if the speckles from the atmosphere and telescope pupil have been reduced in the wavefront sensing plane, then a residual wavefront deformation could probably remain in the science plane owing to a non-common-path problem. Especially at large separations (>0.5 arcsec), the main source of speckles are the surface errors on the telescope primary mirror and internal optics. These speckles can be substantial, and since they arise from the telescope optics themselves, they can persist for a long time, typically many minutes or more. Unfortunately, a telescope speckle has the same appearance as an exoplanet and persistent speckles can easily overwhelm a faint exoplanet image.

There are two approaches for mitigating these wavefront errors: control them or remove their manifestation, speckles, through special data collection and processing techniques. Even better, one could use both approaches, and several new systems are planning to do this. In this section we discuss some processing techniques that

allow us to tackle the problem due to speckle noise. These techniques exploit the characteristics of speckles and in particular their dependence by time, wavelength, and the orientation of the sources of wavefront error that cause them. A fourth quality of speckles that is different from some types of sources is that they are generally not polarized, because starlight is generally not polarized. So we can break these techniques into three main groups each based on the comparison of simultaneous images at different wavelengths or images of the same star at different orientations. In particular we discuss the ADI, the chromatic methods, and the polarization method.

4.7.1 Angular Differential Imaging

ADI is a high-contrast imaging technique that reduces quasi-static speckle noise and facilitates the detection of nearby companions. Another name for this technique is roll deconvolution, a method that has previously applied with success on the HST images (Marois et al. 2006b; Hinkley et al. 2007; Lafrenière et al. 2007b; Artigau et al. 2008).

The ADI technique can overcome internal speckles from the telescope by simply rotating the telescope about the line of sight, or at an altitude-azimuth telescope by allowing the rotating Earth to rotate the apparent sky (except on the celestial equator). A sequence of images is acquired with an altitude-azimuth telescope while the instrument field derotator is switched off. This keeps the instrument and telescope optics aligned and allows the FOV to rotate with respect to the instrument. Since the detector remains fixed with respect to the telescope, the non-common-path speckles also remain fixed on the detector. This setup improves the stability of the quasi-static PSF structure throughout the sequence. On the contrary there is a slight rotation of the FOV, and only of it, with respect to the instrument. The FOV rotation during an exposure causes that the companion PSFs are smeared azimuthally. This effect increases linearly with angular separation and slightly decreases companion's peak intensity. Short exposures and the use of an optimized aperture photometry box can minimize this effect. For each image, after data reduction and image registration of the whole sequence, a reference PSF obtained from other images of the same sequence is subtracted to remove the quasi-static structure. Given enough FOV rotation during the sequence, this subtraction preserves the signal from any eventual companion. All the image differences are then rotated to align the FOV and are median combined. Since the median is taken over a large number of images, the pixel-to-pixel noise (i.e., PSF, flat field, dark and sky Poisson noises, and detector readout noise) of the reference image is much less than that of any individual image (Marois et al. 2006a). Two methods can be used to subtract the quasi-static PSF structure, subtracting the median of all images or subtracting a reference PSF obtained from a few images acquired as close in time as possible. These two methods can be also combined to optimize speckle subtraction and minimize pixel-to-pixel noise. With the first method, the median of all the images is subtracted from

each individual image. For a sequence $I_i(t_i, \theta_i)$ of n reduced and registered images,[5] where t_i is the mean time of exposure i, and θ_i is the FOV orientation at time t_i; the first reference subtraction is

$$I_i^D = I_i - \text{med}(I_1, I_2, \ldots, I_n).$$

The second method consists of obtaining an optimized reference PSF for each image by median combining four images (two acquired before and two after) that show at least a 1.5 FWHM FOV orientation difference. Eventually difference images are rotated to align the FOV to that of the first image. Finally, a median is taken over all differences. The final combination step is thus

$$I_F^A = \text{med}[I_1^A, \text{rot}(I_2^A, \Delta\theta_{2-1}), \ldots, \text{rot}(I_n^A, \Delta\theta_{n-1})].$$

This technique offers a number of advantages over more classical ground-based observations, since the target observations themselves are used to construct a reference PSF. This means that the reference PSF has the same spectrum and brightness as the target and that no time is lost in acquiring reference observations of a different target. Ghost images from optical reflections and the sky flux are also removed by the subtraction. The detector flat-field errors are averaged, since the FOV is integrated with different pixels as it rotates on the detector. The ADI technique attenuates the PSF noise in two steps. First of all the subtraction of a reference image removes the correlated speckles. Secondly, the combination of all residual images after FOV alignment allows us to average the residual noise.

4.7.2 Chromatic Methods

Chromatic methods include simultaneous differential imaging (SDI) and spectral deconvolution (SD). SDI makes use of the different spectral energy distributions of the star and its companion while SD is an evolution of the speckles suppression technique described by Racine et al. (1999).

4.7.2.1 Simultaneous Differential Imaging

The simultaneous differential imaging (SDI) technique was described for the first time by Racine et al. (1999) and afterward applied and developed by different authors (Lenzen et al. 2004; Marois et al. 2005). The general principle of SDI is that two images of the star acquired simultaneously at close wavelengths can

[5] Before combining the images they should be reduced in the usual way: normalization of flat field and bad-pixel correction.

be subtracted to remove most of the stellar halo and speckle pattern but not the companion. In fact speckles are located at an angular distance from the star in proportion to minimize their wavelength. If images are taken at wavelengths very close together, so the PSF is as similar as possible in the two images, and the two images are also taken simultaneously, in order to avoid that the speckle pattern will change from one to the next, and they are radially scaled to a common wavelength, then the difference of images should cause the fixed-pattern speckles to drop out. If an exoplanet is in the field, it will show up as a radially shifting positive and negative feature. An additional leverage factor arises if the exoplanet has a strong absorption feature in its spectrum, different from its star. For example, planetary spectrum has deep absorption features like the methane band at 1.7 μm. The fact that the planet is relatively faint in this band gives it an extra possibility for detecting it (Racine et al. 1999; Marois et al. 2005; Biller et al. 2006a). By taking simultaneous images in and out of the methane absorption band, and scaling and subtracting the two images, one effectively removes the residual starlight. The starlight is removed whether in the form of rapidly time-varying speckles, slowly varying super-speckles, or any other form of scattered light within the telescope and instrument, while leaving the planetary light intact or almost intact. This technique exploits a beam splitter positioned in the pupil of the instrument and two narrow-band filters (see Fig. 4.19).

Biller et al. (2006b) successfully applied SDI in finding cool companions with the NACO instrument on the ESO-VLT. The SDI has been also applied in other high-contrast instruments, for example, HiCIAO at Subaru telescope with which (Janson et al. 2013) detected the deep methane absorption of the planetary companion they discovered orbiting the star GJ504 (see Fig. 4.20).

The method produces a gain in dynamic range of about one or two magnitudes. SDI is, however, limited in application as it relies on an intrinsic feature of the companion spectrum; the CH_4 absorption feature is only found in objects with $T_{\rm eff} < 1200$ K. In addition, there is a need for follow-up spectroscopy of the candidate object detected—both to confirm its nature (often via common proper motion with the parent) and to characterize it in detail. Increasing of the spectral resolution, initially suggested by Sparks and Ford (2002), can vastly increase the power of this technique, and it becomes a somewhat different speckle suppression method called spectral deconvolution (SD).

4.7.2.2 Spectral Deconvolution

Several authors (e.g., Sparks and Ford 2002; Fusco et al. 2005; Berton et al. 2006) have suggested that the wealth of spectral information available in an integral field spectrograph (IFS) data cube (see Sect. 4.8) can be utilized to remove scattered starlight and identify the presence of a close-in companion, and extract its spectrum (typically at a spectral resolution of 30–80) with enhanced SNR, thus maximizing the contrast. The use of an IFS produces as output a data cube of images that has two of the three dimensions that are spatial ones, while the third is along the spectral dimension. Each image is a monochromatic representation of the scene that the

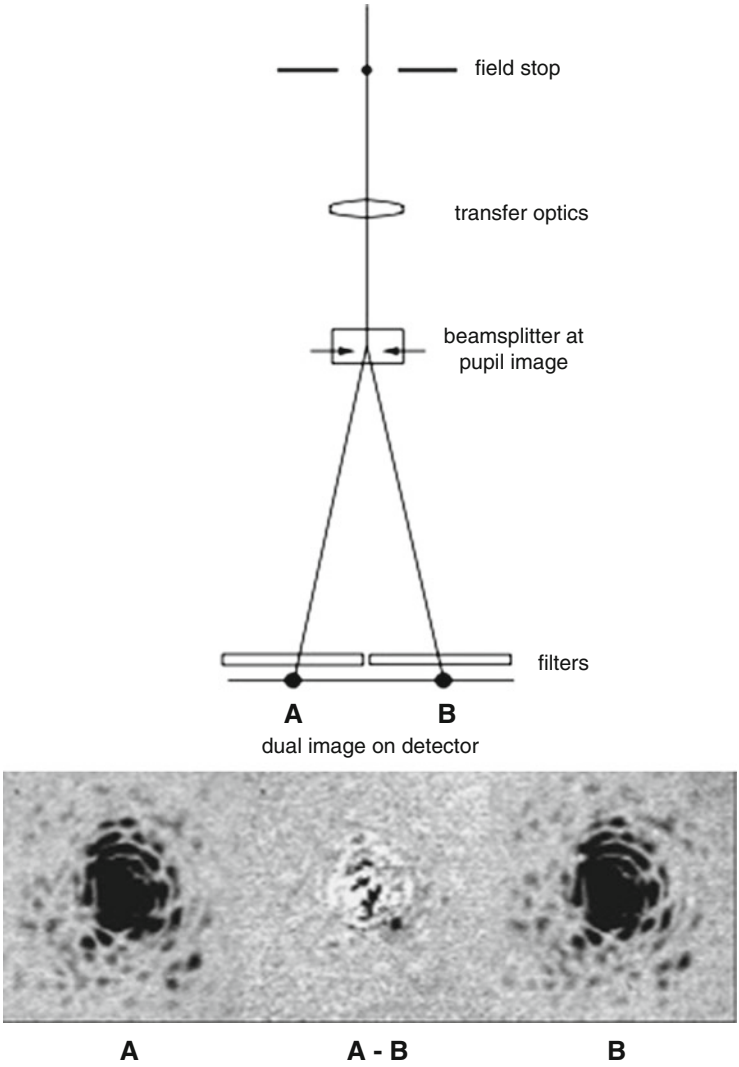

Fig. 4.19 Scheme of the set up of a dual-imaging instrument together with the resulting images and their difference. The figure is taken from Racine et al. (1999) and reprinted with permission from the authors

instrument is observing. Each image of the data cube is affected by speckles and PSF artifacts. The speckle noise pattern is chromatic. Because the typical dimension of a speckle is about λ/D they will move radially in an image as a function of wavelength. In this type of data, the speckles follow diagonal paths through a data cube, while any genuine astrophysical structures have fixed positions with wavelength.

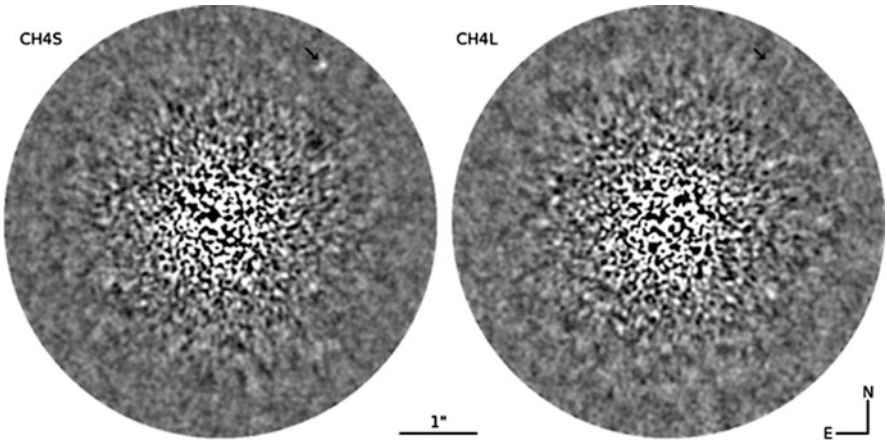

Fig. 4.20 Two simultaneous images of the star GJ504 taken in the CH_4S (*left*, from 1.468 μm to 1.628 μm, outside the CH_4 absorption band) and CH_4L (*right*, from 1.643 μm to 1.788 μm, inside the CH_4 absorption band) intermediate band filters of HiCIAO. In the *left panel* the planet (*right-top* of the image) is bright but disappears in the *right panel* showing the methane absorption of GJ504b. The figure is taken from Janson et al. (2013) and reprinted with permission from the authors

Spectral deconvolution (SD), this is the name of the technique, takes advantage of this wavelength dependence by subtracting all wavelength-dependent artifacts in the stellar PSF, thus unmasking the presence of real physical objects. For a data cube obtained using an IFS, spatially rescaling of each individual slice, proportionally to its wavelength, aligns the speckles but makes the planet move inward with increasing wavelength. Speckles are now well fitted by a smooth (e.g., a low-order polynomial) function to each pixel along the cube's dispersion axis while the planet produces a narrow bump while traveling through the pixel at a certain wavelength range. Because this bump is badly fitted by the smooth function, the subtraction of the fit removes most of the speckles and leaves the planet. Finally, each frame is rescaled back to its original dimension so that the companion is always in the same position. This latter method has been used to reach 9 mag contrast at 0.2 arcsec separation without a coronagraph (Thatte et al. 2007). At separation less than the bifurcation radius (Thatte et al. 2007) the spectrum is completely covered by an eventual companion that, for this reason, would be completely canceled so that the method is not effective at these small separations. To derive this limit (the bifurcation radius) consider a companion object located at radial distance r (expressed in angular units) from the primary object, imaged with a telescope of diameter D. The distance from the first null of the Airy pattern is given by the usual formulation

$$\Theta_0 = 1.22\frac{\lambda_0}{D}$$

where λ_0 is the shortest wavelength in the IFS data cube. Defining the object extent as equal to $2\Theta_0$, and the wavelength range of the IFS as extending from λ_0 (shortest) to λ_1 (longest), we obtain an expression for the movement of the companion's center in the scaled data cube as

$$\Delta r = r - r\frac{\lambda_0}{\lambda_1} = r\frac{\Delta\lambda}{\lambda_1}.$$

Noting that the extent of the object stays constant in the scaled data cube, we can then express the bifurcation point as

$$\Delta r = r\frac{\Delta\lambda}{\lambda_1} = 2\epsilon \times 1.22\frac{\lambda_0}{D},$$

where ϵ is a factor slightly greater than 1. It is obvious from the relation that extended wavelength coverage by the IFS is crucial for removing the flux from a close-in companion in the scaled data cube. For example, the bifurcation radius in the case of an 8 m telescope and typical values of ϵ (1.1–1.2) are equal to 516 mas for H band and 246 mas for $H + K$ bands (Thatte et al. 2007).

The SD technique holds great promise for direct imaging of exoplanets, as it simultaneously detects and spectrally characterizes any faint companion, thus removing the need for expensive and time-consuming follow-up observations, either to detect common proper motion or to obtain a spectrum of the faint companion. In addition, the technique does not require any assumptions about the companion's spectral characteristics (e.g., presence of a CH_4 feature), and can therefore be applied to any high-contrast application.

4.7.3 Polarization

Perhaps the most successful of all speckle suppression techniques to date, dual-mode polarimetric imaging exploits the fact that in general starlight is very weakly polarized. If one is attempting to image an object or material around a star that exhibits large fractional polarization, the starlight and speckles can be removed with almost arbitrary precision. Images are formed using a Wollaston prism, which sends light with perpendicular polarization vectors in slightly different directions. Two images can form and sense simultaneously in this manner. When they are subtracted, only light, which is actually polarized, remains in the image. If the starlight is not polarized, it will be completely removed. Speckles are formed from unpolarized starlight. This technique has been used very successfully to image disks of dust that polarize light through the scattering process (Kuhn et al. 2001; Perrin et al. 2004; Oppenheimer et al. 2008). Another motivation that makes polarimetry a very attractive method for investigation of the reflected light from extrasolar planets is because the expected scattering polarization of an extra solar planet is typically high,

>10 % for large apparent separations when the phase angle is in the range 60°–120° (e.g., Seager et al. 2000; Stam et al. 2004). The level of scattering polarization can be substantially higher.

4.7.3.1 Other Methods

Up to now, other new methods, different from those we previously described, became popular for people working in high-contrast imaging. Most of them are evolutions of SDI and SD. Others can be used in conjunction with chromatic methods such as, for example, locally optimized combination of images (LOCI, Lafrenière et al. 2007b) or the principal component analysis (PCA, see, e.g., Amara and Quanz 2012; Soummer et al. 2012; Oppenheimer et al. 2013). Discussions of the pros and contras of the different methods are voluminous. We simply renunciate them and give references for whoever will be interested.

4.8 Instrumentation

There are a lot of currently operating and proposed projects around the world that have exoplanet imaging and spectroscopy as a major part of their scientific justification. Here we briefly describe some of these experiments. Of course, future projects will likely be different in some respects. From a very general perspective the observational field of comparative exoplanetary science via direct detection is in a nascent stage and with the realization of new generation high-contrast imagers, we are just now on the verge of routinely observing such objects with both photometry and spectroscopy. A lot of these systems are listed in Table 4.1 at the very beginning of this chapter. Among them, there are some instruments that discovered most of the known extrasolar planets via direct imaging. So, it is worth mentioning the past generation of high-contrast imagers.

First of all the HST. The space telescope has several coronagraphic capabilities, thanks to three instruments: ACS, NICMOS, and STIS. The Advanced Camera for Surveys, High Resolution Camera (ACS-HRC, from 2002 to 2007), was repaired after an electronics malfunction in 2007. The ACS is a third generation HST instrument and includes three channels: a wide field channel (WFC), with a FOV of 202 × 202 square arcsec covering the range from 350 to 1100 nm and a plate-scale of 0.05 arcsec/pixel; a high resolution channel (HRC), with a FOV of 29 × 26 square arcsec covering the range from 170 to 1100 nm and a plate-scale of 0.027 arcsec/pixel; a solar blind channel (SBC), with a FOV of 34.6 × 30.5 arcsec FOV, spanning the range from 115 to 170 nm and a plate-scale of 0.032 arcsec/pixel. NICMOS: Near Infrared Camera and Multi-Object Spectrometer (from 1997 to 1999, and 2002 to 2008, with a possible restart in the future) provides imaging capabilities in broad-, medium-, and narrow-band filters, broad-band imaging polarimetry, coronographic imaging, and slitless grism spectroscopy,

in the wavelength range 0.8–2.5 μm. NICMOS has three adjacent but not contiguous cameras, designed to operate independently, each with a dedicated array at a different magnification scale. STIS: Space Telescope Imaging Spectrograph (from 1997 to 2004, and 2009 to present) was installed onboard HST during Servicing Mission (SM) 2 in 1997 and operated until an electronic failure in 2004. It resumed operations with all ultraviolet and optical channels in 2009 after it was successfully repaired during SM4. STIS is a versatile imaging spectrograph, providing spatially resolved spectroscopy in the UV and optical, high spatial resolution echelle spectroscopy in the UV, solar-blind imaging in the UV, and direct and coronagraphic imaging in the optical.

Among the three, the most significant for purposes of exoplanet detection is the coronagraph in the NICMOS instrument. This produces a moderate level of coronagraphic suppression of scattered starlight, but optical errors still scatter light into a residual halo. However, this halo is very stable. This allows successive images taken of different stars or the same star at different orientations to be subtracted, enhancing contrast by a factor of 5–10. The final contrast is then limited by the time evolution of optical errors as the temperature of the telescope changes. At visible wavelengths this had little sensitivity to warm planets but excellent sensitivity to scattered light from circumstellar dust, particularly using PSF subtraction. Both major HST coronagraphs have relatively large IWAs (400–800 marcsec).

Concerning ground-based instruments, NACO (short for NAOS–CONICA) is an AO system consisting of the infrared camera CONICA (Lenzen et al. 2003) assisted by the AO facilities NAOS (Rousset et al. 2003) placed at UT4 from 2001 up to 2013 and now it is placed at UT1 of VLT. NACO provides adaptive optics assisted imaging, imaging polarimetry, coronagraphy, sparse aperture masking, and spectroscopy. NAOS, the adaptive optics (AO) front end, has been designed to work with NGSs and moderately extended objects ($<4''$) and is equipped with one infrared (0.8–2.5 μm) and one visual WFS (0.45–1.0 μm). For a point-like reference source with a visual brightness of $V = 12$ mag, NAOS can provide Strehl ratios as high as 50 % in the K band. The magnitude limit for correction depends on the spectral type of the NGS; for reference partial correction can be obtained for targets as faint as $V = 16.7$ mag. The AO reference star can be either the science object itself or a close-by star (within $55''$). In 2004 it was equipped with an SDI instrument (Lenzen et al. 2004), which is now decommissioned. A number of direct imaging surveys devoted to planets and BD were undertaken by other investigators using the VLT telescopes. In these surveys a classical Lyot coronagraph is fitted behind an AO system. Typically these projects have much smaller fields of view than those described in the previous subsection, but they have more effective starlight suppression due to the presence of a coronagraph. The NACO instrument (Kasper et al. 2009), which operates at 4 μm with a coronagraph and ADI starlight suppression, has been particularly effective in imaging objects that may be planets or brown dwarfs, as it is not easy to distinguish them when they are very young (Chauvin et al. 2005a; Neuhäuser et al. 2005; Lagrange et al. 2009; Thalmann et al. 2009).

A new generation of high-contrast imaging instruments specifically designed for direct imaging of extrasolar planets is now operative, such as the Project 1640 at the 5 m Palomar Telescope (Crepp et al. 2011) which provides important science results (see, e.g., Oppenheimer et al. 2013), or the Gemini planet finder (GPI, Macintosh et al. 2014) at the Gemini South Telescope, that just concluded its commissioning phase and it is now providing scientific results (Galicher et al. 2014). A fourth instrument, the coronagraphic high angular resolution imaging spectrograph (CHARIS, Peters-Limbach et al. 2013), is expected to be operative at the Subaru Telescope at the end of 2015. For the time being the Subaru telescope is outfitted with a new AO coronagraphic imaging device called HiCIAO (Tamura et al. 2006), which includes speckle suppression modes using dual-mode polarimetry and SDI. In Europe, the Spectro-Polarimetric High-contrast Exo-planet REsearch instrument (SPHERE) just had its first light at VLT (Beuzit et al. 2008).

In the following paragraph Project 1640 and GPI will be briefly described while SPHERE will be described in more details. For the first two imagers, references are given in order to dig out details of these instruments.

Project 1640, the successor to the Lyot project at Palomar's 5-m Hale Telescope, combines AO correction with an apodized-pupil and hard-edged mask coronagraph as well as an integral field hyperspectral imaging device, simultaneously obtaining coronagraphic images at 30 different wavelengths over the 1.0–1.8 μm range (Hinkley et al. 2011). This system has begun surveying nearby stars and has found faint stellar companions already. It employs chromatic speckle suppression as well as ADI and SDI, being the first instrument to be able to attempt all three such speckle removal techniques on the same set of data. Project 1640 is due to be upgraded to have a full 3217 actuator AO system for far superior wavefront control, and a second-stage WFS (called a "science-arm WFS"), which obtains the wavefront distortion due to the optical system on a timescale of roughly 1 s. This science-arm WFS is designed to sense and control, through periodic feedback to the AO system, the long-lived speckles that are removed so efficiently by the polarimetric technique, allowing the chromatic speckle suppression to act on much fainter speckles at the 10^{-7} level. The system employs an apodized-pupil Lyot coronagraph (APLC) and the same hyperspectral imaging device as in the current Phase I. The IFS is a TIGER-type (Bacon et al. 1995) microlenses based one with an FOV of 4″ that will allow to obtain low-resolution spectra ($R \sim 33$–58) at all 4×10^4 image samplings. This may allow objects as faint as 10^{-8} at $6\lambda/D$ to be detected and spectra to be extracted, although the FOV is only 4 arcsec wide. This begins to open the exoplanet characterization phase space to the study of many Jupiter-mass exoplanets.

The Gemini planet imager (GPI, Macintosh et al. 2008, 2014), with a goal of 10^{-7} contrast at $6\lambda/D$ in raw contrast, is a large instrument consisting of an AO system with 1500 actuators using a MEMS deformable mirror, an adopted-pupil Lyot chronograph, a high-accuracy IR interferometer calibration system, and a near-infrared IFS to allow detection and characterization of self-luminous extrasolar planets. GPI is designed for near-infrared imaging and spectroscopy of young (10–1000 Myr), massive (1–10 M_J) self-luminous planets. GPI's direct coronagraphic images will be sensitive to planets in wide orbits, inaccessible to Doppler tech-

niques. It will be capable of obtaining moderate resolution ($R \sim 10$–100) of detectable planets, allowing their temperature and gravity to be measured.

GPI consists of six major subsystems (Macintosh et al. 2008). The adaptive optics (AO) system controls the wavefront errors induced by the atmosphere, telescope, and GPIs own optics. The coronagraph masks control the diffraction of coherent starlight. The calibration interferometer (CAL) is one of the most unique features—a post-coronagraphic WFS—it exploits the optical properties of the coronagraphic process to sense the wavefront at the focal plane occultor, the key location. The cryogenic TIGER-Type IFS is the only science instrument. Finally, an optomechanical superstructure (OMSS) supports all the subsystems and components. Top-level software coordinates the operation of the instrument and interfaces to the observatory. GPI had its first light at the end of 2013 observing β Pictoris in the H band (1.65 μm) obtaining more than 30 individual 60 s images in coronagraphic mode. The planet β Pic b (see Fig. 4.21) was immediately visible in a single raw 60 s exposure (Macintosh et al. 2014).

The SPHERE (see Fig. 4.22) planet-finder instrument installed at the VLT (Beuzit et al. 2008) is a highly specialized instrument dedicated to high-contrast imaging, built by a wide consortium of European laboratories. The instrument is made of four subsystems: the Common Path Optics and three science channels, a differential imaging camera (IRDIS), an IFS, and a visible imaging polarimeter (ZIMPOL). The common path includes pupil stabilizing for optics (tip–tilt and derotator), the SAXO ExAO system with a visible WFS, and NIR coronagraphic devices in order to feed IRDIS and IFS with highly stable coronagraphic images. The principal goal of the instrument is to find and characterize giant, gaseous,

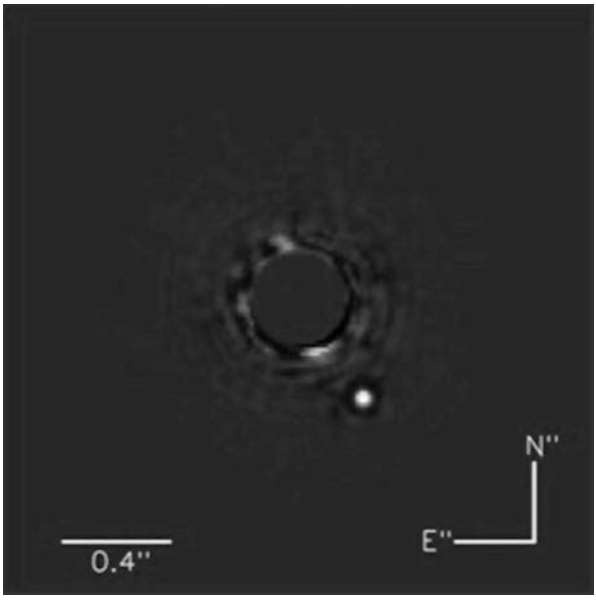

Fig. 4.21 Combined 30 min image of β Pictoris. The image has been obtained after it was reduced with ADI and SDI techniques. The figure is taken from Macintosh et al. (2014) and reprinted with permission from the authors

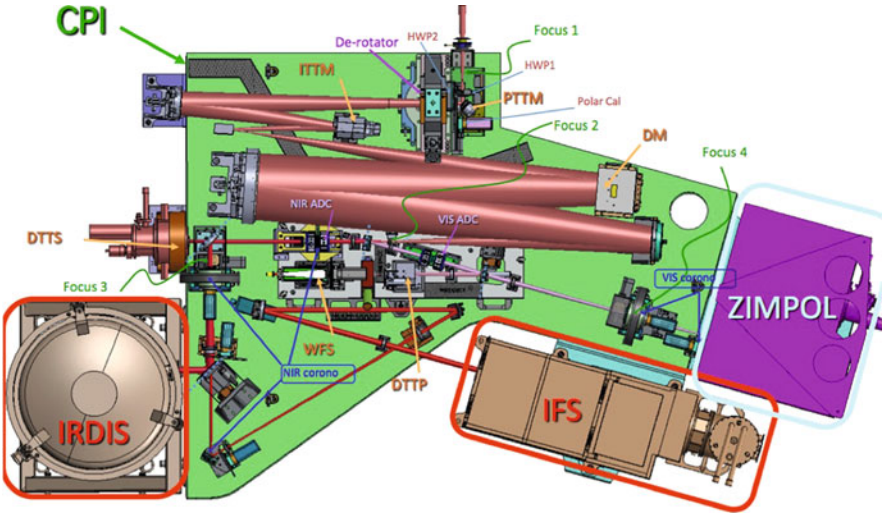

Fig. 4.22 Mechanical and optical scheme of Spectro-Polarimetric High-contrast Exo-planet REsearch (SPHERE) instrument. The three scientific instruments are fully visible

long-period planets within the solar neighborhood. As formation models predict, young planets are hot and self-luminous, making their direct detection possible by masking the primary star with a coronagraph. Evolved systems can also be detected through their reflected, polarized, light. From the ground, ExAO systems are required to correct for the atmospheric turbulence at very high frequency. Classical differential imagers and IFS are then clearly complementary in their properties, and an instrument where both these science modules are available may be extremely powerful for planet search. IFS explores the stellar neighborhood in order to find planetary spectral features. This quest is conducted searching for strong CH_4 absorption bands in both the stellar light reflected by gaseous Jupiter-like planets and in thermal emission from young-warm planets. Moreover it will be possible to have a first order characterization of the low mass companion itself. Additional science topics addressed by SPHERE include the study of protoplanetary disks, brown dwarfs, evolved massive stars, Solar system, and extragalactic science.

The AO module on board on SPHERE, SAXO (Fusco et al. 2006; Petit et al. 2014), includes a 41×41 actuator high-order deformable mirror from CILAS with a maximum stroke $> \pm 3.5\,\mu m$, and a 40×40 lenslet visible Shack–Hartmann WFS, based on the dedicated 240×240 pixel electron multiplying CCD220 from EEV achieving a temporal sampling frequency of 1.2 kHz with a read-out noise <1 electron and a 1.4 excess photon noise factor. The WFS is equipped with a focal plane spatial filter for aliasing control. At the heart of the AO system is the ESO standard real-time computer platform called SPARTA providing a global AO loop delay <1 ms. SPARTA allows control of the system loops while also providing turbulent parameters and system performance estimation as well all the relevant data

for an optimized PSF reconstruction and a clever signal extraction from scientific data. After the XAO module the light path goes towards the scientific instruments. After the splitting due to a NIR–visible beam splitter, several coronagraphic devices could be inserted in the NIR part of the beam. The baseline coronagraph suite will include several coronagraphs: three apodized Lyot coronagraphs (Carbillet et al. 2011) of different dimensions (145, 185, and 240 mas of diameter), two 4-quadrants phase-mask coronagraphs (see Boccaletti et al. 2008), as well as two classical Lyot coronagraphs.

The common path and infrastructure (CPI) brings the telescope light to the three scientific modules (IFS, IRDIS, and ZIMPOL). The CPI contains the deformable mirror (DM), relay optics such as toric mirrors (Hugot et al. 2012), derotator, atmospheric dispersion compensators, and coronagraphs (Dohlen et al. 2011). The NIR beam is then subdivided in two branches, one feeds IRDIS and the second part IFS.

The IRDIS science module (Dohlen et al. 2008) covers a spectral range from 0.95–2.32 μm with an image scale of 12.25 mas consistent with Nyquist sampling at 950 nm. The FOV is $11 \times 12.5\,\mathrm{arcsec}^2$, both for direct and dual imaging. Dual band imaging is the main mode of IRDIS, providing images in two neighboring spectral channels with <10 nm RMS differential aberrations. Two parallel images are projected onto the same 2k × 2k detector with 18 μm square pixels, of which they occupy about half the available area. There is a selection of 12 filters available for imaging, in broad-, medium-, or narrow-band, and five different filter pairs are dedicated to the DBI mode (Vigan et al. 2010). The classical imaging mode allows high-resolution coronagraphic imaging of the circumstellar environment through broad-, medium-, and narrow-band filters throughout the NIR bands including K_s. In addition to these modes, long-slit spectroscopy at resolving powers of 50 and 500 is provided coupled to simple Lyot coronagraphy for the characterization of detected companions (Vigan et al. 2008), as well as a dual polarimetric imaging mode. A pupil-imaging mode for system diagnosis is also implemented.

IFS is a very versatile instruments, well adapted for spectroscopic differential imaging as needed for detection of planets around nearby stars (Claudi et al. 2008, 2014). On the contrary of the other high-contrast imagers described, the integral field unit (IFU) of the SPHERE IFS is a BIGRE lenslet array (Antichi et al. 2009).The main advantage of IFS is that differential aberrations can be kept at a very low level; this is true in particular for lenslet-based systems, where the optical paths of light rays at different wavelengths within the IFS itself can be extremely close to each other. Additionally, IFS provides wide flexibility in the selection of the wavelength channels for differential imaging, and the possibility to perform spectral subtraction, which in principle allows recovering full information on the planet spectra, and not simply the residual of channel subtraction, as in classical differential imagers. The main drawback of IFS is that they require a large number of detector pixels, resulting in a limitation in the FOV, which is more severe for lenslet-based systems. It has two observing modes: YJ-mode (0.95–1.35 μm) with a two-pixels resolving power of 50 and YJH-mode (0.95–1.65 μm) with a two-pixels resolving power of 30.

IRDIS and IFS can operate in parallel on infrared light in the range 0.95–2.32 μm. To allow for a parallel operation of the two NIR instruments, the light entering the telescope is split in two beams downstream of the coronagraphic mask, each instrument having its own set of Lyot stops. Two dichroic plates are available to allow for two different observing modes: IRDIFS mode, where IRDIS performs DBI observations in H band, while IFS works in YJ-mode; and IRDIFS_EXT mode, where IRDIS performs DBI in K_s band, and IFS observes in YJH-mode (Beuzit et al. 2006).

Fed by the visible part of the optical beam, ZIMPOL is located behind the SPHERE visible coronagraph. Among its main specifications are a bandwidth of 600–900 nm and an instantaneous FOV of 3×3 arcsec^2 with access to a total FOV of 8 arcsec in diameter by an internal field selector (Thalmann et al. 2008). The ZIMPOL optical train contains a common optical path that is split with the aid of a polarizing beam splitter in two optical arms, each with its own detector. The common path contains common components for both arms like calibration components, filters, a rotatable half wave plate, and a ferroelectric liquid crystal polarization modulator. The two arms have the ability to measure simultaneously the two complementary polarization states in the same or in distinct filters. The images on both ZIMPOL detectors are Nyquist sampled at 600 nm. The basic ZIMPOL principle for high precision polarization measurements includes a fast polarization modulator with a modulation frequency in the kHz range, combined with an imaging photometer that demodulates the intensity signal in synchronism with the polarization modulation. The polarization modulator and the associated polarizer convert the degree-of-polarization signal into a fractional modulation of the intensity signal, which is measured in a demodulating detector system by a differential intensity measurement between the two modulator states. Each active pixel measures both the high and low states of the intensity modulation and dividing the differential signal by the average signal eliminates essentially all gain changes, notably changes of atmospheric transparency or electronic gain drifts.

The new planet-finder instrument at VLT in Chile, SPHERE, just had its first light during the spring 2014 (see Fig. 4.23).

Several instruments specifically designed to image extrasolar planets are under construction or in design. These instruments will provide a wealth of new data about exoplanets, and ultimately we hope that direct imaging will allow us to answer fundamental questions about planetary systems. On longer timescales, several proposals exist for 20–40 m extremely large telescopes (ELTs), including Extreme AO coronagraphs. Lastly, there are three projects for ELTs that are being actively pursued by the North American (TMT and GMT) and European (E-ELT) astronomical communities. These programs appear feasible in 8–10 years from now. All these projects consider instruments for the direct imaging of exoplanets, either in the near-IR: EPCS at E-ELT (the evolution of the previously named EPICS) (Kasper et al. 2008); PFI at TMT, and HRCAM at GMT or in the mid-IR METIS at E-ELT: Brandl et al. (2008); MIRES at TMT, and MIISE at GMT. E-ELT, with its 42 m diameter, is the most ambitious among these projects. Although even these instruments will be unable to detect Earth-sized planets, it is predicted that they can

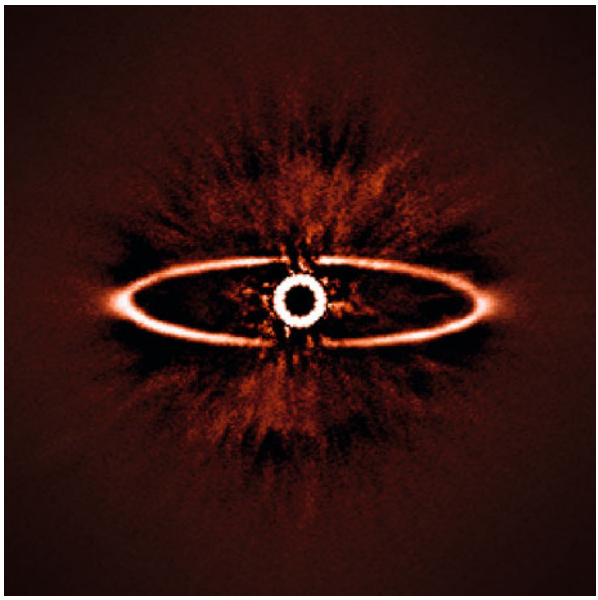

Fig. 4.23 First light of SPHERE. Dust ring around the star HR 4796A (ESO PR1417)

achieve contrast levels in the 10^{-8} to 10^{-9} range and IWAs as small as 0.03 arcsec (Gratton et al. 2010). The combination of small IWA and higher contrast would allow spectral characterization of mature Jovian or ice-giant planets in reflected starlight at ∼1–3 au scales and studies of zodiacal circumstellar dust in inner parts of Solar systems. An IWA of 0.03 arcsec corresponds to ∼4–5 au in nearby star-forming regions, such as the Taurus and Ophiuchus associations, making possible the direct study of the formation of planetary systems in their first few million years. Overall, though, ground-based AO coronagraphs are unlikely ever to image Earth-sized planets; on the other hand, such telescopes will be capable of imaging giant planets in the moderately near future, producing both the first census of giant planets in the 5–20 au range and the first spectra of such planets. Around 2025, ELT systems will push this into the 10^{-20} earth mass range.

Going back to space facilities, the James Webb Space Telescope is another project that is currently funded and expected to see first light. There are two primary instruments being constructed for JWST that address exoplanet imaging directly. The NIRCAM instrument has several coronagraphic modes operating at 1–5 μm and should achieve contrasts of about 10^{-7} to within a few λ/D, which will allow imaging the thermal emission from exoplanets. It employs apodization on the segments of the telescope to reduce speckles pinned to the diffraction pattern as well as several other options for starlight suppression. Grisms will allow us to perform low-resolution spectroscopy as well. The TFI instrument has Fabry–Perot etalons operating from 1.5 to 2.4 μm and 3.1 to 5.0 μm, with several hard-edged circular occulting spots and a non-redundant masking mode. Another instrument of interest in direct imaging of exoplanets with JWST is the Mid-InfraRed Imager

(MIRI), which has several coronagraphic options and images at wavelengths from 9 to 12 μm. This system is predicted to detect planets as cool as 300 K within an arcsecond of a star, using four-quadrant phase masks or a traditional Lyot-style coronagraph. The combination of all these instruments will provide a suite of measurements across a broad range of wavelengths. Due to the smaller telescope size and lesser control of the wavefront errors, NIRCAM will not allow higher contrasts and smaller inner fields of view with respect to SPHERE and GPI; however, it will not be limited to bright targets, and will then have an important niche for nearby extreme M stars. MIRI will open a new window, outperforming by far existing ground-based mid-IR imagers (Gratton et al. 2010).

4.9 Results

Within the coming years, direct imaging will represent the only viable technique distinct from microlensing for probing the existence of extrasolar planets and brown dwarfs companion at large (≤5–10 au) separations from their parent stars. Up to date, about 60 low mass companions (both brown dwarfs and planets) have been discovered by observing nearby stars using AO alone (see Table 4.5). These surveys have proven to be productive, but they generally do not operate in the high-contrast regime (Chauvin et al. 2002; Brandeker et al. 2003; Neuhäuser et al. 2003; Beuzit et al. 2004; Masciadri et al. 2005; Lafrenière et al. 2007a; Nielsen et al. 2008). For example, these surveys have discovered several very young companions of stars or brown dwarfs that may have very low mass, but whose nature lying at the brown dwarf/planet boundary has been extensively debated: 2MASSWJ 1207334-393254 B (Chauvin et al. 2005a), GQ Lup B (Neuhäuser et al. 2005), and AB Pic B (Chauvin et al. 2005b) just to do some examples. For technical motives rather than scientific ones, most surveys have targeted young and nearby stars. This search space is already limited, and imaging surveys have just scratched the surface, strongly limited by contrast and IWA capabilities. The primary goal of high-contrast imaging is to discover planetary systems. This technique, as we have shown, is also unique for the characterization of planetary atmospheres that are not strongly irradiated by the host of the planetary system (Janson et al. 2010; Bowler et al. 2010; Barman et al. 2011a,b; Bonnefoy et al. 2010, 2013; Konopacky et al. 2013). The first planetary mass companions were detected at large distances (≥100 au and/or at a small mass ratio with their primaries, indicating a probable star-like or gravitational disk instability formation mechanism (Chauvin et al. 2005a; Lafrenière et al. 2008). The breakthrough discoveries of closer and/or lighter planetary mass companions like Fomalhaut b (<1$M_{\rm Jup}$ at 177 au; Kalas et al. 2008, 2013), HR 8799 bcde (giant planets with masses, respectively, 10, 10, 10, and 7 $M_{\rm Jup}$ at 14, 24, 38, and 68 au; Marois et al. 2008, 2010), β Pictoris b (8 $M_{\rm Jup}$ at 8 au; Lagrange et al. 2009), or more recently κ And b ($14^{+25}_{-2} M_{\rm Jup}$ at 55 au; Carson et al. 2013; Bonnefoy et al. 2013), HD 95086 b (4–5$M_{\rm Jup}$ at 56 au; Rameau et al. 2013a,b), GJ 504 b ($4^{+4.5}_{-1} M_{\rm Jup}$ at 43.5 au; Kuzuhara et al. 2013), and HD 106906 (Bailey

Table 4.5 Low mass companions discovered by direct imaging up to now

NAME	Mass (M_J)	a (au)	Disc.	D (pc)	Sp	Age (Gyr)	Instr.	Ref.
Kepler-70 c	0.0021	0.0076	2011	1180	sdB	0.0184	Kepler	1
Kepler-70 b	0.014	0.006	2011	1180	sdB	0.0184	Kepler	1
Fomalhaut b	3	115	2008	7.704	A3 V	0.44	HST	2
2M1207 b	4	46	2004	52.4	M8	0.008	VLT/NACO	3
GJ 504 b	4	43.5	2013	17.56	GOV	0.16	Subaru/HiCIAO	4
HD 95086 b	5	61.5	2013	90.4	A8III	0.017	VLT/NACO	5
LKCA 15 b	6	15.7	2011	145	K5V	0.002	KECK II/NIRC2	6
CFBDS 1458 b	6.5	2.6	2011	23.1	T9.5	3	KECK II/NIRC2	7
HR 8799 b	7	68	2008	39.4	A5V	0.06	KECK/GEMINI	8
beta Pic b	7	9.2	2008	19.3	A6V	0.012	VLT/NACO	9
51 Eri b	7	14	2015	29.4	F0IV	0.02	GPI	10
2M 044144 b	7.5	15	2010	140	M8.5	0.001	HST+GEMINI-N/NIRI	11
WD 0806-661B b	8	2500	2011	19.2	DQ D	1.5	CTIO/ISPI	12
HR 8799 e	9	14.5	2010	39.4	A5V	0.06	KECK/GEMINI	13
ROXs 42B b	9	140	2013	135	M0 D	0.0068	KECK II/NIRC2	14
HR 8799 c	10	42.9	2008	39.4	A5V	0.06	KECK/GEMINI	8
HR 8799 d	10	27	2008	39.4	A5V	0.06	KECK/GEMINI	8
FW Tau b	10	330	2013	145	M4	0.0018	KECK II/NIRC2	15
DH Tau b	11	330	2005		M0.5V	0.001	Subaru/CIAO	16
HD 106906 b	11	654	2013	92	F5V	0.013	MAGELLAN/MGAO	17
GU Psc b	11	2000	2014	48	M3	0.1	GEMINI-S/GMOS +	18
							GEMINI-N/GNIRS	
							CFHT/WIRCAM	
VHS 1256-1257 b	11.2	102	2015	12.7	M7.5	0.2	VISTA/VIRCAM	19

(continued)

Table 4.5 (continued)

NAME	Mass (M_J)	a (au)	Disc.	D (pc)	Sp	Age (Gyr)	Instr.	Ref.
Ross 458(AB) c	11.3	1168	2010	11.7	M0.5	0.475	UKIRT	21
CHXR 73 b	12	200	2006		M3.25	0.002	HST	22
SR 12 AB c	13	1083	2011	125	K4-M2.5	0.001	IRFS/SIRIUS	22
							Subaru/CIAO	
WISE 0458+6434 b	13	5	2011	10.5	T 8.5	1	WISE	23
2M 0103(AB) b	13	84	2013	47.2	M	0.03	VLT/NACO	24
AB Pic b	13.5	275	2005	47.3	K2 V	0.03	ESO 3.6/ADONIS	25
							VLT/NACO	
2M 0219-3925 b	13.9	156	2015	39.4	M6		CTIO/SIMON	26
UScoCTIO 108 b	14	670	2007	145	M7	0.011	UKST	27
1RXS 1609 b	14	330	2008	145	K7V	0.011	GEMINI-N/NIRI+ALTAIR	28
kappa And b	14	55	2013	51.6	B9IV	0.03	Subaru/HiCIAO	29
FU Tau b	15	800	2009	140	M7.25	0.001	SPITZER SPACE TEL.	30
HN Peg b	16	795	2007	18.4	G0V	0.2	SPITZER SPACE TEL.	31
ROXs 12 b	16	210	2013	120	M0	0.0076	KECK II/NIRC2	15
CT Cha b	17	440	2008	165	K7	0.002	VLT/NACO	32
GSC 6214-210 b	17	320	2010	145	M1	0.011	PALOMAR 200	32
							KECK II/NIRC2	
WISE 1711+3500 b	18	15	2012	19	T8	3	KECK II/NIRC2	34
2M 2140+16 b	20	3.53	2010	25			KECK/LGS–AO	35
2M 0122-2439 b	20	52	2013	36	M3.5	0.12	SUBARU/IRCS	36
KECK II/NIRC2 HIP 77900 b	20	3200	2013		B6		UKIRT	37
USco1610-1913 b	20	840	2013		K7		UKIRT	38

(continued)

Table 4.5 (continued)

Oph 11 b	21	243	2007	145	M9	0.011	GEMINI-N/NIRI	38
HIP 73990 b	21	20	2015	125	A9V	0.015	KECK II/NIRC2	39
GQ Lup b	21.5	103	2005	140	K7eV	0.001	VLT/NACO	40
WISE 1217+16A b	22	7.6	2012	10	T8.5	6	KECK II/NIRC2	34
HIP 73990 c	22	32	2015	125	A9V	0.015	KECK II/NIRC2	39
HIP 78530 b	23.04	710	2011	156.7	B9V	0.011	GEMINI-N/NIRI	41
TWA 5 A(AB) b	25	86	2009	25	M1.5	0.1	HST/NICMOS	42
UScо1612-1800 b	26	430	2013		M3		UKIRT	37
HIP 74865 b	28	23	2015	115	F3V	0.015	KECK II/NIRC2 +VLT/NACO	39
2M 0746+20 b	30	2.897	2010	12.21			KECK/LGS-AO	35
2M 2206-20 b	30	4.48	2010	26.67			KECK/LGS-AO	35
HD169142 b	30	22.7	2014	145	A7V	12	VLT/NACO	43
CD-35 2722 b	31	67	2011	21.3	M1V	0.1	GEMINI/NICI	44
DE0823-49 b	31.5	0.36	2013	20.69		0.1	VLT/FORS2	45
HD 284149 b	32	400	2014	108.2	F8	0.025	GEMINI-N/NIRI	46
GJ 758 b	35	44.8	2009	15.5	K0V	8.2	Subaru/HiCIAO	47
USco1602-2401 b	47	1000	2013		K4		UKIRT	37

(continued)

Table 4.5 (continued)

NAME	Mass (M_J)	a (au)	Disc.	D (pc)	Sp	Age (Gyr)	Instr.	Ref.
HR 3549 b	47.5	80	2015	92.5	A0V	0.23	VLT/NACO	39
PZ Tel b	62	20	2010	51.49	G6.5	0.023	GEMINI/NICI	48

Reference: **1** Charpinet et al., 2011, Nature, 480, 496; **2** Kalas et al., 2008, Science, 322, 1345; **3** Chauvin et al., 2004, A&A, 425, L29; **4** Janson et al., 2013, ApJL, 778, L4; **5** Rameau et al., 2013, Astroph. J., 779, L26; **6** Kraus et al., 2012 Astroph. J., 745, 5; **7** Liu et al., 2011, Astroph. J., 740, 108; **8** Marois et al. 2008, Science, 322, 1348; **9** Lagrange et al., 2009, Astron. Astroph., 506, 927; **10** Macintosh et al., 2015, Science, In Press; **11** Todorev et al., 2010, Astroph. J., 714, L84; **12** Rodriguez et al., 2011, Astroph. J., 732, L29; **13** Marois et al., 2010, Nature, 468, 1080; **14** Currie et al., 2014, Astroph. J., 787, 104; **15** Kraus et al., 2014, Astroph. J., 781, 20; **16** Itoh et al., 2005, Astroph. J., 620, 984; **17** Bailey et al., 2014, Astroph. J., 780, L4; **18** Naud et al., 2014, Astroph. J., 787, 5; **19** Gauza et al., 2015, Astroph. J., 804, 96; **20** Goldman et al., 2010, Mon. Not. R. Astron. Soc., 404,1140; **21** Luhman et al., 2006, Astroph. J., 649, 894; **22** Kuzuhara et al., 2011, Astron. J., 141, 119; **23** Mainzer et al., 2011, Astroph. J., 726, 30; **24** Delorme et al., 2013, Astron. Astroph., 553, L5; **25** Chauvin et al., 2005, Astron. Astroph., 438, L29; **26** Artigau et al., 2015 in Pubbl.; **27** Bejar et al., 2008, Astroph. J., 673, L185; **28** Lafreniére et al., 2008, Astroph. J., 689, L153; **29** Carson et al., 2013, Astroph. J., 763, L32; **30** Luhman et al., 2009, Astroph. J., 691, 1265; **31** Luhman et al., 2007, Astroph. J., 654, 570 **32** Schmidt et al., 2008, Astron. Astroph., 491, 311; **33** Ireland et al., 2011, Astroph. J., 726, 113; **34** Liu et al., 2012, Astroph. J., 758, 57; **35** Konopacki et al., 2011, ApJ, 711, 1087; **36** Bowler et al., 2013, Astroph. J., 774, 55; **37** Aller et al., 2013, Astroph. J., 773, 63; **38** Close et al., 2007, Astroph. J., 660, 1492; **39** Hinkley et al., 2015, Astroph. J. in press; **40** Neuhäuser et al., 2005, Astron. Astroph., 345, l13; **41** Lafreniére et al., 2011, Astroph. J., 730, 42; **42** Lagrange et al., 1999, Astroph. J., 512, l69 **43** Reggiani et al., 2014, Astroph. J., 792, L23; **44** Wahhaj et al., 2011, Astroph. J., 729, 139; **45** Sahlmann et al., 2013, Astron. Astroph., 556, A133 **46** Bonavita et al., 2014, Astroph. J., 791, 40; **47** Thalman et al., 2009, Astroph. J., 707, L123; **48** Biller et al., 2010, Astroph. J., 720, L82

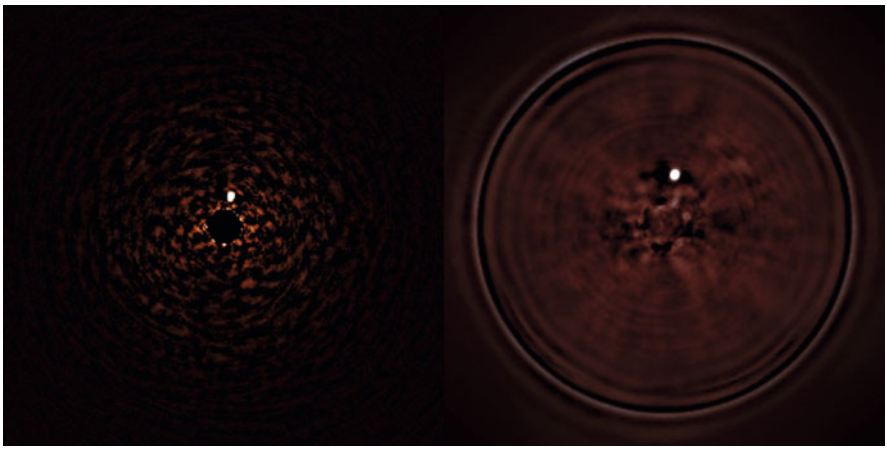

Fig. 4.24 The companion of ι Sgr is clearly visible in the images taken with both IFS and IRDIS. The central star is 4×10^3 times brighter than the M dwarf, but its image has been well canceled by the subtraction procedures

et al. 2014), indicate that we are just initiating the characterization of the outer part of planetary systems. The new generation instrumentation SPHERE and GPI represents a unique opportunity for a major breakthrough in this field, thanks to the small IWA due to new kind of coronagraphs and the extreme contrast for detection of faint companions, which has already been demonstrated by the first scientific observations of these instruments.

As an example of the potentiality of these instruments, SPHERE observed the star ιSgr (HR7581, $J = 2.29$, K giant, $D = 55.7 \pm 0.6$ pc) that was previously known as an astrometric binary (Makarov and Kaplan 2005) but the companion was never detected before. The IFS observing mode was Y–J. Once the IFS data cube was reduced with Spectra Deconvolution and ADI, the faint companion was clearly detected (SNR = 50) at an angular separation of $\theta = 0.249$ arcsec (see Fig. 4.24) and a contrast of 9 mag. The object was simultaneously seen by IRDIS.

Using $\Delta m = 9$ mag the calibration by Delfosse et al. (2000) gives a corresponding main-sequence mass is $0.36\,M_{\odot}$. From the IFS and IRDIS data it was also possible to get the spectrum of the faint companion (see Fig. 4.25), which results essentially without features, as expected for an early M-type star on IFS spectra.

So, it seems that we are entering in a new era, in which we can discover many new objects and investigate the parameter space of faint/small planets close to their parent stars. The new sample is critical to exoplanet science because it will shed some light on planet formation mechanisms close to the snow line, and help to build a bridge between the populations of close planets, discovered by radial velocity or transit techniques, and the free-floating planets, discovered by microlensing observations (Quanz et al. 2012).

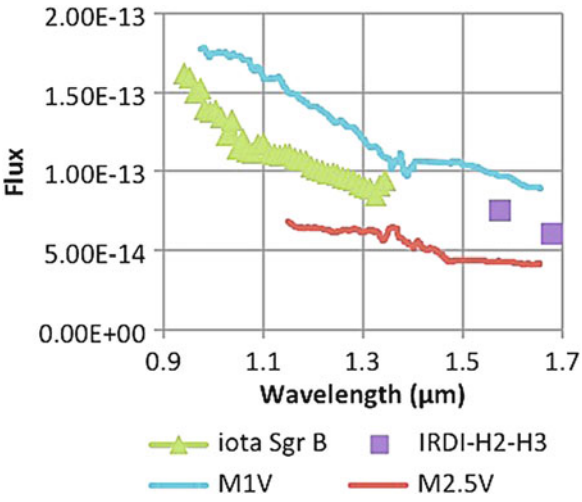

Fig. 4.25 The spectrum (in $W/m^2/\mu m$) of ι Sgr B as obtained by the IFS data (*green triangles*) and IRDIS (*violet squares*) compared with those of two M stars of different spectral types, scaled at the distance of this system

Acknowledgements The author acknowledges support from the "Progetti Premiali" WOW funding scheme of the Italian Ministry of Education, University, and Research. Furthermore I would like to thank Anthony Boccaletti and Raffaele Gratton for the useful discussion and material for the front lectures. Besides I would like to thank also all those unaware colleagues from who I took some help for preparing the lectures and these lecture notes. A particular thanks goes to the organizers of the school (V. Bozza, L. Mancini, and A. Sozzetti) and also to the patient young colleagues and students without who these lecture notes could not exist.

References

Absil, O., Mawet, D.: Formation and evolution of planetary systems: the impact of high-angular resolution optical techniques. Astron. Astrophys. Rev. **18**, 317–328 (2010)

Alibert, Y., Mordasini, C., Benz, W., Winisdoerffer, C.: Models of giant planet formation with migration and disc evolution. Astron. Astrophys. **434**, 343–353 (2005)

Amara, A., Quanz, S.P.: PYNPOINT: an image processing package for finding exoplanets. Mon. Not. R. Astron. Soc. **427**, 948–955 (2012)

Antichi, J., Dohlen, K., Gratton, R.G., Mesa, D., Claudi, R.U., Giro, E., Boccaletti, A., Mouillet, D., Puget, P., Beuzit, J.-L.: BIGRE: a Low cross-talk integral field unit tailored for extrasolar planets imaging spectroscopy. Astroph. J. **695**, 1042–1057 (2009)

Artigau, E., Biller, B.A., Wahhaj, Z., Hartung, M., Hayward, T.L., Close, L.M., Chun, M.R., Liu, M.C., Trancho, G., Rigaut, F., Toomey, D.W., Ftaclas, C.: NICI: combining coronagraphy, ADI and SDI. In: Proceedings of SPIE, vol. 7014, p. 70141Z (2008)

Babcock, H.W.: The possibility of compensating astronomical seeing. Publ. Astron. Soc. Pac. **65**, 229–236 (1953)

Bacon, R., Adam, G., Baranne, A., Courtes, G., Dubet, D., Dubois, J.P., Emsellem, E., Ferruit, P., Georgelin, Y., Monnet, G., Pécontal, E., Rousset, A., Sayéde, F.: 3D spectrography at high spatial resolution. I. Concept and realisation of the integral field spectrograph TIGER. Astron. Astrophys. Suppl. Ser. **113**, 347–357 (1995)

Bailey, V., Meshkat, T., Reiter, M., Morzinski, K., Males, J., Su, K.Y.L., Hinz, P.M., Kenworthy, M., Stark, D., Mamajek, E., Briguglio, R., Close, L.M., Follette, K.B., Puglisi, A., Rodigas, T.,

Weinberger, A.J., Xompero, M.: HD 106906 b: a planetary-mass companion outside a massive debris disk. Astrophys. J. Lett. **780**, L4–L9 (2014)

Baraffe, I., Chabrier, G., Allard, F., Hauschildt, P.H.: Evolutionary models for solar metallicity low-mass stars: mass-magnitude relationships and color-magnitude diagrams. Astron. Astrophys. **337**, 403–412 (1998)

Baraffe, I., Chabrier, G., Allard, F., Hauschildt, P.H.: Evolutionary models for low-mass stars and brown dwarfs: uncertainties and limits at very young ages. Astron. Astrophys. **382**, 563–572 (2002)

Baraffe, I., Chabrier, G., Barman, T.S., Allard, F., Hauschildt, P.H.: Evolutionary models for cool brown dwarfs and extrasolar giant planets. The case of HD 209458. Astron. Astrophys. **402**, 701–712 (2003)

Barman, T.S., Macintosh, B., Konopacky, Q.M., Marois, C.: Clouds and chemistry in the atmosphere of extrasolar planet HR8799b. Astrophys. J. **733**, 65–83 (2011a)

Barman, T.S., Macintosh, B., Konopacky, Q.M., Marois, C.: The young planet-mass object 2M1207b: a cool, cloudy, and methane-poor atmosphere. Astrophys. J. Lett. **735**, L39–L43 (2011b)

Beckers, J.M.: Adaptive optics for astronomy: principles, performance, and applications. Ann. Rev. Astron. Astrophys. **31**, 13–62 (1993)

Beckers, J.M.: Increasing the size of the isoplanatic patch with multiconjugate adaptive optics. In: Ulrich, M.-H. (ed.) Very Large Telescopes and Their Instrumentation, p. 693. ESO, Garching (1988)

Berton, A., Gratton, R.G., Feldt, M., Henning, T., Desidera, S., Turatto, M., Schmid, H.M., Waters, R.: Detecting extrasolar planets with integral field spectroscopy. Publ. Astron. Soc. Pac. **118**, 1144–1164 (2006)

Beuzit, J.-L., Ségransan, D., Forveille, T., Udry, S., Delfosse, X., Mayor, M., Perrier, C., Hainaut, M.-C., Roddier, C., Roddier, F., Mart'n, E.L.: New neighbours. III. 21 new companions to nearby dwarfs, discovered with adaptive optics. Astron. Astrophys. **425**, 997–1008 (2004)

Beuzit, J.-L., Feldt, M., Dohlen, K., Mouillet, D., Puget, P., Antichi, J., Baruffolo, A., Baudoz, P., Berton, A., Boccaletti, A., Carbillet, M., Charton, J., Claudi, R., Downing, M., Feautrier, P., Fedrigo, E., Fusco, T., Gratton, R., Hubin, N., Kasper, M., Langlois, M., Moutou, C., Mugnier, L., Pragt, J., Rabou, P., Saisse, M., Schmid, H.M., Stadler, E., Turatto, M., Udry, S., Waters, R., Wildi, F.: SPHERE: A 'Planet Finder' instrument for the VLT. Messenger **125**, 29 (2006)

Beuzit, J.-L., Feldt, M., Dohlen, K., Mouillet, D., Puget, P., Wildi, F., Abe, L., Antichi, J., Baruffolo, A., Baudoz, P., Boccaletti, A., Carbillet, M., Charton, J., Claudi, R., Downing, M., Fabron, C., Feautrier, P., Fedrigo, E., Fusco, T., Gach, J.-L., Gratton, R., Henning, T., Hubin, N., Joos, F., Kasper, M., Langlois, M., Lenzen, R., Moutou, C., Pavlov, A., Petit, C., Pragt, J., Rabou, P., Rigal, F., Roelfsema, R., Rousset, G., Saisse, M., Schmid, H.-M., Stadler, E., Thalmann, C., Turatto, M., Udry, S., Vakili, F., Waters, R.: SPHERE: a 'Planet Finder' instrument for the VLT. In: SPIE Conference Series, vol. 7014, p. 701418 (2008)

Biller, B.A., Close, L.M., Masciadri, E., Lenzen, R., Brandner, W., McCarthy, D., Henning, T., Nielsen, E.L., Hartung, M., Kellner, S., Geissler, K., Kasper, M.: Contrast limits with the simultaneous differential extrasolar planet imager (SDI) at the VLT and MMT. In: SPIE Conference Series, vol. 6272, p. 62722D (2006a)

Biller, B.A., Close, L.M., Lenzen, R., Brandner, W., McCarthy, D., Nielsen, E., Kellner, S., Hartung, M.: Suppressing speckle noise for simultaneous differential extrasolar planet imaging (SDI) at the VLT and MMT, IAU Coll. 200. In: Direct Imaging of Exoplanets: Science and Techniques, p. 571. Cambridge University Press, Cambridge (2006b)

Boccaletti, A., Riaud, P., Baudoz, P., Baudrand, J., Rouan, D., Gratadour, D., Lacombe, F., Lagrange, A.-M.: The four-quadrant phase mask coronagraph. IV. First light at the very large telescope. Publ. Astron. Soc. Pac. **116**, 1061–1071 (2004)

Boccaletti, A., Abe, L., Baudrand, J., Daban, J.-B., Douet, R., Guerri, G., Robbe-Dubois, S., Bendjoya, P., Dohlen, K., Mawet, D.: Prototyping coronagraphs for exoplanet characterization with SPHERE. In: SPIE Conference Series, vol. 7015, p. 70151B, 10 pp. (2008)

Boccas, M., Rigaut, F., Bec, M., Irarrazaval, B., James, E., Ebbers, A., d'Orgeville, C., Grace, K., Arriagada, G., Karewicz, S., Sheehan, M., White, J., Chan, S.: Laser guide star upgrade of Altair at Gemini North. In: Advances in Adaptive Optics II. Proceedings of SPIE Conference Series, vol. 6272, p. 62723L (2006)

Bonaccini Calia, D., Allaert, E., Alvarez, J.L., Araujo Hauck, C., Avila, G., Bendek, E., Buzzoni, B., Comin, M., Cullum, M., Davies, R., Dimmler, M., Guidolin, B., Hackenberg, W., Hippler, S., Kellner, S., van Kesteren, A., Koch, F., Neumann, U., Ott, T., Popovic, D., Pedichini, F., Quattri, M., Quentin, J., Rabien, S., Silber, A., Tapia, M.: First light of the ESO laser guide star facility. In: Advances in Adaptive Optics II. Proceedings of SPIE Conference Series, vol. 6272, p. 627207 (2006)

Bonnefoy, M., Chauvin, G., Rojo, P., Allard, F., Lagrange, A.-M., Homeier, D., Dumas, C., Beuzit, J.-L.: Near-infrared integral-field spectra of the planet/brown dwarf companion AB Pictoris b. Astron. Astrophys. **512**, A52 (2010)

Bonnefoy, M., Boccaletti, A., Lagrange, A.-M., Allard, F., Mordasini, C., Beust, H., Chauvin, G., Girard, J.H.V., Homeier, D., Apai, D., Lacour, S., Rouan, D.: The near-infrared spectral energy distribution of β Pictoris b. Astron. Astrophys. **555**, A107 (2013)

Born, M., Wolf, E.: Principles of Optics. Pergamon Press, London (1999)

Boss, A.P.: Giant planet formation by gravitational instability. Science **276**, 1836–1839 (1997)

Bowler, B.P., Liu, M.C., Dupuy, T.J., Cushing, M.C.: Near-infrared spectroscopy of the extrasolar planet HR8799 b. Astrophys. J. **723**, 850–868 (2010)

Brandeker, A., Jayawardhana, R., Najita, J.: Keck adaptive optics imaging of nearby young stars: detection of close multiple systems. Astron. J. **236**, 2009–2014 (2003)

Brandl, B.R., Lenzen, R., Pantin, E., Glasse, A., Blommaert, J., Venema, L., Molster, F., Siebenmorgen, R., Boehnhardt, H., van Dishoeck, E., van der Werf, P., Henning, T., Brandner, W., Lagage, P.-O., Moore, T.J.T., Baes, M., Waelkens, C., Wright, C., Käufl, H.U., Kendrew, S., Stuik, R., Jolissaint, L.: METIS: the mid-infrared E-ELT imager and spectrograph. In: SPIE Conference Series, vol. 7014, p. 70141N, 15 pp. (2008)

Brown, T.M.: Transmission spectra as diagnostics of extrasolar giant planet atmospheres. Astrophys. J. **553**, 1006–1026 (2001)

Burrows, A., Marley, M., Hubbard, W.B., Lunine, J.I., Guillot, T., Saumen, D., Freedman, R., Sudarski, D., Sharp, C.: A nongray theory of extrasolar giant planets and brown dwarfs. Astrophys. J. **491**, 856–875 (1997)

Burrows, A., Hubbard, W.B., Lunine, J.I., Liebert, J.: The theory of brown dwarfs and extrasolar giant planets. Rev. Mod. Phys. **73**, 719–765 (2001)

Burrows, A., Sudarsky, D., Hubeny, I.: L and T dwarf models and the L to T transition. Astrophys. J. **640**, 1063–1077 (2006)

Carbillet, M., Bendjoya, P., Abe, L., Guerri, G., Boccaletti, A., Daban, J.-B., Dohlen, K., Ferrari, A., Robbe-Dubois, S., Douet, R., Vakili, F.: Apodized Lyot coronagraph for SPHERE/VLT. I. Detailed numerical study. Exp. Astron. **30**, 39–58 (2011)

Carson, J., Thalmann, C., Janson, M., Kozakis, T., Bonnefoy, M., Biller, B., Schlieder, J., Currie, T., McElwain, M., Goto, M., Henning, T., Brandner, W., Feldt, M., Kandori, R., Kuzuhara, M., Stevens, L., Wong, P., Gainey, K., Fukagawa, M., Kuwada, Y., Brandt, T.; Kwon, J., Abe, L., Egner, S., Grady, C., Guyon, O., Hashimoto, J., Hayano, Y., Hayashi, M., Hayashi, S., Hodapp, K., Ishii, M., Iye, M., Knapp, G., Kudo, T., Kusakabe, N., Matsuo, T., Miyama, S., Morino, J., Moro-Martin, A., Nishimura, T., Pyo, T., Serabyn, E., Suto, H., Suzuki, R., Takami, M., Takato, N., Terada, H., Tomono, D., Turner, E., Watanabe, M., Wisniewski, J., Yamada, T., Takami, H., Usuda, T., Tamura, M.: Direct imaging discovery of a "Super-Jupiter" around the late B-type star κ And. Astrophys. J. Lett. **763**, L32–L37 (2013)

Chabrier, G., Baraffe, I., Allard, F., Haudschildt, P.: Evolutionary models for very low-mass stars and brown dwarfs with dusty atmospheres. Astrophys. J. **542**, 464–472 (2000)

Charbonneau, D., et al.: Detection of thermal emission from an extrasolar planet. Astrophys. J. **626**, 523–529 (2005)

Chauvin, G., Ménard, F., Fusco, T., Lagrange, A.-M., Beuzit, J.-L., Mouillet, D., Augereau, J.-C.: Adaptive optics imaging of the MBM 12 association. Seven binaries and an edge-on disk in a quadruple system. Astron. Astrophys. **394**, 949–956 (2002)

Chauvin, G., Lagrange, A.-M., Dumas, C., Zuckerman, B., Mouillet, D., Song, I., Beuzit, J.-L., Lowrance, P.: A giant planet candidate near a young brown dwarf. Direct VLT/NACO observations using IR wavefront sensing. Astron. Astrophys. **425**, L29–L32 (2004)

Chauvin, G., Lagrange, A.-M., Dumas, C., Zuckerman, B., Mouillet, D., Song, I., Beuzit, J.-L., Lowrance, P.: Giant planet companion to 2MASSW J1207334–393254. Astron. Astrophys. **438**, L25–L28 (2005a)

Chauvin, G., Lagrange, A.-M., Zuckerman, B., Dumas, C., Mouillet, D., Song, I., Beuzit, J.-L., Lowrance, P., Bessell, M.S.: A companion to AB Pic at the planet/brown dwarf boundary. Astron. Astrophys. **438**, L29–L32 (2005b)

Chauvin, G., Lagrange, A.-M., Bonavita, M., Zuckerman, B., Dumas, C., Bessell, M.S., Beuzit, J.-L., Bonnefoy, M., Desidera, S., Farihi, J., Lowrance, P., Mouillet, D., Song, I.: Deep imaging survey of young, nearby austral stars VLT/NACO near-infrared Lyot-coronographic observations. Astron. Astrophys. **509**, A52 (2010)

Claudi, R.U., Turatto, M., Gratton, R.G., Antichi, J., Bonavita, M., Bruno, P., Cascone, E., De Caprio, V., Desidera, S., Giro, E., Mesa, D., Scuderi, S., Dohlen, K., Beuzit, J.L., Puget, P.: SPHERE IFS: the spectro differential imager of the VLT for exoplanets search. In: SPIE Conference Series, vol. 7014, p. 70143E, 11 pp. (2008)

Claudi, R., Giro, E., Turatto, M., Baruffolo, A., Bruno, P., Cascone, E., DeCaprio, V., Desidera, S., Dorn, R., Fantinel, D., Finger, G., Gratton, R., Lessio, L., Lizon, J.L., Maire, A.L., Mesa, D., Salasnich, B., Scuderi, S., Zurlo, A., Dohlen, K., Beuzit, J.L., Mouillet, D., Puget, P., Wildi, F., Hubin, N., Kasper, M.: The SPHERE IFS at work. In: SPIE Conference Series, vol. 9147, p. 91471L, 13 pp. (2014)

Crepp, J.R., Pueyo, L., Brenner, D., Oppenheimer, B.R., Zimmerman, N., Hinkley, S., Parry, I., King, D., Vasisht, G., Beichman, C., Hillenbrand, L., Dekany, R., Shao, M., Burruss, R., Roberts, L.C., Bouchez, A., Roberts, J., Soummer, R.: Speckle suppression with the project 1640 integral field spectrograph. Astrophys. J. **729**, 132–139 (2011)

Currie, T., Burrows, A., Madhusidhan, N., Fukagawa, M., Girard, J.H., Dawson, R., Murray-Clay, R., Kenyon, M., Matsumura, S., Jayawardhana, R., Chambers, J., Bromley, B.: A combined very large telescope and Gemini study of the atmosphere of directly imaged planet β Pictoris b. Astrophys. J. **776**, 15–33 (2013)

Davies, R., Kasper, M.: Adaptive optics for astronomy. Ann. Rev. Astron. Astrophys. **50**, 305–351 (2012)

Dekany, R., Roberts, J., Burruss, R., Truong, T., Palmer, D., Guiwits, S., Hale, D., Angione, J., Baranec, C., Croner, E., Davis, J.T.C., Zolkower, J., Henning, J., McKenna, D., Bouchez, A.H., Dekany, R.: Palm-3000 on-sky results. In: Presented at 2nd International Conference on Adaptive Optics for Extremely Large Telescopes, Victoria. Online at http://ao4elt2.lesia.obspm.fr id. 4 (2011)

Delfosse, X., et al.: Astron. Astrophys. **364**, 217 (2000)

Deming, D., Seager, S., Richardson, L.J., Harrington, J.: Infrared radiation from an extrasolar planet. Nature **434**, 740–743 (2005)

de Pater, I., Lissauer, J.J.: Planetary Sciences, 2nd edn. Cambridge University Press, Cambridge (2010)

Dicke, R.H.: Phase-contrast detection of telescope seeing errors and their correction. Astrophys. J. **198**, 605–615 (1975)

Dohlen, K., Langlois, M., Saisse, M., Hill, L., Origne, A., Jacquet, M., Fabron, C., Blanc, J.-C., Llored, M., Carle, M., Moutou, C., Vigan, A., Boccaletti, A., Carbillet, M., Mouillet, D., Beuzit, J.-L.: The infra-red dual imaging and spectrograph for SPHERE: design and performance. In: SPIE Conference Series, vol. 7014, p. 70143L, 10 pp. (2008)

Dohlen, K., Wildi, F.P., Puget, P., Mouillet, D., Beuzit, J.-L.: SPHERE: confronting in-lab performance with system analysis predictions. In: Second International Conference on Adaptive Optics for Extremely Large Telescopes. Online at http://ao4elt2.lesia.obspm.fr id.75 (2011)

Doyle, L.R., Carter, J.A., Fabrycky, D.C., Slawson, R.W., Howell, S.B., Winn, J.N., Orosz, J.A., Prÿsa, A., Welsh, W.F., Quinn, S.N., Latham, D., Torres, G., Buchhave, L.A., Marcy, G.W., Fortney, J.J., Shporer, A., Ford, E.B., Lissauer, J.J., Ragozzine, D., Rucker, M., Batalha, N., Jenkins, J.M., Borucki, W.J., Koch, D., Middour, C.K., Hall, J.R., McCauliff, S., Fanelli, M.N., Quintana, E.V., Holman, M.J., Caldwell, D.A., Still, M., Stefanik, R.P., Brown, W.R., Esquerdo, G.A., Tang, S., Furesz, G., Geary, J.C., Berlind, P., Calkins, M.L., Short, D.R., Steffen, J.H., Sasselov, D., Dunham, E.W., Cochran, W.D., Boss, A., Haas, M.R., Buzasi, D., Fischer, D.: Kepler-16: a transiting circumbinary planet. Science, **333**, 1602–1606 (2011)

Ducourant, C., Teixeira, R., Chauvin, G., Daigne, G., Le Campion, J.-F., Song, I., Zuckerman, B.: An accurate distance to 2M1207Ab. Astron. Astrophys. **477**, L1–L4 (2008)

Dumusque, X., et al.: The HARPS search for southern extra-solar planets. XXX. Planetary systems around stars with solar-like magnetic cycles and short-term activity variation. Astron. Astrophys. **535**, A55 (2011a)

Dumusque, X., Santos, N.C., Udry, S., Lovis, C., Bonfils, X.: Planetary detection limits taking into account stellar noise. II. Effect of stellar spot groups on radial-velocities. Astron. Astrophys. **527**, A82 (2011b)

Dyson, F.W., Eddington, A.S., Davidson, C.: A determination of the deflection of light by the Sunś gravitational field, from observations made at the total eclipse of May 29, 1919. Phil. Trans. R. Soc. Lond. **220**, 291–333 (1920)

Eckart, A., Hippler, S., Glindemann, A., Hackenberg, W., Quirrenbach, A., Kalas, P., Kasper, M., Davies, R.I., Ott, T., Rabien, S., Butler, D., Holstenberg, H.-C., Looze, D., Rohloff, R.-R., Wagner, K., Wilnhammer, N., Hamilton, D., Beckwith, S.V.W., Appenzeller, I., Genzel, R.: ALFA: the MPIA/MPE laser guide star AO system. Exp. Astron. **10**, 1 (2000)

Esposito, S., Riccardi, A., Fini, L., Puglisi, A.T., Pinna, E., Xompero, M., Briguglio, R., Quirs-Pacheco, F., Stefanini, P., Guerra, J.C., Busoni, L., Tozzi, A., Pieralli, F., Agapito, G., Brusa-Zappellini, G., Demers, R., Brynnel, J., Arcidiacono, C., Salinari, P.: First light AO (FLAO) system for LBT: final integration, acceptance test in Europe, and preliminary on-sky commissioning results. In: Proceedings of SPIE, vol. 7736, p. 773609 (2010)

Esposito, S., Mesa, D., Skemer, A., Arcidiacono, C., Claudi, R.U., Desidera, S., Gratton, R., Mannucci, F., Marzari, F., Masciadri, E., Close, L., Hinz, P., Kulesa, C., McCarthy, D., Males, J., Agapito, G., Argomedo, J., Boutsia, K., Briguglio, R., Brusa, G., Busoni, L., Cresci, G., Fini, L., Fontana, A., Guerra, J.C., Hill, J.M., Miller, D., Paris, D., Pinna, E., Puglisi, A., Quiros-Pacheco, F., Riccardi, A., Stefanini, P., Testa, V., Xompero, M., Woodward, C.: LBT observations of the HR 8799 planetary system first detection of HR 8799e in H band. Astron. Astrophys. **549**, A52 (2013)

Fortney, J.J., Marley, M.S., Lodders, K., Saumon, D., Freedman, R.: Comparative planetary atmospheres: model of TrES–1 and HD 209458b. Astrophys. J. **627**, L69–L72 (2005)

Foy, R., Labeyrie, A.: Feasibility of adaptive telescope with laser probe. Astron. Astrophys. **152**, L29–L31 (1985)

Fressin, F., Torres, G., Rowe, J.F., Charbonneau, D., Rogers, L.A., Ballard, S., Batalha, N.M., Borucki, W.J., Bryson, S.T., Buchhave, L.A., Ciardi, D.R., Désert, J.-M., Dressing, C.D., Fabrycky, D.C., Ford, E.B., Gautier, T.N., III, Henze, C.E., Holman, M.J., Howard, A.H., Steve, B., Jenkins, J.M., Koch, D.G., Latham, D.W., Lissauer, J.J., Marcy, G.W., Quinn, S.N., Ragozzine, D., Sasselov, D.D., Seager, S., Barclay, T., Mullally, F., Seader, S.E., Still, M., Twicken, J.D., Thompson, S.E., Uddin, K.: Two Earth-sized planets orbiting Kepler-20. Nature **482**, 195–198 (2012)

Fugate, R.Q.: High bandwidth laser guide star adaptive optics at starfire optical range. In: Ulrich, M.-H. (ed.) Progress in Telescope and Instrumentation Technologies. ESO Conference and Workshop Proceedings, vol. 42, p. 407. ESO, Garching (1992)

Fusco, T., Rousset, G., Beuzit, J.-L., Mouillet, D., Dohlen, K., Conan, R., Petit, C., Montagnier, G.: Conceptual design of an extreme AO dedicated to extra-solar planet detection by the VLT-Planet Finder instrument. In: SPIE Conference Series, vol. 5903, pp. 178–189 (2005)

Fusco, T., Rousset, G., Sauvage, J.-F., Petit, C., Beuzit, J.-L., Dohlen, K., Mouillet, D., Charton, J., Nicolle, M., Kasper, M., Baudoz, P., Puget, P.: High-order adaptive optics requirements for

direct detection of extrasolar planets: application to the SPHERE instrument. Opt. Express **14**, 7515 (2006)

Galicher, R., Rameau, J., Bonnefoy, M., Baudino, J.-L., Currie, T., Boccaletti, A., Chauvin, G., Lagrange, A.-M., Marois, C.: Near-infrared detection and characterization of the exoplanet HD95086 b with the Gemini planet imager. Astron. Astrophys. **565**, L4–L7 (2014)

Gratton, R., Bonavita, M., Desidera, S., Boccaletti, A., Kasper, M., Kerber, F.: Scientific output of single aperture imaging of exoplanets. In: Coude du Foresto, V., Gelino, D.M., Ribas, I. (eds.) Pathways Towards Habitable Planets. ASP Conference Series, vol. 430, pp. 37–44 (2010)

Guyon, O., Pluzhnik, E.A., Kuchner, M.J., Collins, B., Ridgway, S.T.: Theoretical limits on extrasolar terrestrial planet detection with coronagraphs. Astrophys. J. Suppl. Ser. **167**, 81–99 (2006)

Hayano, Y., Takami, H., Oya, S., Hattori, M., Saito, Y., Watanabe, M., Guyon, O., Minowa, Y., Egner, S.E., Ito, M., Garrel, V., Colley, S., Golota, T., Iye, M.: Commissioning status of Subaru laser guide star adaptive optics system. In: Adaptive Optics Systems II. Proceedings SPIE Conference Series, vol. 7736, p. 77360N (2010)

Hinkley, S., Oppenheimer, B.R., Soummer, R., Sivaramakrishnan, A., Roberts, L.C. Jr., Kuhn, J., Perrin, M.D., LLoyd, J.P., Kratter, K., Brenner, D.: Temporal evolution of coronagraphic dynamic range and constraints on companions to Vega. Astrophys. J. **654**, 633–640 (2007)

Hinkley, S., Oppenheimer, B.R., Zimmerman, N., Brenner, D., Parry, I.R., Crepp, J.R., Vashist, G., Ligon, E., King, D., Soummer, R., Sivaramakrishnan, A., Beichman, C., Shao, M., Roberts, L.C. Jr., Bouchez, A., Dekany, R., Pueyo, L., Roberts, J.E., Lockhart, T.: A new high contrast imaging program at Palomar observatory. Publ. Astron. Soc. Pac. **123**, 74 (2011)

Hugot, E., Ferrari, M., El Hadi, K., Costille, A., Dohlen, K., Rabou, P., Puget, P., Beuzit, J.-L.: Active optics methods for exoplanet direct imaging (Research Note) Stress polishing of supersmooth aspherics for VLT-SPHERE planet finder. Astron. Astrophys. **538**, A139 (2012)

Janson, M., Bergfors, C., Goto, M., Brandner, W., Lafreniére, D.: Spatially resolved spectroscopy of the exoplanet HR8799 c. Astrophys. J. **710**, L35–L38 (2010)

Janson, M., Brandt, T.D., Kuzuhara, M., Spiegel, D.S., Thalmann, C., Currie, T., Bonnefoy, M., Zimmerman, N., Sorahana, S., Kotani, T., Schlieder, J., Hashimoto, J., Kudo, T., Kusakabe, N., Abe, L., Brandner, W., Carson, J.C., Egner, S., Feldt, M., Goto, M., Grady, C.A., Guyon, O., Hayano, Y., Hayashi, M., Hayashi, S., Henning, T., Hodapp, K.W., Ishii, M., Iye, M., Kandori, R., Knapp, G.R., Kwon, J., Matsuo, T., McElwain, W., Mede, K., Miyama, S., Morino, J.-I., Moro-Martin, A., Nakagawa, T., Nishimura, T., Pyo, T.-S., Serabyn, E., Suenaga, T., Suto, H., Suzuki, R., Takahashi, Y., Takami, M., Takato, N., Terada, H., Tomono, D., Turner, E.L., Watanabe, M., Wisniewski, J., Yamada, T., Takami, H., Usuda, T., Tamura, M.: Direct imaging of methane in the atmosphere of GJ 504 b. Astrophys. J. Lett. **778**, L4–L6 (2013)

Kalas, P., Graham, J.R., Chiang, E., Fitzgerald, M.P., Clampin, M., Kite, E.S., Stapelfeldt, K., Marois, C., Krist, J.: Optical images of an exosolar planet 25 light-years from Earth. Science **322**, 1345–1348 (2008)

Kalas, P., Graham, J.R., Fitzgerald, M.P., Clampin, M.: STIS coronagraphic imaging of Fomalhaut: main belt structure and the orbit of Fomalhaut b. Astrophys. J. **775**, 56–86 (2013)

Kaltenegger, L., Selsis, F.: ESA white paper: atmospheric modeling: setting biomarkers in context. Eprint arXiv: 0809.4042 (2008)

Kaltenegger, L., Selsis, F., Fridlund, M., Lammer, H., Beichman, Ch., Danchi, W., Eiroa, C., Henning, T., Herbst, T., Léger, A., Liseau, R., Lunine, J., Paresce, F., Penny, A., Quirrenbach, A., Röttgering, H., Schneider, J., Stam, D., Tinetti, G., White, G.J.: Deciphering spectral fingerprints of habitable extrasolar planets. Astrobiology **10**, 89–102 (2010)

Kasting, J.F., Catling, D.: Evolution of a habitable planet. Ann. Rev. Astron. Astrophys. **41**, 429–463 (2003)

Kasper, M.E., Beuzit, J.-L., Verinaud, C., Yaitskova, N., Baudoz, P., Boccaletti, A., Gratton, R.G., Hubin, N., Kerber, F., Roelfsema, R., Schmid, H.M., Thatte, N.A., Dohlen, K., Feldt, M., Venema, L., Wolf, S.: EPICS: the exoplanet imager for the E-ELT. In: SPIE Conference Series, vol. 7015, p. 70151S, 12 pp. (2008)

Kasper, M., Amico, P., Pompei, E., Ageorges, N., Apai, D., Argomedo, J., Kornweibel, N., Lidman, C.: Direct imaging of exoplanets and brown dwarfs with the VLT: NACO pupil-stabilised Lyot coronagraphy at 4 μm. Messenger **137**, 8–13 (2009)

Knutson, H.A., Charbonneau, D., Allen, L.E., Fortney, J.J., Agol, E., Cowan, N.B., Showman, A.P., Cooper, C.S., Megeath, S.T.: A map of the day-night contrast of the extrasolar planet HD 189733b. Nature **447**, 183–186 (2007)

Konopacky, Q.M., Barman, T.S., Macintosh, B.A., Marois, C.: Detection of carbon monoxide and water absorption lines in an exoplanet atmosphere. Science **339**, 1398–1401 (2013)

Kuchner, M.J., Traub, W.A.: A coronagraph with a band-limited mask for finding terrestrial planets. Astrophys. J. **570**, 900–908 (2002)

Kuhn, J.R., Potter, D., Parise, B.: Imaging polarimetric observations of a new circumstellar disk system. Astrophys. J. **553**, L189–L191 (2001)

Kuzuhara, M., Tamura, M., Kudo, T., Janson, M., Kandori, R., Brandt, T.D., Thalmann, C., Spiegel, D., Biller, B., Carson, J., Hori, Y., Suzuki, R., Burrows, A., Henning, T., Turner, E. L., McElwain, M.W., Moro-Martín, A., Suenaga, T., Takahashi, Y.H., Kwon, J., Lucas, P., Abe, L., Brandner, W., Egner, S., Feldt, M., Fujiwara, H., Goto, M., Grady, C.A., Guyon, O., Hashimoto, J., Hayano, Y., Hayashi, M., Hayashi, S.S., Hodapp, K.W., Ishii, M., Iye, M., Knapp, G.R., Matsuo, T., Mayama, S., Miyama, S., Morino, J.-I., Nishikawa, J., Nishimura, T., Kotani, T., Kusakabe, N., Pyo, T.-S., Serabyn, E., Suto, H., Takami, M., Takato, N., Terada, H., Tomono, D., Watanabe, M., Wisniewski, J.P., Yamada, T., Takami, H., Usuda, T.: Direct imaging of a cold Jovian exoplanet in orbit around the Sun-like Star GJ 504. Astrophys. J. **774**, 11–28 (2013)

Lafrenière, D., Doyon, R., Marois, C., Nadeau, D., Oppenheimer, B.R., Roche, P.F., Rigaut, F., Graham, J.R., Jayawardhana, R., Johnstone, D., Kalas, P.G., Macintosh, B., Racine, R.: The gemini deep planet survey. Astrophys. J. **670**, 770–780 (2007a)

Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., Artigau, E.: A new algorithm for point-spread function subtraction in high-contrast imaging: a demonstration with angular differential imaging. Astrophys. J. **660**, 1367–1390 (2007b)

Lafrenière, D., Jayawardhana, R., van Kerkwijk, M.H.: Direct imaging and spectroscopy of a planetary-mass candidate companion to a young solar analog. Astrophys. J. **689**, L153–L156 (2008)

Lagrange, A.-M., Gratadour, D., Chauvin, G., Fusco, T., Eherenreich, D., Mouillet, D., Rousset, G., Rouan, D., Allard, F., Gendron, É., Charton, J., Mugnier, L., Rabou, P., Montri, J., Lacombe, F.: A probable giant planet imaged in the β Pictoris disk VLT/NaCo deep L'-band imaging. Astron. Astrophys. **493**, L21–L25 (2009)

Laughlin, G., Marcy, G.W., Vogt, S.S., Fisher, D.A., Butler, R.P.: On the eccentricity of HD209458b. Astrophys. J. **629**, L121–L124 (2005)

Lenzen, R., Hartung, M., Brandner, W., Finger, G., Hubin, N.N., Lacombe, F., Lagrange, A.-M., Lehnert, M.D., Moorwood, A.F.M., Mouillet, D.: NAOS-CONICA first on sky results in a variety of observing modes. In: SPIE Conference Series, vol. 4841, pp. 944–952 (2003)

Lenzen, R., Close, L., Brandner, W., Biller, B., Hartung, M.: A novel simultaneous differential imager for the direct imaging of giant planets. In: SPIE Conference Series, vol. 5492, p. 970 (2004)

Lissauer, J.J., Fabrycky, D.C., Ford, E.B., Borucki, W.J., Fressin, F., Marcy, G.W., Orosz, J.A., Rowe, J.F., Torres, G., Welsh, W.F., Batalha, N.M., Bryson, S.T., Buchhave, L.A., Caldwell, D.A., Carter, J.A., Charbonneau, D., Christiansen, J.L., Cochran, W.D., Desert, J.-M., Dunham, E.W., Fanelli, M.N., Fortney, J.J., Gautier, III T.N., Geary, J.C., Gilliland, R.L., Haas, M.R., Hall, J.R., Holman, M.J., Koch, D.G., Latham, D.W., Lopez, E., McCauliff, S., Miller, N.,

Morehead, R.C., Quintana, E.V., Ragozzine, D., Sasselov, D., Short, D.R., Jason, H., Steffen, J.H.: A closely packed system of low-mass, low-density planets transiting Kepler-11. Nature **470**, 53–58 (2011)

Lissauer, J.J., Dawson, R.I., Tremaine, S.: Advances in exoplanet science from Kepler. Nature **513**, 336–344 (2014)

Lunine, J.I., Fischer, D., Hammel, H.B., Henning, T., Hillenbrand, L., Kasting, J., Laughlin, G., Macintosh, B., Marley, M., Melnick, G., Monet, D., Noecker, C., Peale, S., Quirrenbach, A., Seager, S., Winn, J.N.: Worlds beyond: a strategy for the detection and characterization of exoplanets. Executive summary of a report of the exoplanet task force astronomy and astrophysics advisory committee, Washington, DC June 23, 2008. Astrobiology **8**, 875–881 (2008)

Lyot, M.B.: A study of the Solar corona and prominences without eclipses. Mon. Not. R. Astron. Soc. **99**, 580–594 (1939)

Macintosh, B.A., Graham, J.R., Palmer, D.W., Doyon, R., Dunn, J., Gavel, D.T., Larkin, J., Oppenheimer, B., Saddlemyer, L., Sivaramakrishnan, A., Wallace, J.K., Bauman, B., Erickson, D.A., Marois, C., Poyneer, L.A., Soummer, R.: The Gemini planet imager: from science to design to construction, adaptive optics systems. In: Hubin, N., Max, C.E., Wizinowich, P.L. (eds.) Proceedings of the SPIE, vol. 7015, pp. 701518 (2008)

Macintosh, B., Graham, J.R., Ingraham, P., Konopacky, Q., Marois, C., Perrin, M., Poyneer, L., Bauman, B., Barman, T., Burrows, A.S., Cardwell, A., Chilcote, J., De Rosa, R.J., Dillon, D., Doyon, R., Dunn, J., Erikson, D., Fitzgerald, M.P., Gavel, D., Goodsell, S., Hartung, M., Hibon, P., Kalas, P., Larkin, J., Maire, J., Marchis, F., Marley, M.S., McBride, J., Millar-Blanchaer, M., Morzinski, K., Norton, A., Oppenheimer, B.R., Palmer, D., Patience, J., Pueyo, L., Tantakyro, F., Sadakuni, N., Saddlemyer, L., Savransky, D., Serio, A., Soummer, R., Sivaramakrishnan, A., Song, I., Thomas, S., Wallace, J.K., Wiktorowicz, S., Wolff, S.: First light of the gemini planet imager. Proc. Nat. Acad. Sci. **111**, 12661–12666 (2014)

Makarov, V.V., Kaplan, G.H.: Statistical constraints for astrometric binaries with nonlinear motion. Astron. J. **129**, 2420–2427 (2005)

Marchetti, E., Brast, R., Delabre, B., Donaldson, R., Fedrigo, E., Frank, C., Hubin, N., Kolb, J., Lizon, J.-L., Marchesi, M., Oberti, S., Reiss, R., Soenke, C., Tordo, S., Baruffolo, A., Bagnara, P., Amorim, A., Lima, J.: MAD on sky results in star oriented mode. In: Adaptive Optics Systems. Proceedings of SPIE Conference Series, vol. 7015, p. 70150F. SPIE, Bellingham (2008)

Marley, M.S., Fortney, J.J., Hubickyj, O., Bodenheimer, P., Lissauer, J.J.: On the luminosity of young Jupiters. Astrophys. J. **655**, 541–549 (2007)

Marois, C., Doyon, R., Nadeau, D., Racine, R., Riopel, M., Vallée, P., Lafreniére, D.: TRIDENT: an infrared differential imaging camera optimized for the detection of methanated substellar companions. Publ. Astron. Soc. Pac. **117**, 745–756 (2005)

Marois, C., Lafreniére, D., Macintosh, B., Doyon, R.: Accurate astrometry and photometry of saturated and coronographic point spread functions. Astrophys. J. **647**, 612–619 (2006a)

Marois, C., Lafreniére, D., Doyon, R., Macintosh, B., Nadeau, D.: Angular differential imaging: a powerful high-contrast imaging technique. Astrophys. J. **641**, 556–564 (2006b)

Marois, C., Macintosh, B., Barman, T., Zuckerman, B., Song, I., Patience, J., Lafreniére, D., Doyon, R.: Direct imaging of multiple planets orbiting the Star HR8799. Science **322**, 1348–1352 (2008)

Marois, C., Zuckerman, B., Konopacky, Q.M., Macintosh, B., Barman, T.: Images of a fourth planet orbiting HR 8799. Nature **468**, 1080–1083 (2010)

Masciadri, E., Mundt, R., Henning, Th., Alvarez, C., Barrado y Navascués, D.: A search for hot massive extrasolar planets around nearby young Stars with the adaptive optics system NACO. Astrophys. J. **625**, 1004–1018 (2005)

Mawet, D., Riaud, P., Absil, O., Surdej, J.: Annular groove phase mask coronagraph. Astrophys. J. **633**, 1191–1200 (2005)

Mawet, D., Absil, O., Delacroix, C., Girard, J.H., Milli, J., O'Neal, J., Baudoz, P., Boccaletti, A., Bourget, P., Christiaens, V., Forsberg, P., Gonte, F., Habraken, S., Hanot, C., Karlsson, M., Kasper, M., Lizon, J.-L., Muzic, K., Olivier, R., Peña, E., Slusarenko, N., Tacconi-Garman, L.E., Surdej, J.: L'-band AGPM vector vortex coronagraph's first light on VLT/NACO, Discovery of a late-type companion at two beamwidths from a F0V star. Astron. Astrophys. **552**, L13 (2013)

Max, C.E., Olivier, S.S., Friedman, H.W., An, J., Avicola, K., Beeman, B.V., Bissinger, H.D., Brase, J.M., Erbert, G.V., Gavel, D.T., Kanz, K., Liu, M.C., Macintosh, B., Neeb, K.P., Patience, J., Waltjen, K.E.: Image improvement from a sodium-layer laser guide star adaptive optics system. Science **277**, 1649 (1997)

Mayer, L., Quinn, T., Wadsley, J., Stadel, J.: Formation of giant planets by fragmentation of protoplanetary disks. Science **298**, 1756–1759 (2002)

Mayor, M., Lovis, C., Santos, N.C.: Doppler spectroscopy as a path to the detection of Earth-like planets. Nature **513**, 328 (2014)

Meadows, V., Seager S.: Terrestrial planet atmospheres and biosignatures. In: Seager, S. (ed.) Exoplanets, pp. 441–470. University of Arizona Press, Tucson (2010). ISBN 978-0-8165-2945-2

Merkle, F., Kern, P., Léna, P., Rigaut, F., Fontanella, J.C., Rousset, G., Boyer, C., Gaffard, J.P., Jagourel, P.: Successful tests of adaptive optics. Messenger **58**, 1–4 (1989)

Mizuno, H.: Formation of the giant planets. Prog. Theor. Phys. **64**, 544–557 (1980)

Neuhäuser, R., Guenther, E.W., Alves, J., Huélamo, N., Ott, Th., Eckart, A.: An infrared imaging search for low-mass companions to members of the young nearby β Pic and Tucana/Horologium associations. Astron. Nachr. **324**, 535–542 (2003)

Neuhäuser, R., Guenther, E.W., Wuchterl, G., Mugrauer, M., Bedalov, A., Hauschildt, P.H.: Evidence for a co-moving sub-stellar companion of GQ Lup. Astron. Astrophys. **435**, L13–L16 (2005)

Nielsen, E.L., Close, L.M., Biller, B.A., Masciadri, E., Lenzen, R.: Constraints on extrasolar planet populations from VLT NACO/SDI and MMT SDI and direct adaptive optics imaging surveys: giant planets are rare at large separations. Astrophys. J. **674**, 466–481 (2008)

Oppenheimer, B.R., Hinkley, S.: High contrast observations in optical and infrared astronomy. Ann. Rev. Astron. Astrophys. **47**, 253–289 (2009)

Oppenheimer, B.R., Golimowski, D.A., Kulkarni, S.R., Matthews, K., Nakajima, T., Creech-Eakman, M., Durrance, S.T.: A coronagraphic survey for companions of stars within 8 Parsecs. Astron. J. **121**, 2189–2211 (2001)

Oppenheimer, B.R., Brenner, D., Hinkley, S., Zimmerman, N., Sivaramakrishnan, A., Soummer, R., Kuhn, J., Graham, J.R., Perrin, M., LLoyd, J.P., Roberts, L.C. Jr, Harrington, D.M.: The Solar-system-scale disk around AB Aurigae. Astrophys. J. **679**, 1574–1581 (2008)

Oppenheimer, B.R., Baranec, C., Beichman, C., Brenner, D., Burruss, R., Cady, E., Crepp, J.R., Dekany, R., Fergus, R., Hale, D., Hillenbrand, L., Hinkley, S., Hogg, D.W., King, D., Ligon, E.R., Lockhart, T., Nilsson, R., Parry, I.R., Pueyo, L., Rice, E., Roberts, J.E., Roberts, L.C. Jr., Shao, M., Sivaramakrishnan, A., Soummer, R., Truong, T., Vasisht, G., Veicht, A., Vescelus, F., Wallace, J.K., Zhai, C., Zimmerman, N.: Reconnaissance of the HR 8799 exosolar system. i. Near-infrared spectroscopy. Astrophys. J. **768**, 24–41 (2013)

Pepe, F.A., Lovis, C.: From HARPS to CODEX: exploring the limits of Doppler measurements. Phys. Scr. **T130**, 014007 (2008)

Perrin, M.D., Graham, J.R., Kalas, P., Lloyd, J.P., Max, C.E., Gavel, D.T., Pennington, D.M., Gates, E.L.: Laser guide star adaptive optics imaging polarimetry of Herbig Ae/Be stars. Science **303**, 1345–1348 (2004)

Peters-Limbach, M.A., Groff, T.D., Kasdin, N.J., Driscoll, D., Galvin, M., Foster, A., Carr, M.A., LeClerc, D., Fagan, R., McElwain, M.W., Knapp, G., Brandt, T., Janson, M., Guyon, O., Jovanovic, N., Martinache, F., Hayashi, M., Takato, N.: The optical design of CHARIS: an exoplanet IFS for the Subaru telescope. In: SPIE Conference Series, vol. 8864, p. 88641N (2013)

Petit, C., Sauvage, J.-F., Fusco, T., Sevin, A., Suarez, M., Costille, A., Vigan, A., Soenke, C., Perret, D., Rochat, S., Barrufolo, A., Salasnich, B., Beuzit, J.-L., Dohlen, K., Mouillet, D., Puget, P., Wildi, F., Kasper, M., Conan, J.-M., Kulcsár, C., Raynaud, H.-F.: SPHERE eXtreme AO control scheme: final performance assessment and on sky validation of the first auto-tuned LQG based operational system. In: SPIE Conference Series, vol. 9148, p. 91480O, 17 pp. (2014)

Pollack, J.B., Hubickyj, O., Bodenheimer, P., Lissauer, J.J., Formation of the giant planets by concurrent accretion of solids and gas. Icarus **124**, 62–85 (1996)

Racine, R., Walker, G.A.H., Nadeau, D., Doyon, R., Marois, C.: Speckle noise and the detection of faint companions. Publ. Astron. Soc. Pac. **111**, 587–594 (1999)

Ragazzoni, R.: Pupil plane wavefront sensing with an oscillating prism. J. Mod. Opt. **43**, 289 (1996)

Rameau, J., Chauvin, G., Lagrange, A.-M., Boccaletti, A., Quanz, S.P., Bonnefoy, M., Girard, J.H., Delorme, P., Desidera, S., Klahr, H., Mordasini, C., Dumas, C., Bonavita, M.: Discovery of a Probable 4-5 Jupiter-Mass exoplanet to HD 95086 by direct imaging. Astrophys. J. Lett. **772**, L15–L20 (2013a)

Rameau, J., Chauvin, G., Lagrange, A.-M., Meshkat, T., Boccaletti, A., Quanz, S.P., Currie, T., Mawet, D., Girard, J.H., Bonnefoy, M., Kenworthy, M.: Confirmation of the planet around HD 95086 by direct imaging. Astrophys. J. Lett. **779**, L26–L30 (2013b)

Rigaut, F., Rousset, G., Kern, P., Fontanella, J.C., Gaffard, J.P., Merkle, F., Léna, P.: Adaptive optics on a 3.6-m telescope: results and performance. Astron. Astrophys. **250**, 280–290 (1991)

Rousset, G., Fontanella, J.C., Kern, P., Gigan, P., Rigaut, F., Léna, P., Boyer, C., Jagourel, P., Gaffard, J.P., Merkle, F.: First diffraction-limited astronomical images with adaptive optics. Astron. Astrophys. **230**, L29–L32 (1990)

Rousset, G., Lacombe, F., Puget, P., Hubin, N.N., Gendron, E., Fusco, T., Arsenault, R., Charton, J., Feautrier, P., Gigan, P., Kern, P.Y., Lagrange, A.-M., Madec, P.-Y., Mouillet, D., Rabaud, D., Rabou, P., Stadler, E., Zins, G.: NAOS, the first AO system of the VLT: on-sky performance. In: SPIE Conference Series, vol. 4839, pp. 140–149 (2003)

Quanz, S.P., Lafrenière, D., Meyer, M.R., Reggiani, M.M., Buenzli, E.: Direct imaging constraints on planet populations detected by microlensing. Astron. Astrophys. **541**, 133–136 (2012)

Schneider, J., Deldieu, C., Le Sidaner, P., Savalle, R., Zolotukhin, I.: Defining and cataloguing exoplanets: the exoplanet.eu database. Astron. Astrophys. **532**, A79 (2013)

Seager, S.: Exoplanetary Atmospheres: Physical Process, By Sara Seager. Princeton University Press, Princeton (2010)

Seager, S., Sasselov, D.D.: Theoretical transmission spectra during extrasolar giant planet transits. Astrophys. J. **537**, 916–921 (2000)

Seager, S., Whitney, B.A., Sasselov, D.D.: Photometric light curves and polarization of close-in extrasolar giant planets. Astrophys. J. **540**, 504–520 (2000)

Sivaramakrishnan, A., Oppenheimer, B.R.: Astrometry and photometry with coronagraphs. Astrophys. J. **647**, 620–629 (2006)

Sivaramakrishnan, A., Koresko, C.D., Makidon, R.B., Berkefeld, T., Kuchner, M.J.: Ground based coronagraphy with high order adaptive optics. Astrophys. J. **552**, 397–408 (2001)

Soummer, R., Ferrari, A., Aime, C., Jolissaint, L.: Speckle noise and dynamic range in coronagraphic images. Astrophys. J. **669**, 642–656 (2007)

Soummer, R., Pueyo, L., Larkin, J.: Detection and characterization of exoplanets and disks using projections on Karhunen–Loéve Eigenimages. Astrophys. J. **755**, L28–L41 (2012)

Sparks, W.B., Ford, H.C.: Imaging spectroscopy for extrasolar planet detection. Astrophys. J. **578**, 543–564 (2002)

Spiegel, D.S., Burrows, A.: Spectral and photometric diagnostics of giant planet formation scenarios. Astrophys. J. **745**, 174–188 (2012)

Spitzer, L. Jr.: Space telescopes and components. Astron. J. **65**, 242–263 (1960)

Stam, D.M., Hovenier, J.W., Waters, L.B.F.M.: Using polarimetry to detect and characterize Jupiter-like extrasolar planets. Astron. Astrophys. **428**, 663–672 (2004)

Tamura, M., Hodapp, K., Takami, H., Abe, L., Suto, H., Guyon, O., Jacobson, S., Kandori, R., Morino, J.-I. Murakami, N., Stahlberger, V., Suzuki, R., Tavrov, A., Yamada, H., Nishikawa, J.,

Ukita, N., Hashimoto, J., Izumiura, H., Hayashi, M., Nakajima, T., Nishimura, T.: Concept and science of HiCIAO: high contrast instrument for the Subaru next generation adaptive optics. In: SPIE Conference Series, vol. 6269, p. 62690V (2006)

Thalmann, C., Schmid, H.M., Boccaletti, A., Mouillet, D., Dohlen, K., Roelfsema, R., Carbillet, M., Gisler, D., Beuzit, J.-L., Feldt, M., Gratton, R., Joos, F., Keller, C.U., Kragt, J., II, Pragt, J.H., Puget, P., Rigal, F., Snik, F., Waters, R., Wildi, F.: SPHERE ZIMPOL: overview and performance simulation. In: SPIE Conference Series, vol. 7014, p. 70143F, 12 pp. (2008)

Thalmann, C., Carson, J., Janson, M., Goto, M., McElwain, M., Egner, S., Feldt, M., Hashimoto, J., Hayano, Y., Henning, T., Hodapp, K.W., Kandori, R., Klahr, H., Kudo, T., Kusakabe, N., Mordasini, C., Morino, J.-I., Suto, H., Suzuki, R., Tamura, M.: Discovery of the coldest imaged companion of a Sun-like star. Astrophys. J. Lett. **707**, L123–L127 (2009)

Thatte, N., Abuter, R., Tecza, M., Nielsen, E.L., Clarke, F.J., Close, L.M.: Very high contrast integral field spectroscopy of AB Doradus C: 9-mag contrast at 0.2 arcsec without a coronagraph using spectral deconvolution. Mon. Not. R Astron. Soc. **378**, 1229–1236 (2007)

Tinetti, G., Encrenaz, T., Coustenis, A.: Spectroscopy of planetary atmospheres in our Galaxy. Astron. Astrophys. Rev. **21**, 63 (2013)

Traub, W.A., Oppenheimer, B.R.: Direct imaging of exoplanets. In: Seager, S. (ed.) Exoplanets, 526 pp. University of Arizona Press, Tucson (2010)

Turnbull, M.C., Traub, W.A., Jucks, K.W., Woolfs, N.J., Meyer, M.R., Gorlova, N., Skrutskie, M.F., Wilson, J.C.: Spectrum of habitable world: Earthshine in the near-infrared. Astrophys. J. **644**, 551–559 (2006)

Vanderbei, R.J., Cady, E., Kasdin, N.J.: Optimal occulter design for finding extrasolar planets. Astrophys. J. **665**, 794–798 (2007)

Vidal-Madjar, A., Désert, J.-M., Lecavelier des Etangs, A., Hébrard, G., Ballester, G.E., Ehrenreich, D., Ferlet, D., McConnell, J.C., Mayor, M., Parkinson, C.D.: Detection of oxygen and carbon in the hydrodynamically escaping atmosphere of the extrasolar planet HD 209458B. Astrophys. J. **604**, L69–L72 (2004)

Vigan, A., Langlois, M., Moutou, C., Dohlen, K.: Exoplanet characterization with long slit spectroscopy. Astron. Astrophys. **489**, 1345–1354 (2008)

Vigan, A., Moutou, C., Langlois, M., Allard, F., Boccaletti, A., Carbillet, M., Mouillet, D., Smith, I.: Photometric characterization of exoplanets using angular and spectral differential imaging. Mon. Not. R. Astron. Soc. **407**, 71–82 (2010)

Wakeford, H.R., Sing, D.K.: Transmission spectral properties of clouds for hot Jupiter exoplanets. Astron. Astrophys. **573**, A122 (2015)

Wizinowich, P.L., Le Mignant, D., Bouchez, A.H., Campbell, R.D., Chin, J.C.Y., Contos, A.R., van Dam, M.A., Hartman, S.K., Johansson, E.M., Lafon, R.E., Lewis, H., Stomski, P.J., Summers, D.M., Brown, C.G., Danforth, P.M., Max, C.E., Pennington, D.M.: The W. M. Keck observatory laser guide star adaptive optics system: overview. Publ. Astron. Soc. Pac. **118**, 297–309 (2006)