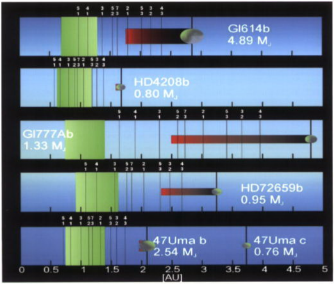
PHYSICS TEXTBOOK

Edited by Rudolf Dvorak

WILEY-VCH

Extrasolar Planets

Formation, Detection and Dynamics



Extrasolar Planets

Edited by
Rudolf Dvorak

Related Titles

S.N. Shore

Astrophysical Hydrodynamics

An Introduction

2007
ISBN 978-3-527-40669-2

M. Salaris

Evolution of Stars and Stellar Populations

2005
E-Book
ISBN: 978-0-470-09222-4

S.N. Shore

The Tapestry of Modern Astrophysics

2003
ISBN 978-0-471-16816-4

Extrasolar Planets

Formation, Detection and Dynamics

Edited by
Rudolf Dvorak

WILEY-VCH Verlag GmbH & Co. KGaA

The Editor

Prof. Dr. Rudolf Dvorak
University of Vienna
Institute for Astronomy
Tuerkenschanzstrasse 17
1180 Wien
Austria

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de

Printed in the Federal Republic of Germany
Printed on acid-free paper

Typesetting Uwe Krieg, Berlin
Printing betz-druck GmbH, Darmstadt
Bookbinding Litges & Dopf GmbH, Heppenheim

ISBN: 978-3-527-40671-5

Contents

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

Prolog

The plans for this book on extrasolar planets go back to 2005, when the publisher WILEY-VCH approached me to edit a book on this hot topic in astronomy. My contacts to colleagues from all over the world made the decision to accept this task easy. Nevertheless, it took quite a time to persuade them to contribute to the book with a competent review in their field of research. The main problem is that this topic of astronomy is developing very rapidly, maybe even faster than other fields. This is due to the development of new and better instruments on the ground and – in future – from space. Nevertheless, we have tried our best and I may say that all the chapters are top of the line with regard to the topic with which they are dealing. Nothing more needs to be said, because in the next chapter, a kind of prolog, one of the leading scientists in the world dealing with extrasolar planets, Dr. Jean Schneider, who maintains the exocatalogue website, highlights this exciting field of Astronomy and Astrophysics.

Rudolf Dvorak
AstroDynamicsGroup
Institute of Astronomy
University of Vienna

Preface

Exoplanets, an infinite territory.
Jean Schneider – Paris Observatory

In this book, the reader will find current theoretical studies and observational results about exoplanets.

Exoplanets occupy a particular position in science. Everybody is able to grasp the general idea of this domain of research, but at the same time it constitutes one of the most difficult research fields in astronomy, since extra solar planets are very hard to detect and investigate. The goal of astronomers and planetologists is clear: comparative planetology will be expanded beyond the boundaries of our solar system and extended to different planetary systems. Another purpose of this new science is to correlate planetary system properties to their parent star characteristics. To date, we have a very modest understanding of exoplanets. In most cases we know essentially the mass and some orbital parameters. Only for a few transiting planets do we have some insight about their atmospheres and radii. But even at this early stage, surprises and unexplained discoveries have occurred: planets around pulsars, for example, semi-major axes a hundred times smaller than expected, high eccentric orbits for half of the planets, unexpectedly large radii. No doubt more surprises will arrive with future observations, especially with direct imaging.

Lots of new questions can naturally be raised. How frequent are moons, rings, Trojan planets? Are there binary planets out there? All of these objects would have a significant impact on the investigation of planetary systems. Moons and Trojan planets are an interesting class of potentially habitable abodes, rings may perturb the determination of the planetary size (especially by direct imaging), binary planets may affect the histogram of planet masses/radii. Interestingly, these objects are not much harder to detect than the planet itself. Are there special orbits (e.g., “eccentric resonant”)? A new class of objects is particularly intriguing: the “free-floating” planets, i.e., planets without a parent star. Are there two types of free-floating planets? Those formed in situ in dense clouds and those ejected dynamically from a planetary

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

system? Since planets have been detected around two pulsars, it is reasonable to ask whether planets can exist around white dwarfs. If so, we can speculate about "second generation planets" (B. Lopez).

To be honest, these items are only a list of curiosities which do not open up a new branch of physics, such as gamma ray bursts, dark matter, universal cosmic repulsion ("dark energy") or the cosmological singularity. The real interest in and motivation for exoplanet search is the possibility of finding another form of life. I will not enter the debate of a "definition" of the essence of life (if there is any). All that astronomers can do is to propose and use scientific criteria to search for biosignatures. In the solar system, only two or three places might be inhabited, and, if there is a possible form of life, it will be very simple and unevolved (microbe-like). A neglected research topic is the modeling of forms of life not based on carbon chemistry (auto-organizing structures in different media) and the corresponding biosignatures.

What is encouraging with exoplanets is that there are at least 15 billion giant planets in the galaxy. And, most probably, there are even more terrestrial planets, some of them in the habitable zone of their parent star. In addition, there presently is no observational evidence against the hypothesis of an evolved form of life on some of these planets (with the exception of the embarrassing Fermi Paradox which I will not discuss here). Note that today biosignatures such as oxygen and ozone are accepted without great criticism. But, as W. Benz puts it "Extraordinary claims (i.e., the discovery of an exolife) require extraordinary evidence". We can already anticipate that, if oxygen and ozone spectral bands are detected in 20–25 years in a habitable planet, biochemists, chemists and atmospherists will engage a serious technical debate on their exobiological significance. Indeed, the problem is the following: we have no idea of how rare life is on an exoplanet; therefore, if rare, unusual abiotic mechanisms can mimic biosignatures, we will have to compare two unknown probabilities. We finally have convincing reasons to be optimistic in the near and mid-term future. In the next 50 years, it is even sensible to think of searching for "technosignatures", i.e., signs of technological activities (e.g., "city lights"). This will be possible with large interferometers combined with high-resolution spectrographs able to record the spectrocartography of exoplanets. So even more surprises are waiting for us in the future, but also some considerations to bear in mind. We might be able to detect spectral biosignatures such as the colour of the "exovegetation" or atmospheric byproducts such as oxygen, but with any foreseeable technology it will be impossible to detect the morphology of any living creature. The detection of any morphological detail on a planet with a resolution of 1 meter at a distance of 5 pc in 10 hours, would require an interferometer with a collecting area of 10^7 km^2 and a baseline of 1 AU. For moving objects ("animals") the situation is even more challenging. We could consider going there, but then other difficulties emerge. At a speed of 0.1 c, an interstellar dust grain of 100µ has the same kinetic energy as a body weighing 1 ton at 100 km per hour. So either we find an efficient protection mechanism for an interstellar probe, or the speed must be reduced to a few hundred kilometers per second, forcing the travel to last 30 000 years.

A final word about SETI (Search for Extra-Terrestrial Intelligence). Whatever "Intelligence" means, we have to overcome the following difficulty: how can we detect non-human minds with a human-bound way of thinking? This is undoubtedly contrary to what naive scientists still believe 200 years after Kant's *Kritik der reinen Vernunft*, i.e., that concepts do not derive from experience and physical constitution.

In any case, exoplanet research has a brilliant future for the next decades and even centuries – assuming that we will last that long.

Jean Schneider

List of Contributors

Christian Beaugé
Observatorio Astronómico
Universidad Nacional de Córdoba
Laprida 854
(X5000B GR)
Córdoba
Argentina
beauge@oac.uncor.edu

Christine Bounama
Potsdam-Institut für
Klimafolgenforschung (PIK)
Telegrafenberg A62
14473 Potsdam
Germany
Christine.Bounama@pik-potsdam.de

Rudolf Dvorak
University of Vienna
Institute for Astronomy
Türkenschanzstrasse 17
1180 Vienna
Austria
dvorak@astro.univie.ac.at

Anders Erikson
Institut für Planetenforschung, DLR
Rutherfordstrasse 2
12489 Berlin
Germany
pks6n@viginia.edu

Sylvio Ferraz-Mello
Instituto de Astronomia
Geofísica e Ciências Atmosféricas
Universidade de São Paulo
Rua do Matão 1226
05508-900 São Paulo
Brazil
sylvio@usp.br

Siegfried Franck
Potsdam-Institut für
Klimafolgenforschung (PIK)
Telegrafenberg A62
14473 Potsdam
Germany
Siegfried.Franck@pik-potsdam.de

Malcolm Fridlund
Astrophysics Mission Division
Research and Scientific Support
Department ESA, ESTEC, SCI-SA
P.O. Box 299
Keplerlaan 1
NL-2200 AG, Noordwijk
The Netherlands
malcolm.fridlund@esa.int

Lisa Kaltenegger
Harvard-Smithsonian Center
for Astrophysics, MS-20
60 Garden Street
Cambridge, MA 02138
USA
lkaltenegger@cfa.harvard.edu

Maxim L. Khodachenko
Austrian Academy of Sciences
Space Research Institute
Schmiedlstrasse 6
8042 Graz
Austria
maxim.khodachenko@oeaw.ac.at

Wilhelm Kley
University of Tübingen
Astronomy and Astrophysics
Auf der Morgenstelle 10
72076 Tübingen
Germany
wilhelm.kley@uni-tuebingen.de

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

Yuri N. Kulikov
Russian Academy of Sciences
Polar Geophysical Institute (PGI)
Khalturina Str. 15
183010 Murmansk
Russian Federation
kulikov@pgi.ru

Helmut Lammer
Space Research Institute
Austrian Academy of Sciences
Schmiedlstrasse 6
8042 Graz
Austria
helmut.lammer@oeaw.ac.at

Herbert I. M. Lichtenegger
Space Research Institute
Austrian Academy of Sciences
Schmiedlstrasse 6
8042 Graz
Austria
herbert.lichtenegger@oeaw.ac.at

Tatiana A. Michtchenko
Instituto de Astronomia
Geofísica e Ciências Atmosféricas
Universidade de São Paulo
Rua do Matão 1226
05508-900 São Paulo
Brazil
tatiana@astro.iag.usp.br

Elke Pilat-Lohinger
University of Vienna
Institute for Astronomy
Türkenschanzstrasse 17
1180 Vienna
Austria
Lohinger@astro.univie.ac.at

Heike Rauer
Institut für Planetenforschung, DLR
Rutherfordstrasse 2
12489 Berlin
Germany
heike.rauer@dlr.de

Zsolt Sándor
Department of Astronomy
Loránd Eötvös University
Pázmány Péter sétány 1/A
1117 Budapest
Hungary
Zs.Sandor@astro.elte.hu

Jean Schneider
Observatoire de Paris
Section de Meudon
5, Place Jules Janssen
92195 Meudon Cedex
France
Jean.Schneider@obspm.fr

P. Kenneth Seidelmann
University of Virginia
Department of Astronomy
P.O. Box 400325
Charlottesville, VA 22904-4325
USA
pks6n@viginia.edu

Franck Selsis
CRAL-ENS
46, Allée d'Italie
69364 Lyon Cedex 07
France
franck.selsis@ens-lyon.fr

Werner von Bloh
Potsdam-Institut für
Klimafolgenforschung (PIK)
Telegrafenberg A31
14412 Potsdam
Germany
Werner.von.Bloh@pik-potsdam.de

1
Planetary Masses and Orbital Parameters from Radial Velocity Measurements

Christian Beaugé, Sylvio Ferraz-Mello, and Tatiana A. Michtchenko

Abstract

So far, practically all detections of extrasolar planets have been obtained from radial velocity data in which the presence of planetary bodies is deduced from temporal variations in the radial motion of the host star. To perform any dynamical study for these systems, it is necessary to specify: (i) initial conditions (mass and orbital elements) and (ii) an adequate coordinate system from which to construct the equations of motion. This chapter discusses both of these points.

In the first part, we introduce the reader to the process of orbital determination from Doppler data for both single and multiple exoplanetary systems. We distinguish between primary parameters (which are obtainable directly from the observational data) and secondary parameters requiring additional information about the system, such as the stellar mass or inclination of the orbital plane. For multiple planetary systems we also discuss the differences between Keplerian fits, in which the mutual perturbations between the planets are neglected, and dynamical (or N-body) fits.

The second part of the chapter is devoted to the construction of the equations of motion in different coordinate systems. Special attention is given to the Hamiltonian formalism in barycentric, Jacobi and Poincaré coordinates, and we explain how to obtain orbital elements in each case. Finally, we discuss some of their advantages and disadvantages, particularly with respect to orbital fits and general dynamical studies.

1.1
Exoplanet Detection

Planets are very dim objects, and their direct observation is an extremely difficult task. Even Jupiter, the biggest planet in our own solar system, has only about 10^{-9} times the luminosity of the sun, making a similar exoplanet unobservable to us by present techniques. The first direct observation of an exo-

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

planet (*GQ Lupi*) only occurred in 2005 with VLT and, as of July 2006, three other exoplanets have also been imaged (*2M1207*, *AB Pic* and *SCR 1845*). However, most of the exoplanetary bodies have never been seen at all.

If an exoplanet cannot be seen, how can we know it is there? The basic idea is that, even if invisible, the presence of a planetary body may affect the luminosity of the star, or its motion with respect to background objects. Thus, we may deduce the existence of a planet by analyzing changes in some observable aspect of the star it revolves around. Such "indirect" detection methods are the main backbone in current discovery strategies of exoplanets. Five different techniques have been proposed and developed in recent years: (i) Stellar Transit, (ii) Radial Velocity Curves (Doppler), (iii) Gravitational Micro-lensing, (iv) Stellar Interferometry and (v) Astrometry. For details on these and other proposed methods, the reader is referred to Perryman [1] for a very comprehensive review.

Although very promising, Micro-lensing and Astrometry are still far from fulfilling their potential. Only four planetary candidates have been proposed based on micro-lensing techniques (*OGLE-05-071L*, *OGLE-05-169L*, *OGLE-05-390L* and *OGLE235-MOA53*), and even though some estimate may be obtained concerning mass and orbital period, there is no information about the remaining orbital elements. A similar picture can be given for interferometric techniques, and at present only four positive detections are counted. Astrometry has yet to yield a discovery, although the projected launch of several space telescopes (e.g., Kepler, TPF) will almost certainly change this picture. Consequently, and at least at present, practically all the currently exoplanet population has been obtained either by Stellar Transit or Doppler. The former is the subject of another chapter of this book, while the latter is the main objective of the present text.

1.2
Radial Velocity in Astrocentric Elements

The observation of a Doppler shift of the spectral lines of a star denounces a change in the velocity of the star with respect to the observer. Since the observer himself is moving with a velocity $\sim 30\ \mathrm{km\,s^{-1}}$, variable in direction, it is necessary to subtract this motion from the observational data and reduce it to the barycenter of the solar system (for a description of the necessary operations see Ferraz-Mello et al. [2]).

The velocity of an isolated star, with respect to the barycenter of the solar system, is constant, at least for times short as compared to the timescale of galactic motion. However, if it has N planetary companions, the star will display a motion around the common barycenter of the system (see Fig. 1.1).

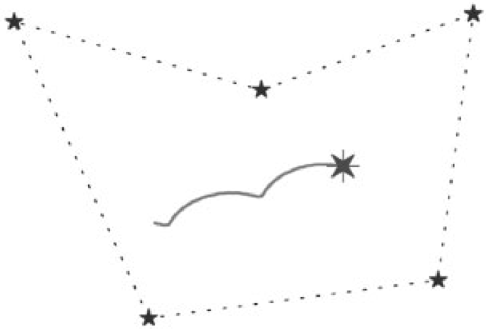


Fig. 1.1 The sinusoidal motion of a star due to the presence of a planetary companion. The faint field stars are used as a reference frame.

In order to understand this effect, we begin studying the kinematics of a single planet in elliptic orbit around the star. In an astrocentric reference frame, the position and velocity vectors of the planet are given by Brouwer and Clemence [3] and Murray and Dermott [4]:

$$\mathbf{r} = r\cos f\,\hat{\imath} + r\sin f\,\hat{\jmath} \tag{1.1}$$

$$\mathbf{v} = -\frac{2\pi a}{T\sqrt{1-e^2}}\left[\sin f\,\hat{\imath} - (e+\cos f)\,\hat{\jmath}\right]$$

where

$$r = \frac{a(1-e^2)}{1+e\cos f} \qquad \frac{df}{dt} = \frac{2\pi a^2}{Tr^2}\sqrt{1-e^2} \tag{1.2}$$

Here r is the magnitude of the radius vector $\mathbf{r}$, the velocity vector is denoted by $\mathbf{v}$, f is the true anomaly, a is the astrocentric semi-major axis, e is the eccentricity, and $\hat{\imath}$ and $\hat{\jmath}$ are two unit vectors in the orbital plane. The first is orientated in the direction of the pericenter, and the latter is orthogonal to it. The orbital period T can be obtained directly from Kepler's third law. Denoting by m_0 the mass of the star and m the mass of the planet, we have:

$$n^2a^3 = \mathcal{G}(m_0+m) \tag{1.3}$$

where the mean motion $n = 2\pi/T$ is the mean angular velocity along the orbit.

We must now transform these vectors to a new coordinate system that is independent of the plane of orbital motion. For exoplanets it is customary to use a modification of the so-called Herschel astrocentric coordinates, which were first developed for studies of visual double stars. It uses the sky (i.e., a plane tangent to the celestial sphere) as the reference plane, and a schematic view is presented in Fig. 1.2. The x-axis is taken along the intersection line between the orbital plane and the sky. Its direction is chosen towards γ, which is the node where the motion of the planet is directed towards the observer. The y-axis is also tangent to the celestial sphere, and is such that the resulting system is right-handed. Finally, the z-axis is directed along the line of sight, away from the observer.

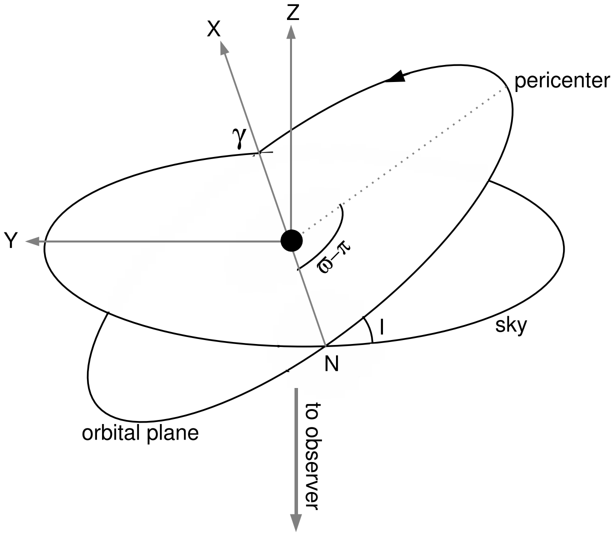


Fig. 1.2 The rotated astrocentric reference frame showing the orbital plane and the plane tangent to the celestial sphere (sky). The origin of the angles is the point γ.

In this coordinate system, the unit vectors $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$ of the orbital plane have components:

$$\hat{\imath} = \begin{pmatrix} \cos\omega \\ \sin\omega\cos I \\ -\sin\omega\sin I \end{pmatrix} \qquad \hat{\jmath} = \begin{pmatrix} -\sin\omega \\ \cos\omega\cos I \\ -\cos\omega\sin I \end{pmatrix} \qquad \hat{k} = \begin{pmatrix} 0 \\ \sin I \\ \cos I \end{pmatrix} \tag{1.4}$$

$\hat{k}$ is perpendicular to the orbital plane of the planet. This decomposition of the unit vectors is analogous to the transformations commonly used in celestial mechanics to pass to coordinates with respect to the ecliptic (except that here we fix $\Omega = \pi$). I is the inclination of the orbital plane with respect to the sky, and the argument of the pericenter is given by $\omega + \pi$. The addition of the angle π is due to the direction of the x-axis, which is chosen opposite to the "ascending node" N.

We can now obtain the components of **r** and **v** in this new reference frame. After a few simple algebraic manipulations, we obtain the velocity $\mathbf{v} = (v_x, v_y, v_z)$ where:

$$
\begin{aligned}
v_x &= -\frac{2\pi a}{T\sqrt{1-e^2}}\left[\sin(f+\omega) + e\sin\omega\right] \\
v_y &= \frac{2\pi a\ \cos I}{T\sqrt{1-e^2}}\left[\cos(f+\omega) + e\cos\omega\right] \\
v_z &= -\frac{2\pi a\ \sin I}{T\sqrt{1-e^2}}\left[\cos(f+\omega) + e\cos\omega\right]
\end{aligned}
\tag{1.5}
$$

Having the astrocentric velocity vector of the planet in the desired reference frame, we can pass to barycentric coordinates. Calling **V** the barycentric velocity vector of the planet, and $\mathbf{V}_*$ that of the star, we have that $\mathbf{v} = \mathbf{V} - \mathbf{V}_*$. On the other hand, since the barycenter is fixed in this reference frame, we have $m_0\mathbf{V}_* + m\mathbf{V} = 0$. Solving for $\mathbf{V}_*$, we obtain:

$$
\boldsymbol{V}_* = -\frac{m}{m_0 + m}\boldsymbol{v} \tag{1.6}
$$

which represents the velocity of the motion of the star around the center of mass of the system. To calculate the velocity actually detected by the observer, we must add the velocity $\boldsymbol{V}_0$ of the barycenter itself with respect to background stars.

It is useful to decompose this observable velocity into the tangential velocity component V_t and the radial velocity $V_r = V_{*z} + V_{0z}$. The former causes a displacement of the position of the star with respect to background stars. Its measurement is the role of astrometry but, as mentioned before, telescopes on earth are currently not able to detect these variations except in a few cases. The radial velocity V_r is far easier to detect, even with ground-based instruments, due to changes in the frequency (Doppler shift) of spectral lines from the star's spectrum. The best stellar candidates are those that, on one hand, contain a fair amount of absorption lines in the visible spectrum (i.e., must not be too hot) but, on the other hand, the number of lines must not be too large (i.e., the star must not be too cold). Thus, the best candidates are stars of spectral type F or G; in other words, similar to our own sun. A complete expression for the radial velocity can be found simply by substituting V_{*z} from Eqs. (1.5) and (1.6), and yields:

$$
V_r = \frac{2\pi a}{T\sqrt{1-e^2}}\frac{m\sin I}{(m+m_0)}\left[\cos(f+\omega) + e\cos\omega\right] + V_{r0} \tag{1.7}
$$

where $V_{r0} = V_{0z}$ is the (constant) reference radial velocity of the barycenter.

The extension to N planets is straightforward and follows the same lines, as long as we neglect mutual perturbations and assume Keplerian solutions. We can then write the complete radial velocity of the star at a given time t as:

$$V_r(t) = \sum_{i=1}^{N} \mathcal{K}_i \Big[\cos(f_i + \omega_i) + e_i \cos\omega_i\Big] + V_{r0} \tag{1.8}$$

where

$$\mathcal{K}_i = \frac{m_i \sin I_i}{\mathcal{M}} \frac{2\pi a_i}{T_i\sqrt{1-e_i^2}} \tag{1.9}$$

and

$$\mathcal{M} = \sum_{i=0}^{N} m_i \tag{1.10}$$

is the total mass of the system (star and planets). Transforming from semi-major axis to mean motions via Kepler's third law, we can then rewrite the coefficients $\mathcal{K}_i$ as:

$$\mathcal{K}_i = \Big(\mathcal{G}(M + m_i)\Big)^{1/3} \frac{m_i \sin I_i}{\mathcal{M}} n_i^{1/3} (1 - e_i^2)^{-1/2} \tag{1.11}$$

or, more succinctly, as

$$\mathcal{K}_i = \mathcal{F}_i(M, m_i, I_i) n_i^{1/3} (1 - e_i^2)^{-1/2} \tag{1.12}$$

where $\mathcal{F}_i(M, m_i, I_i)$ is sometimes called the "mass function" and groups all the terms that depend explicitly on the stellar and planetary masses, as well as the orbital inclination. This expression is valid only for astrocentric orbital elements. If Jacobian coordinates are employed, the expression given by Eq. (1.38) must be used.

The basis of the Doppler method is then to build an observational data base of the changes in the radial velocity of a target star. These radial velocity data points represent a discretized representation of the radial component of the left-hand side of Eq. (1.8). The idea now is to deduce, from this data set, the masses and orbital parameters of all the planetary companions that make up the right-hand member of the same equation.

Note that $V_r(t)$ is the sum of N periodic terms, each with semi-amplitude $\mathcal{K}_i$. However, the true anomaly f_i is only a linear function of time in the case of circular orbits $e_i = 0$. In the general elliptic case, only the mean anomaly ℓ_i has a constant derivative (given by the mean motion n_i). The relationship between f and ℓ is given in terms of the (intermediate) eccentric anomaly u, and via the following two equations:

$$\tan(f/2) = \sqrt{\frac{1+e}{1-e}}\tan(u/2) \tag{1.13}$$
$$u - e\sin u = \ell = n(t-\tau)$$

The second expression is the classical Kepler equation, and must be solved iteratively to obtain the passage from ℓ (or the time) to the eccentric anomaly. The quantity τ is sometimes referred to as the time of passage through the pericenter. Finally, the mean motion n is related to the semi-major axis and masses through Kepler's third law.

As an example, Fig. 1.3 shows the shape of two fictitious radial velocity curves, constructed from Eq. (1.8) with only one planet. The continuous line shows the case of a circular orbit ($e = 0$), while the dashed line presents an example of a highly elliptic body ($e = 0.6$). Although both periods and semi-amplitudes are the same, the second curve shows distinctive peaks each time the planet crosses the pericenter of its orbital motion. Another noticeable effect of the eccentricity is a change in the averaged value of V_r. Once again, this is due to the nonlinear behavior of the true anomaly f for noncircular orbits.

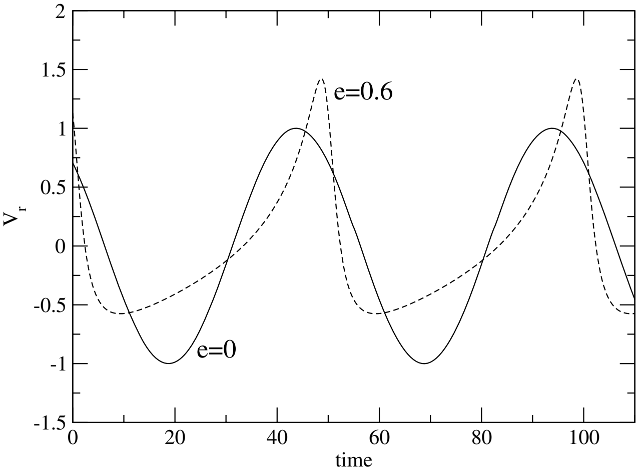


Fig. 1.3 Fictitious radial velocity curves, using Eq. (1.8) with $K = 1, \omega = 180$ degrees and $T = 50$ in arbitrary time units. Continuous line corresponds to a circular orbit, while the dashed curve was calculated with $e = 0.6$.

As a final important point, it must be stressed that there is no free angle equivalent to the longitude of the ascending node Ω that can be simply added to the orbital elements of the planets. Ω measures the angular distance from the x-axis and the "ascending node" $\mathbf{N}$, and in Herschel's modified coordinate system, as shown in Fig. 1.2, it is set to $\Omega = \pi$. Expressions (1.5) for the velocity components (v_x, v_y, v_z) were derived for this orientation of the x-axis, and thus implicitly depend on this choice of Ω. Any other value for Ω would be inconsistent with the orbit issued from the observations.

1.3
Orbital Fits from Radial Velocity Curves

1.3.1
Primary Parameters

Until recently, radial velocity data were zealously guarded by the observational teams and not available to the general scientific community. Fortunately this picture is changing (albeit slowly), and some information is currently available from the on-line versions of the published papers. This information is already pre-processed, in the sense that all the necessary steps have been taken to reduce the velocities to the barycenter of our solar system.

A real example of a radial velocity data set can be seen in Fig. 1.4 (symbols). Each point corresponds to discrete values $V_r(t_k)$ of *HD 82943*, a system known to contain two planets in a 2/1 mean-motion resonance (see Mayor et al. [5], Ferraz-Mello et al. [6]). Doppler data is usually presented in a multi-column format giving, among other information, the times of observation t_k (usually in Julian days), radial velocities $V_{rk} = V_r(t_k)$ (usually in meters per second) and the expected uncertainties ϵ_k (also in $\mathrm{m\,s^{-1}}$). These later values correspond to the size of the error bar of the radial velocity data, and a Gaussian error distribution is usually assumed. Current instrumentation and reduction techniques have lowered the values of ϵ_k to the order of a few $\mathrm{m\,s^{-1}}$. When observations include data from more than one instrument, the origin of each data segment is also included in the files.

With the numerical data in hand, we first assume that the temporal variations of V_r are caused by the presence of one (or more) exoplanets, and therefore correspond to time-discrete values of a function of type (1.8). That being the case, our second task is to develop a numerical algorithm to deduce the number of periodic terms contained in the signal (i.e., number of planets N), and for each to estimate the values of the set

$$(\mathcal{K}_i, n_i, e_i, \omega_i, \tau_i) \qquad (i = 1, \ldots, N) \tag{1.14}$$

plus the barycentric radial velocity V_{r0}. These are sometimes referred to as the “primary parameters” of an orbital fit. The individual planetary masses (multiplied by $\sin I_i$) are derived from the calculated value of $\mathcal{K}_i$ and the mass function. Notice that the number of free parameters is equal to $5N + 1$, consisting of five orbital parameters per planet plus the radial velocity V_{r0} of the

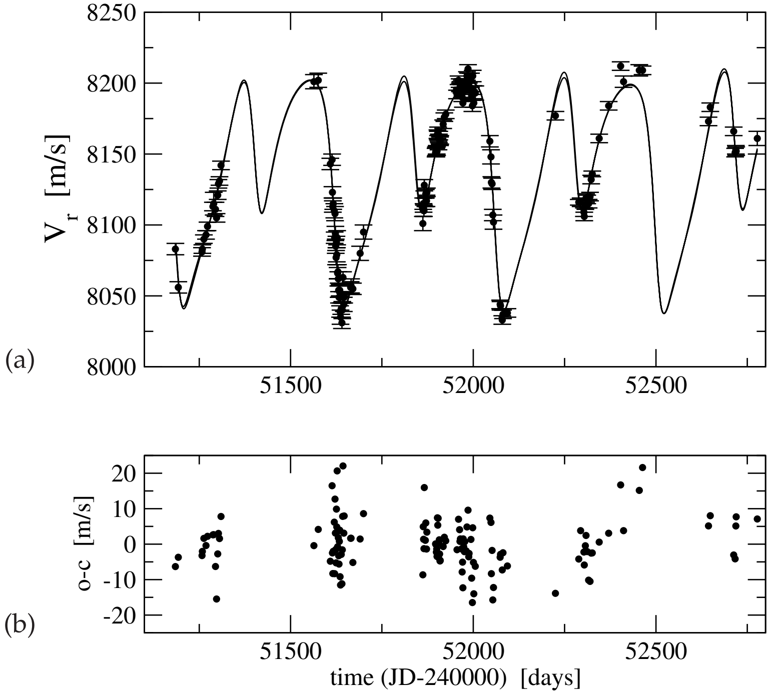


Fig. 1.4 (a) Radial velocity data points of the *HD82943* star, together with an orbital fit with two planets. (b) Residuals from the fit. Figure obtained with a dynamical two-planet fit.

barycenter of the extrasolar system. As we shall show later on, V_{r0} does not necessarily correspond to the time-averaged value of V_r, unless all exoplanets move in circular orbits. In the case where the data includes values from different instruments and observatories, individual values of V_{r0} are usually assigned. As a final note of caution, in the case of more than one planet, the calculated orbital periods (or mean motions) are not osculating, but apparent (see [6]).

The mass of the star is taken from sophisticated stellar models. However, one must keep in mind that, even for Hipparcos stars having the best available spectroscopy and astrometry, the more accurate models do not allow to know the masses better than $\simeq$ 8 percent (Allende Prieto et al. [7]). This fact supersedes some discussions on the nature of the published planetary elements, if astrocentric or barycentric. The difference between coordinates in these systems is usually much smaller than the uncertainty in our knowledge of the stellar mass.

Even though the functional form of $V_r(t)$ given by (1.8) is the sum of periodic terms, it is not usually convenient to attempt an orbital fit using a direct

Fourier analysis. The reasons are twofold. First, a precise identification of the leading frequencies with Fourier decomposition requires that the observational data should cover several periods. This is not usually the case, specially for planets with large semi-major axis. Second, radial velocity data is not evenly spaced in the time axis and, even worse, usually contains months-long gaps where observations are not favorable. Both problems can be overcome using a more general Fourier method, such as the Date-Compensated Fourier Transform [8] or the CLEANest algorithm [9], which were specifically developed for nonequidistant data points and arbitrary frequencies. However, a least-squares algorithm is usually more precise and requires less fine-tuning of the results. Thus, practically all orbital fits have been calculated using this approach.

We then search for adequate coefficients (1.14) of a fitting function $y(t)$, of type (1.8), such that the residual function

$$Q^2 = \sum_{t_n} \frac{[y(t_n) - V_r(t_n)]^2}{\epsilon_n^2} \tag{1.15}$$

is minimum. Notice that this definition includes the uncertainty of each data point $V_r(t_n)$ and has the advantage of considering different precisions among the data. This is particularly important when mixing observations from different instruments. In the case where the ϵ_n correspond to the standard deviation of the data, Q^2 is related to the χ^2 of the data modelization.

Initially, deterministic versions of nonlinear least-squares were used, such as hill-climbing techniques or the Levenberg–Marquardt method (see [5, 10, 11]). The main drawback with these methods is that they are unable to distinguish between local and global minima of Q^2; consequently, there is no guarantee that the calculated orbital elements correspond to the best fit of the data sets. Since the number of free parameters can be large, the shape of the residual function may be complex and contain numerous local minima, several of them possibly with similar values. Moreover, since the problem is highly nonlinear, the result may be highly sensitive to the initial values of the parameters. An example of this behavior was given by Mayor et al. [5] for the two *HD 82943* planets. The authors presented two different fits: in the first the orbital eccentricities of the planets were $(e_1, e_2) = (0.4, 0.0)$ and for the second $(e_1, e_2) = (0.4, 0.18)$. Although the eccentricity of the outer planet changed significantly, the value of Q^2 only varied by $\simeq$ 0.1 percent. What is more worrisome in this case is that the best-fit solutions found by several authors actually corresponds to orbits which are dynamically unstable in timescales of the order of 10^5–10^6 years (see [6]). Thus, given a limited set of observations, the orbital configuration of the real planets does not necessarily correspond to the best fit.

Since results of orbital fits sometimes seem very sensitive to the numerical method and/or data set, we need to fine-tune our techniques. We need a strategy (or method) that can identify the global extrema of the residual function. Additionally, we must be able to estimate the confidence levels (i.e., errors) in the orbital parameters themselves. Due to the highly nonlinear characteristics of the equations, it is not correct to assume Gaussian distribution errors in $(V_{0r}, \mathcal{K}_i, n_i, e_i, \omega_i, \tau_i)$. Consequently, the standard deviations that are sometimes seen, alongside the best fits, can be misleading and must be considered with utmost care [12].

Considering that the result of classical nonlinear best-fit methods depends on the initial guess, a possible approach towards a global minimum is to apply the same method to a large number of initial conditions, distributed randomly in the parameter space. This so-called Monte Carlo approach was used by Brown [13] to the V_r data from *HD 72659*. A year later, Ferraz-Mello et al. [6] employed a similar approach to study the two-planet system of *HD 82943*. One of the main advantages of this type of Monte Carlo algorithm is the possibility of estimating the confidence region of each of the orbital elements; in other words, the different possible primary parameters that are all compatible with the given data set. For the particular case of *HD 82943*, we found a large set of different orbital fits which yield practically the same value of the residual function (see Fig. 1.5). Thus, in some cases, it is not possible to give a single value of the parameter set as the "correct" orbital fit.

A different strategy for the search of global minima of the orbital fit, is the use of genetic algorithms. This technique is based on natural selection (mimicking the behavior of biological populations), by which an initially random population of initial guesses evolves towards the global minimum. Although this approach can require larger computational resources than deterministic methods, it has proved to be extremely robust in all applications to exoplanetary systems (e.g. [14, 15]). Other advantages of this approach include its simple manipulation, and its ability to introduce non-Gaussian error estimations with no significant complications. A recommended introductory text on genetic algorithms can be found in Charbonneau [16].

1.3.2
Secondary Parameters

Whatever the chosen numerical approach, the orbital fit yields values for V_{0r},$\mathcal{K}_i$, n_i, e_i, ω_i and τ_i. From these we must now estimate the planetary masses and semi-major axes. These quantities are related to the primary parameters through the Eqs. (1.3) and (1.11). Notice that we have two algebraic equations with three unknowns, and it is impossible to separate the value of $\sin I_i$ from the planetary mass. Thus, the values of m and a must be de-

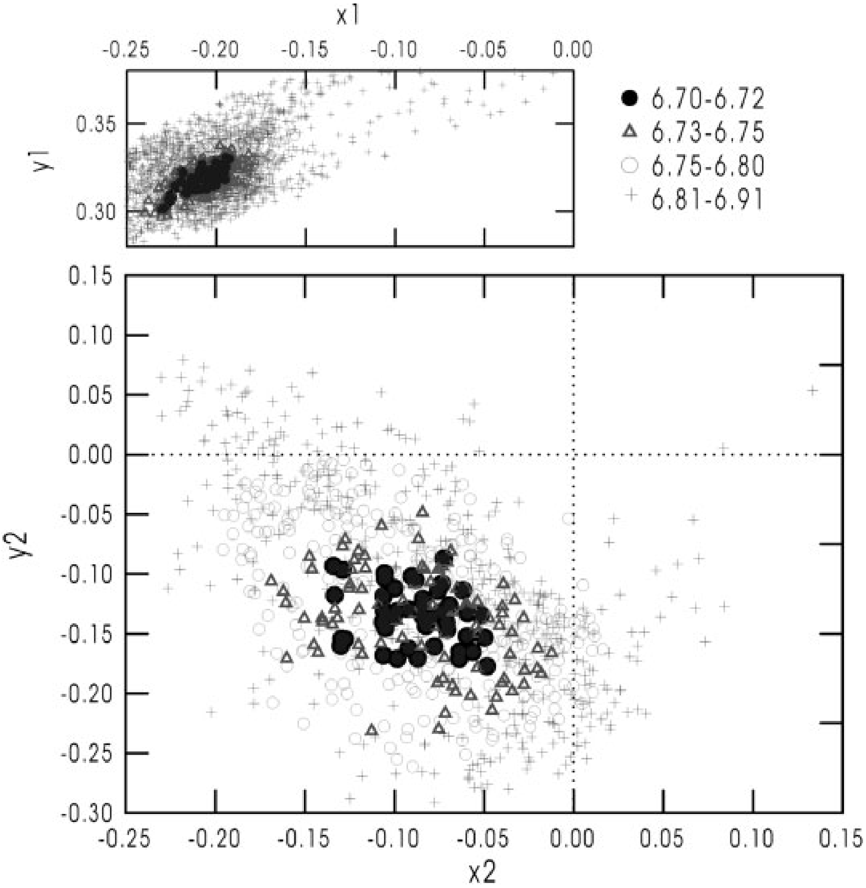


Fig. 1.5 Projections of possible orbital fits (in astrocentric orbital elements) for the two *HD 82943* planets, on the planes $(x_i, y_i) = (e_i \cos\omega_i, e_i \sin\omega_i)$. Different symbols correspond to different values of the of r.m.s. of the residuals (from Ferraz-Mello et al. [6]).

termined assuming some ad hoc value for the orbital inclination. Usually, a value of 90 degrees is chosen, which corresponds to an edge-on orbital fit and minimum planetary masses. The sensitivity of both m and a to different values of I is shown in Fig. 1.6. In this plot we have used the following primary parameters: $\mathcal{F} = 10^{-3}$, $M_* = 1$ and $n = 2\pi/365$ days^{-1}. Notice that both the mass and semi-major axes increase as smaller inclinations are assumed, although the mass is the most sensitive parameter. The change in a is not very important except for small values of I.

A possible determination of the real individual planetary masses occurs when both Doppler data and stellar transits are simultaneously available for

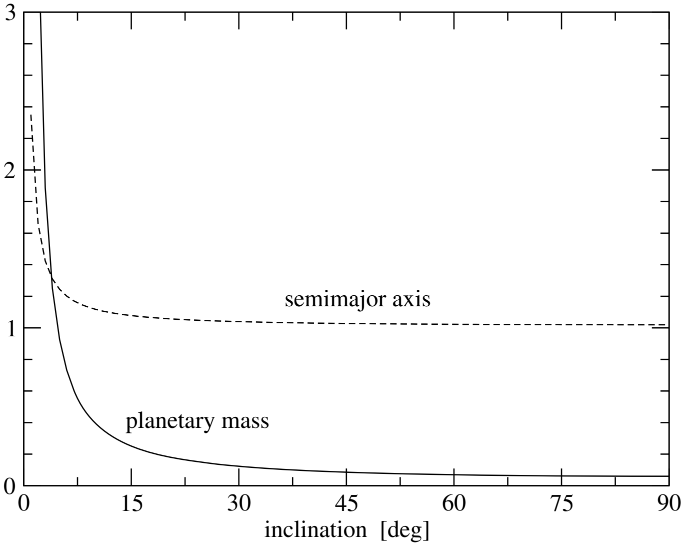


Fig. 1.6 Variation of the planetary mass m (in units of stellar mass) and semi-major axis a (in AU), as a function of the unknown inclination of the orbital plane I, for fixed values of the primary parameters of an orbital fit. The plot was constructed with $\mathcal{F} = 10^{-3}$, $M_* = 1$ and $n = 2\pi/365$ days^{-1}.

an exoplanetary system. In this case, the inclination I is known, and m can be uncoupled from this angle. So far, only a handfull of exoplanets have been observed by both techniques (e.g. HD 209458 b) and, for most of the rest, the masses and semi-major axis are still affected by I_i. At first hand, this seems a major limitation for any dynamical analysis, since these are probably the most important parameters. However, if for multiple planetary systems we assume that all planets are co-planar, then the ratios:

$$\frac{m_j}{m_i} \qquad \frac{a_j}{a_i} \qquad\qquad (i, j = 1, \ldots, N) \tag{1.16}$$

are unaffected by the value of the spatial inclination. In other words, although the individual values of m_i and a_i may be unknown, the relative values can be deduced, and used in our dynamical studies.

1.3.3
N-Body Fits

In the previous analysis, we have assumed that the motion of each planet orbiting a given star can be modeled by a Keplerian ellipse. This is an approximation since mutual perturbations will cause the orbital elements to change with time. If the estimated values for the planetary masses (minimum values) are sufficiently small or the mutual separation (i.e., a_i/a_j) are sufficiently large, we can assume that the orbital variations are negligible within the timespan of the observations. In that case, the multi-Keplerian fits presented before are valid approximations to the problem. However, if the mutual perturbations are large, we must modify the orbital fit to accommodate nonconstant orbital elements. This is usually referred to dynamical (or N-Body) orbital fits.

We assume a data file consisting of several observations, starting at time $t = t_0$ and ending at $t = t_M$. In this interval, the orbital elements are allowed to vary with time. A dynamical fit proceeds the same way as the multi-Keplerian version, except for the calculation of the model values of $V_r(t_i)$. For perturbed orbits, it is no longer optimal to use (1.8) to relate the radial velocity with the orbital elements. The procedure can be separated into the following steps:

1. Specify initial conditions $(n_{i0}, e_{i0}, \omega_{i0}, \tau_{i0}, I_{i0})$ for all the planets, which will correspond to the astrocentric orbits at the beginning of the observations. The orbital period of each planet must be osculating, and not apparent [6]. We will also require values for the real planetary masses, unaffected by the inclinations I_i, and the distance between the ascending nodes.

2. Transform the orbital elements to Cartesian coordinates and velocities. We will need Kepler's third law to obtain the semi-major axes, and thus the result will depend on the stellar mass m_0. From this data we can calculate the velocity vector of the star $\mathbf{V}$ (with respect to the barycenter of the system) at t_0. Choosing the reference frame of the coordinate system tangent to the celestial sphere, the first model value of $V_r(t_0)$ will be given by the z-component of $\mathbf{V}$.

3. Using an N-body numerical integrator, calculate the positions of the planets at all the subsequent times of observation (i.e., $t = t_1, \ldots, t_M$).

4. For each t_i, repeat the calculations in Step 2, and obtain the complete set of radial velocities $V_r(t_i)$.

Having all the values of $V_r(t_i)$ for the chosen initial conditions and real planetary masses, we can calculate the residual function. The best fit will then be the set of initial parameters and planetary masses that minimizes Q^2. The use of numerical integrations will obviously increase the amount of CPU time; thus N-body fits are sometimes done as a second-order approximation from initial Keplerian parameters.

A factor to be taken into account is the uncertainty in the value of the stellar mass, and its propagation to other quantities in the orbital fit. To obtain a_i we must use Kepler's third law, and the results depend explicitly on the choice of m_0. A possible way to avoid these problems is to use an adimensional formulation for the equations of motion (see [2]). Even if this approach is not employed, it can easily be seen that the ratios m_i/m_j and a_i/a_j between any two planets are unaffected by the particular value of m_0. The same characteristic was also seen in the dependence of the orbital fits with the inclinations I. Thus, any dynamical study that can be constructed as a function of these ratios will yield results virtually independent of I and m_0.

One of the most important traits of dynamical fits is its theoretical ability to assess the planetary masses independently of other detection methods, thus allowing us to bypass the limitations of radial velocity data. However, this task is not always possible. Even for fixed edge-on systems, the difference between a dynamical and a multi-Keplerian fit is appreciable only if: (i) the planets are under significant mutual perturbations and, (ii) the observational timespan is large. This is true only in a very few cases, the most well known example being *GJ 876* [10]. For almost all other known planetary systems, the distinction is practically unnoticeable.

An example is given in Fig. 1.7 for the *HD 82943* planets. We can see little difference between both fits (Keplerian and dynamical) within the observation interval and, at least for this system, both models yield similar results. However, the divergence between results will increase with time, and longer observational timespans should be able to detect the effects of mutual perturbations between the planets.

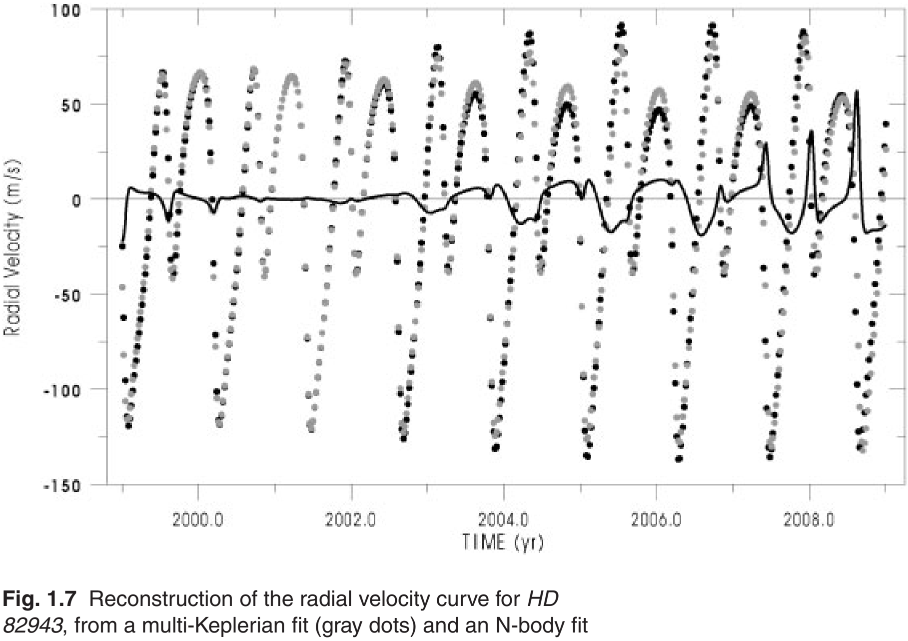


Fig. 1.7 Reconstruction of the radial velocity curve for *HD 82943*, from a multi-Keplerian fit (gray dots) and an N-body fit (black dots). The solid line shows the difference (Keplerian minus dynamical). For more details, see Ferraz-Mello et al. [6].

Since dynamical fits are not necessary for most planetary systems, in practice we are still not able to decouple the planetary masses from the inclinations. The values of I_i are not usually considered variables, but fixed at some initial value, and co-planarity between the planets is assumed. Thus, the true potential of N-body fits is still far from being fulfilled. Once again, however, larger observational timespans will certainly change this picture.

1.4
Coordinate Systems and Equations of Motion

Even with all the limitations and uncertainties, stemming both from the observations and reduction techniques, orbital fits yield (minimum) masses and orbital elements of the planets in a given stellar system. As a first step towards a dynamical study, we must construct their equations of motion.

Suppose a system consisted of a star of mass m_0 and N planets of mass m_i, thus making this a $(N+1)$-body problem. Let $\boldsymbol{X}_i$ denote the position vectors of all bodies with respect to an inertial reference frame centered in the barycenter of the system. Then, from Newton's law of gravitation, we have:

$$\ddot{\boldsymbol{X}}_i = -\mathcal{G} \sum_{\substack{j=0 \\ j \neq i}}^{M} m_j \frac{\boldsymbol{X}_i - \boldsymbol{X}_j}{|\boldsymbol{X}_i - \boldsymbol{X}_j|^3} \tag{1.17}$$

where the double dot denotes the second derivative with respect to the time. Introducing the astrocentric positions of the planets as $\boldsymbol{r}_i = \boldsymbol{X}_i - \boldsymbol{X}_0$, we can write the equations of motion for the planets in astrocentric variables as:

$$\ddot{\boldsymbol{r}}_i = -\mathcal{G} \frac{(m_0 + m_i)}{|\boldsymbol{r}_i|^3} \boldsymbol{r}_i + \mathcal{G} \sum_{\substack{j=1 \\ j \neq i}}^{M} m_j \left(\frac{\boldsymbol{r}_j - \boldsymbol{r}_i}{|\boldsymbol{r}_j - \boldsymbol{r}_i|^3} - \frac{\boldsymbol{r}_j}{|\boldsymbol{r}_j|^3} \right) \tag{1.18}$$

In terms of these coordinates, the barycentric motion of the star is given by:

$$\boldsymbol{X}_0 = -\frac{\sum_{i_1}^{N} m_i \boldsymbol{r}_i}{\sum_{i_1}^{N} m_i} \tag{1.19}$$

The second term inside the brackets is due to the noninertiality of the astrocentric reference frame, and is caused by the perturbations of the planets on the motion of the star.

1.4.1
Barycentric Hamiltonian Equations

Since dynamical studies of extrasolar planets benefit from the Hamiltonian structure of the equations of motion, we devote the rest of this section to presenting three different forms of canonical variables and Hamiltonian functions. Although this is a well established problem in celestial mechanics, the vast majority of papers deal with the so-called restricted problem in which only two bodies have finite masses.

The barycentric Hamiltonian equations of the $(N+1)$-body problem are easy to obtain from Eq. (1.17). Defining $\mathbf{\Pi}_i = m_i \dot{\boldsymbol{X}}_i$ as the linear momenta associated to each position vector $\boldsymbol{X}_i$, these variables are canonical, and the Hamiltonian of the system is the sum of their kinetic and potential energies:

$$\tilde{H} = \frac{1}{2}\sum_{k=0}^{N} \frac{\Pi_k^2}{m_k} - \mathcal{G}\sum_{k=0}^{N}\sum_{j=k+1}^{N} \frac{m_k m_j}{\Delta_{kj}} \tag{1.20}$$

where $\Delta_{kj} = |\boldsymbol{X}_k - \boldsymbol{X}_j|$. This system has, however, $3(N+1)$ degrees of freedom, that is, six equations more than the usual Laplace–Lagrange formulation of the heliocentric equations of motion.

The system can be reduced to $3N$ degrees of freedom through the convenient use of the trivial conservation laws concerning the inertial motion of the barycenter. There are two sets of variables used to reduce to $3N$ the number of degrees of freedom of the above system. Each will be discussed in the following subsections.

1.4.2
Jacobi Hamiltonian Formalism

The most popular reduction, due to Jacobi, is widely used in the study of the general three-body problem and of planetary and stellar systems. In Jacobi's formulation, the position and velocity of the planet m_1 are given in a reference frame with origin in m_0 (equal to the star); the position and velocity of m_2 are given in a reference frame with the origin at the barycenter of m_0 and m_1; the position and velocity of m_3 are given in a reference frame with the origin at the barycenter of m_0, m_1 and m_2, and so on. If we denote with $\boldsymbol{\rho}_k$ $(k = 1, \ldots, N)$ the vectors thus defined, we have

$$\boldsymbol{\rho}_k = \boldsymbol{X}_k - \frac{1}{\sigma_k}\sum_{j=0}^{k} m_j \boldsymbol{X}_j \qquad (k = 1, \ldots, N) \tag{1.21}$$

where

$$\sigma_k = \sum_{j=0}^{k} m_j \tag{1.22}$$

The quantities $\boldsymbol{\rho}_k$ are our new coordinates; we must now search for their canonical momenta $\boldsymbol{\pi}_k$. These can be obtained from the original $\boldsymbol{p}_k$ by means of the simple canonical condition $\sum_{i=1}^{N}(\boldsymbol{\pi}_i d\boldsymbol{\rho}_i - \boldsymbol{\Pi}_i d\mathbf{X}_i) = 0$, and give the implicit relation:

$$\boldsymbol{\Pi}_k = \pi_k - \sum_{j=k+1}^{N} \frac{m_k \pi_j}{\sigma_{j-1}} \tag{1.23}$$

The reader is referred to Ferraz-Mello et al. [2] for details of this construction. A lengthy, but simple calculation shows that nonconstant terms of the total kinetic energy, in these variables, are given by

$$T = \sum_{i=1}^{N} \frac{\pi_i^2}{2\tilde{\beta}_i} \tag{1.24}$$

where $\tilde{\beta}_i$ are the so-called reduced masses of the Jacobian formulation, defined by

$$\tilde{\beta}_i = \frac{m_i \sigma_{i-1}}{\sigma_i} \tag{1.25}$$

The complete Hamiltonian of the relative motion of the N planets, can be written as:

$$H = H_0 + H_1 \tag{1.26}$$

where:

$$\begin{aligned} H_0 &= \sum_{k=1}^{N} \left(\frac{\pi_k^2}{2\tilde{\beta}_k} - \frac{\mathcal{G}\sigma_k \tilde{\beta}_k}{\rho_k} \right) \\ H_1 &= -\mathcal{G} \sum_{k=1}^{N} \sum_{j=k+1}^{N} \frac{m_k m_j}{\Delta_{kj}} - \mathcal{G} \sum_{k=1}^{N} m_k \left(\frac{m_0}{\Delta_{0k}} - \frac{\sigma_{k-1}}{\rho_k} \right) \end{aligned} \tag{1.27}$$

Constant terms were discarded, since they do not contribute to the equations. This function defines a system with $3N$ degrees of freedom in the canonical variables (ρ_k, π_k) with $(k = 1, \ldots, N)$.

Notice that H_0 may be written as the sum of N terms of the form

$$F_k = \frac{\pi_k^2}{\tilde{\beta}_k} - \frac{\mathcal{G}\sigma_k \tilde{\beta}_k}{\rho_k} \tag{1.28}$$

each of which represents the Hamiltonian for the unperturbed motion of m_k around the center of gravity of the first $(k-1)$ mass-points. It is easy to see that it has the same functional form as the two-body Hamiltonian in astrocentric coordinates, except for a change of definition in the masses. Thus, the solution of the unperturbed system with $H_1 = 0$ are also conics, and we can use (1.28) to define new *Jacobian* orbital elements. These will differ from their astrocentric counterparts in the first order of the planetary masses.

1.4.3
Poincaré Hamiltonian Formalism

A different reduction to $3N$ degrees of freedom is due to Poincaré [17]. The resulting equations were not often used in studies of the solar system, perhaps because Poincaré himself mentioned he believed its difficulties outweighted its advantages [18]. However, in recent years this approach has been applied successfully to several problems in planetary dynamics [19–22]. In fact, and as we will show below, Poincaré's formalism is not so complex at all and, when compared to Jacobi's approach, the expressions are significantly simpler, and even easier to use.

The definition of the new canonical variables $(\boldsymbol{r}_i, \boldsymbol{p}_i)$ for the N planets are very simple. The new coordinates $\boldsymbol{r}_i$ are simply equal to the astrocentric position vectors $\boldsymbol{X}_i - \boldsymbol{X}_0$, and the new momenta $\boldsymbol{p}_i$ are the same linear momenta $\boldsymbol{\Pi}_i$ of the barycentric formulation. Hence,

$$\boldsymbol{r}_i = \boldsymbol{X}_i - \boldsymbol{X}_0 \qquad \boldsymbol{p}_i = \boldsymbol{\Pi}_i \qquad (i = 1, 2, \ldots, N) \tag{1.29}$$

It is noteworthy that this definition mixes coordinate systems, the positions being astrocentric while the momenta are barycentric.

We refer the reader to [2] for more details on the construction of these variables, as well as the algebraic manipulations to obtain the Hamiltonian function. The works of Laskar [19] and Laskar and Robutel [20] are also highly recommended references.

The Hamiltonian of the reduced system can once again be written as $H = H_0 + H_1$, where

$$H_0 = \sum_{i=k}^{N} \left(\frac{1}{2} \frac{p_k^2}{\beta_k} - \frac{\mu_k \beta_k}{r_k} \right) \tag{1.30}$$

$$H_1 = \sum_{k=1}^{N} \sum_{j=k+1}^{N} \left(-\frac{\mathcal{G} m_k m_j}{\Delta_{kj}} + \frac{\boldsymbol{p}_k \cdot \boldsymbol{p}_j}{m_0} \right)$$

and

$$\mu_k = \mathcal{G}(m_0 + m_k) \qquad \beta_k = \frac{m_0 m_k}{m_0 + m_k} \tag{1.31}$$

We note that H_0 is of the order of the planetary masses m_k while H_1 is of order two with respect to these masses. Then H_0 may be seen as the new expression for the undisturbed energy while H_1 is the potential energy of the interaction between the planets. It is worth noting that each term

$$F_k = \frac{1}{2}\frac{p_k^2}{\beta_k} - \frac{\mu_k \beta_k}{r_k} \tag{1.32}$$

is the Hamiltonian of a two-body problem in which the mass point m_k is moving around the mass point m_0.

The expression for the perturbation term H_1 is worth a couple of comments. On one hand, it is more compact than its counterpart in Jacobi coordinates (Eq. (1.27)). In fact, it is very similar to the expression of the disturbing function in the astrocentric reference frame. If we add a greater simplicity of the definition of the canonical variables, Poincaré's approach begins to appear more appealing than Jacobi's. On the other hand, Poincaré's expression for H_1 includes terms that depend on the momenta, and this characteristic is baffling to a first-time user. Most of us are used to working with potentials that are only a function of the positions, and mixed variables have the feel of nonconservative systems. The explanation, however, simply lies in the different reference frames chosen for the coordinates and momenta.

1.4.4
Generalized Orbital Elements and Delaunay Variables

The reduced Hamiltonians (1.27) and (1.30) were written in Cartesian coordinates. The purpose of this subsection is to obtain "general" orbital elements and Delaunay variables corresponding to both Jacobi and Poincaré formalisms.

Orbital elements (or their Delaunay canonical counterparts) of each planetary mass m_i are defined as solutions of each F_i making up the unperturbed Hamiltonian H_0. The expression for F_i in each coordinate system is:

$$\begin{aligned}
\text{Astrocentric:} \quad & F_k = \frac{1}{2}\frac{(m_k \dot{\boldsymbol{r}}_k)^2}{m_k} - \frac{\mathcal{G}(m_0 + m_k)m_k}{r_k} \\
\text{Jacobi:} \quad & F_k = \frac{1}{2}\frac{\pi_k^2}{\tilde{\beta}_k} - \frac{\mathcal{G}\sigma_k \tilde{\beta}_k}{\rho_k} \\
\text{Poincaré:} \quad & F_k = \frac{1}{2}\frac{p_k^2}{\beta_k} - \frac{\mu_k \beta_k}{r_k}
\end{aligned} \tag{1.33}$$

Recall, however, that astrocentric coordinates $(\boldsymbol{r}_k, m_k \dot{\boldsymbol{r}}_k)$ are only canonical if $N = 2$, while the Jacobi and Poincaré version are canonical for any number of bodies.

Notice that all three expressions in (1.33) have the same functional form with respect to the coordinates; only the mass parameters are different. This means that the solution in each coordinate system will also have the same form, and their integrals of motion (e.g., orbital elements) can be obtained with the same formulas. In particular, we can write a general expression for F_k in the form:

$$F_k = \frac{1}{2} m v^2 - \frac{\mu}{|\boldsymbol{r}|} \tag{1.34}$$

where the meaning of the set $(\boldsymbol{r}, \boldsymbol{v}, m, \mu)$ in each reference frame is summarized in Table 1.1.

Table 1.1 Correspondence between coordinates and mass parameters defining the unperturbed Hamiltonian H_0 in three different reference frames.

Coordinate system	**Position** (x)	**Velocity** (v)	**Mass** (m)	μ
Astrocentric	$\boldsymbol{r}_k$	$\dot{\boldsymbol{r}}_k$	m_k	$\mathcal{G}(m_0 + m_k)$
Jacobi	$\boldsymbol{\rho}_k$	$\boldsymbol{\pi}_k / \tilde{\beta}_k$	$\tilde{\beta}_k$	$\mathcal{G}\sigma_k$
Poincaré	$\boldsymbol{r}_k$	$\boldsymbol{p}_k / \beta_k$	β_k	μ_k

We can now use the usual two-body formulas to define generalized orbital elements in each reference frame. These expressions can be found in any textbook on celestial mechanics (e.g., [3, 4, 23]). For the semi-major axis and eccentricity, we have:

$$\begin{aligned} a &\overset{\text{def}}{=} \frac{\mu r}{2\mu - r v^2} \\ e &\overset{\text{def}}{=} \sqrt{\left(1 - \frac{r}{a}\right)^2 + \frac{(\boldsymbol{r} \cdot \boldsymbol{v})^2}{\mu a}} \end{aligned} \tag{1.35}$$

The remaining elements also follow the same usual definitions. Kepler's Third Law also reads:

$$n^2 a^3 = \mu \tag{1.36}$$

where, once again, the different definitions of μ yield different relations between the semi-major axis and orbital frequency. Last of all, we need to modify the orbital elements $(a, e, I, l, \omega, \Omega)$ to a canonical set. The usual choice is the so-called mass-weighted Delaunay variables $(L, G, T, l, \omega, \Omega)$, where the new momenta are defined by:

$$\begin{aligned} L &= m\sqrt{\mu a} \\ G &= L\sqrt{1 - e^2} \\ T &= G \cos I \end{aligned} \tag{1.37}$$

It is interesting to note that Eqs. (1.35) and (1.36) are valid for all our reference systems; the only difference lies in the definitions found in Table 1.1 for each individual case.

1.4.5
Comparisons Between Coordinate Systems

We have seen that, at least formally, the Poincaré formalism is simpler and more compact than the Jacobi variables. But how does each perform in practice? We have simulated the short-term orbital evolution of a co-planar system formed by a central star with $m_0 = 0.32 M_\oplus$ and two planets with masses $m_1 = 20 M_{\text{Jup}}$ and $m_2 = 5 M_{\text{Jup}}$. Initial conditions were chosen such that both planets have circular orbits with semi-major axes $a_1 = 0.131$ AU and $a_2 = 0.232$ AU and are in opposition. The large planetary masses were chosen in order to have strong perturbations and to avoid misleading graphics hiding the actual behavior of the orbital parameters.

Figure 1.8 shows the variations of the semi-major axis (a) and eccentricity (b) of the outer planet only. The orbital elements were calculated in each of the three reference frames (astrocentric, Jacobi and Poincaré). The inner planet shows very little difference, and is not shown. The results noted in Fig. 1.8 should be taken with care. The Jacobi variables are those showing the less variable elements in this example, but this is due to the fact that the planet in the innermost orbit is much larger. Thus, the results appearing here can differ from system to system and depend on the arbitrary order in which the planets are chosen in the construction of Jacobian coordinates. When a natural choice is possible, as in the given example, Jacobi elements are those showing the least variations.

The main advantage of having a small temporal variation, at least on short timescales, can be found in the use of orbital fits as initial conditions for simulations. If the two-body values of a and e vary little, then a multi-Keplerian orbital determination from radial velocity data will be more precise than a case where the same parameters show significant variations throughout the observational interval. For these reasons, in recent years Jacobi coordinates have been a popular choice for orbital determination. The only problem lies in the hierarchical structure of Jacobi coordinates. In order to define the variables, we must first know which is the first planet, which is the second, etc. This prior knowledge is not necessary in Poincaré or astrocentric variables, and can lead to confusion or erroneous results if not done with care.

Orbital fits in Jacobi coordinates can be undertaken in the same way as deduced for astrocentric elements, except for a change in the definition of the semi-amplitudes K_i. Thus, the complete radial velocity of the star m_0 is still given by (1.8), but now:

$$\mathcal{K}_i = \frac{m_i \sin I_i}{\sigma_i} \frac{2\pi a_i}{\hat{T}_i \sqrt{1-e_i^2}} \qquad \text{with} \qquad \hat{T}_i = \frac{2\pi a_i^{3/2}}{\sqrt{\mathcal{G}\sigma_i}} \tag{1.38}$$

and where $\sigma_i = \sum_{k=0}^{i} m_i$, and all orbital elements are Jacobian. The reader is referred to Lee and Peale [24] for further details.

In conclusion, Jacobi seems a good choice for orbital representation, especially if N-body fits are not employed. However, the larger sensitivity of the astrocentric coordinates to mutual perturbations has its advantages. If dynamical fits are used in the hope of uncoupling the planetary masses and the inclinations, astrocentric coordinates are preferable, since the orbital varia-

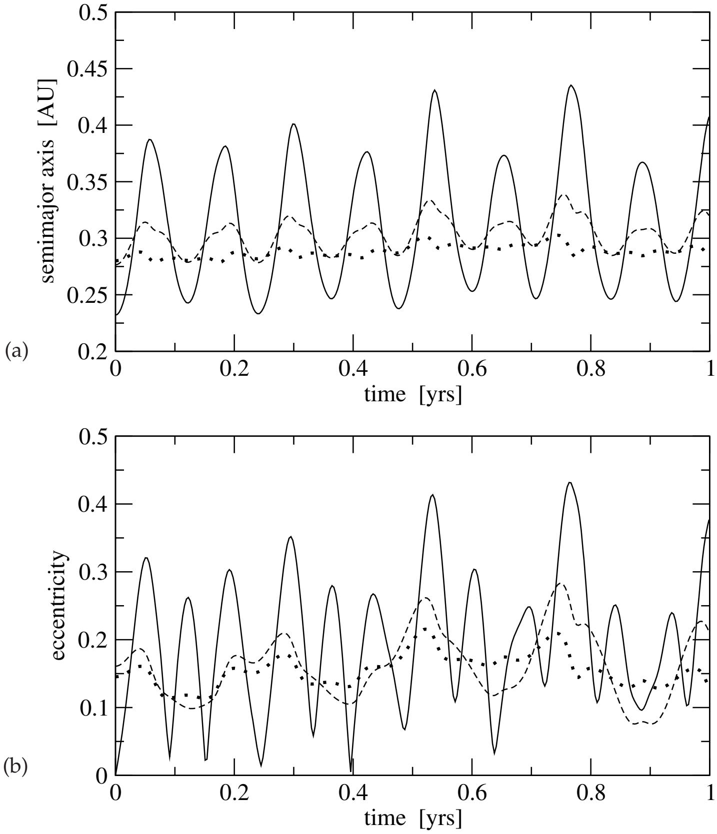


Fig. 1.8 Evolution of semi-major axis and eccentricity of the outer body (two-planet system) in three different coordinate systems. Continuous lines correspond to astrocentric orbital elements, dotted lines to Jacobi variables, and dashed lines to orbital elements deduced from Poincaré canonical variables.

tions will first become appreciable in this reference frame. Thus, the choice between Jacobi and astrocentric for the process of orbital determination depends on the problem at hand, and on the information desired by the researcher. Whatever the choice, the Poincaré canonical variables still stands out as the most adequate reference frame for dynamical studies. The transformation between all systems is straightforward, and there should not be any inhibitions in using different coordinates for different tasks.

1.4.6
The Conservation of the Angular Momentum

If the only forces acting on the $N+1$ bodies are their point-mass gravitational attractions, the angular momentum is conserved:

$$\mathcal{L} = \sum_{i=0}^{N} m_i \boldsymbol{X}_i \times \dot{\boldsymbol{X}}_i \tag{1.39}$$

Since $\sum_{i=0}^{N} m_i \boldsymbol{X}_i = \sum_{i=0}^{N} m_i \dot{\boldsymbol{X}}_i = 0$, the above equation gives

$$\mathcal{L} = \sum_{i=0}^{N} m_i \boldsymbol{r}_i \times \boldsymbol{p}_i \tag{1.40}$$

which, in terms of the orbital elements, yields

$$\mathcal{L} = \sum_{i=1}^{N} \beta_i \sqrt{\mu_i a_i (1 - e_i^2)}\,\hat{\boldsymbol{k}}_i \tag{1.41}$$

where $\boldsymbol{k}_i$ are the unit vectors normal to the orbital planes. This is an exact conservation law. In this equation a_i and e_i are not the usual astrocentric osculating elements but the canonical Poincaré elements.

The conservation law given by (1.40) is also true if Jacobian coordinates are used. It is worth emphasizing that when a_i and e_i are the astrocentric osculating elements, the expression

$$\hat{\mathcal{L}} = \sum_{i=1}^{N} m_i \sqrt{\mu_i a_i (1 - e_i^2)}\,\hat{\boldsymbol{k}}_i \tag{1.42}$$

is no longer an exact conservation law. One may easily see that:

$$\hat{\mathcal{L}} = \mathcal{L} - \sum_{i=1}^{N} m_i \boldsymbol{X}_0 \times \dot{\boldsymbol{X}}_0 \tag{1.43}$$

showing that the quantity $\hat{\mathcal{L}}$ has in fact a variation of order $\mathcal{O}(m_i^2)$. Thus, the conservation of the total angular momentum is better expressed in canonical variables than in astrocentric orbital elements.

References

1 Perryman, M.A.C.: 2000, Extra-solar planets. *Rep. Prog. Phys.*, **63**, 1209–1272.

2 Ferraz-Mello, S., Michtchenko, T.A., Beaugé, C., and Callegari Jr., N.: 2005b, Extrasolar Planetary Systems. In *Chaos and Stability in Planetary Systems*, (R. Dvorak, F. Freistetter and R. Kurths, Eds.), Lect. Notes Phys., **683**, 219–271.

3 Brouwer, D. and Clemence, G.M.: 1961, *Methods of Celestial Mechanics*, Academic Press, NY.

4 Murray, C.D. and Dermott, S.F.: 1999, *Solar System Dynamics*, Cambridge University Press.

5 Mayor, M., Udry, S. Naef, D., Pepe, F., Queloz, D., Santos, N.C., and Burnet, M.: 2004, The CORALIE survey for southern extrasolar planets. XII. Orbital solutions for 16 extrasolar planets discovered with CORALIE. *A&A*, **415**, 391–402.

6 Ferraz-Mello, S., Michtchenko, T.A., and Beaugé, C.: 2005a, The orbits of the extrasolar planets HD 82943c and b. *ApJ*, **621**, 473.

7 Allende Prieto, C. and Lambert, D.L.: 1999, Fundamental parameters of nearby stars from the comparison with evolutionary calculations: masses, radii and effective temperatures. *A&A*, **352**, 555–562.

8 Ferraz-Mello, S.: 1981, Estimation of periods from unequally spaced observations. *AJ*, **86**, 619–624.

9 Foster, G.: 1995, The cleanest Fourier spectrum. *AJ*, **109**, 1889–1902.

10 Laughlin, G. and Chambers, J.E.:2001, Short-term dynamical interactions among extrasolar planets. *ApJ*, **551**, L109–L113.

11 Rivera, E. and Haghighipour, N.: 2007, On the stability of test particles in extrasolar multiple planet systems. *MNRAS*, **374**, 599–613.

12 Ford, E.B.: 2005, Quantifying the uncertainty in the orbits of extrasolar planets. *AJ*, **129**, 1706–1717.

13 Brown, R.A.: 2004, New information from radial velocity data sets. *ApJ*, **610**, 1079–1092.

14 Stepiński, T.F., Malhotra, R., and Black, D.C.: 2000, The upsilon Andromeda system: models and stability. *ApJ*, **545**, 1044–1057.

15 Goździewski, K. and Migaszewski, C.: 2006, About putative Neptune-like extrasolar planetary candidates. *A&A*, **449**, 1219–1232.

16 Charbonneau, P.: 1995, Genetic algorithms in astronomy and astrophysics. *ApJSS*, **101**, 309–334.

17 Poincaré, H.: 1897, Sur une forme nouvelle des équations du problème des trois corps. *Bull. Astron.*, **14**, 53–67.

18 Poincaré, H.: 1905, *Leçcons de Mécanique Céleste*, Gauthuer-Villars, Paris, Vol. I.

19 Laskar, J.: 1991, In *Predictability, Stability and Chaos in N-Body Dynamical Systems*, Plenum Press, NY, pp. 93–114.

20 Laskar, J. and Robutel, P.: 1995, Stability of the planetary three-body problem. I. Expansion of the planetary Hamiltonian. *CMDA*, **62**, 193–217.

21 Michtchenko, T.A. and Ferraz-Mello, S.: 2001, Modeling the 5:2 mean-motion resonance in the Jupiter–Saturn planetary system. *Icarus*, **149**, 357–374.

22 Beaugé, C. and Michtchenko, T.A.: 2003, Modelling the high-eccentricity three-body problem. Application to the GJ876 planetary system. *MNRAS*, **341**, 760–770.

23 Ferraz-Mello, S.: 2007, *Canonical Perturbation Theories, Degenerate Systems and Resonance*, Springer.

24 Lee, M.H. and Peale, S.J.: 2003, Secular evolution of hierarchical planetary systems. *ApJ*, **592**, 1201–1216.

2
Terrestrial Planets in Extrasolar Planetary Systems

Rudolf Dvorak and Elke Pilat-Lohinger

Abstract

In this chapter we deal with the dynamical stability of fictitious terrestrial planets in exosolar planetary systems, and concentrate on possible motions in the habitable zones of these systems (see Chapters 4, 6 and 11). Up to now, all planets found outside our solar system are of the size of several earth masses and therefore cannot be regarded as terrestrial planets (which are the interesting ones for they can build up a biosphere). Theoretical studies of fictitious objects – a widely used technique in astrodynamics – are very important to help observers in their search for "a second Earth". The conditions to be fulfilled to allow for a terrestrial planet are reviewed from the dynamical point of view: the interesting regions of motion around a host star have been studied by many authors with the aid of extensive numerical integrations. Besides the stability of an orbit itself, its dynamical evolution, changes of its semi-major axis, its inclination as well as its eccentricity are studied. These investigations were accomplished in different dynamical models, with the aid of numerical integration methods of the equations of motion and with different methods of analysis (e.g., chaos indicators).

Three types of "habitable" regions for terrestrial planets are the subject of investigation:

a) the Jupiter-like planet is outside its star's habitable zone,
b) there is a hot Jupiter in such a system,
c) the giant planet moves inside the habitable zone.

Many studies have been undertaken for systems with only one giant planet, and with more than one gas-planets present in a system respectively. We present the results of several investigations of different authors and discuss also global results of attempts to compile catalogues where the dynamics of planets in habitable zones for most systems can be estimated according to the characteristics of the extrasolar planetary systems. An interesting conclusion can be drawn: the chance for terrestrial planets to move well inside habitable regions around a host star on stable orbits is not too small. The respective es-

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

timates from different authors are between 10 and 60 percent for the presently known extrasolar planetary systems!

2.1
Introduction

Although we have not found any terrestrial planet outside the solar system yet, the scientific astronomical community is convinced that our galaxy is full of them. With the European satellite CoRoT successfully launched recently[1] we are optimistic to find such planets. (For a detailed description of CoRoT see Chapter 3.)

With the aid of the NASA counterpart, the Terrestrial Planet Finder TPF[2] scientists hope also to extend their knowledge about extraplanetary systems (= EPS) significantly:

> “. . . TPF will revolutionize major areas of both planetary and non-planetary science . . . , prior to the start of TPF, ground- and space based searches will confirm the expectation that terrestrial planets are common around solar-type stars.[3]”

The many planetary systems in our solar environment, which have already been discovered during the last 10 years, show quite plainly that the detection of the first terrestrial planet (= TP) is just a question of observational technique and therefore is only a question of time.

Nevertheless it will be difficult to find such a “second Earth”, because there are many constraints for a planet to fulfill to develop and keep its biosphere: e.g., a surface with plate tectonics, a similar atmosphere to the one of our earth, etc. All these critical circumstances are – to a large part – dependent on the host star, its spectral type and age, but also on the orbital dynamics of the TP. Even when moving in a low eccentric orbit with small perturbations it must stay at the “right” distance to the star, to be able to maintain an earth-like environment and thus it should remain in the so called habitable zone (= HZ)[4] (see Chapters 4, 6 and 11 in this book).

When we look at our solar system, its structure seems to be quite particular: inside the quasi-circular orbits of the gas-giants Jupiter, Saturn, Uranus and Neptune, the four terrestrial planets Mercury, Venus, Earth and Mars are orbiting the sun. Additionally other earth-sized planets are satellites of Jupiter

---

[1] CoRoT was launched on the 27th december 2006 into its halo orbit around the Lagrange point L_2.

[2] The Terrestrial Planet Finder will be launched by NASA, probably in 2010; see Chapter 10.

[3] NASA: Earth-Like Exoplanets: The Science of NASA’s Navigator programme, 2006.

[4] A somewhat rough outline of the habitable zone is *where water could exist on the surface of a planet in liquid form;* for a more precise definition see Kasting et al. [1] and other chapters of the book.

(e.g., Europa) and Saturn (e.g., Titan) or reside in orbits outside Neptune in the Kuiper-belt[5)].

Out of more than 200 planets known up to now (February 2007) we can observe only large planets from six earth masses (Gliese 876 d) up to several Jupiter masses. Apart from the problem of stability of the observed planets, the search for stability regions of still hypothetical planets is interesting and can be studied with tools of astrodynamics. The constraints have already been mentioned: stable low eccentric orbits of the TPs in the HZ of the host star are needed.

In Extrasolar Planetary Systems (= EPS) we can distinguish four different groups in respect of the location of the HZ; the classification depends on where the Jovian planet(s) (Gas-giant = GG) is (are) moving (see Fig. 2.1):

- **G1**: if the GG is very close to the star (= hot Jupiter), stable orbits of TPs can exist in the HZ for time scales long enough to develop a biosphere.
- **G2**: if this GG moves far away from the central star (like Jupiter), then stable low eccentric orbits for TPs in the HZ can occur.
- **G3**: if the GG itself moves in the HZ, a terrestrial satellite (e.g., Titan in the system of Saturn) can have a stable orbit.
- **G4**: if the GG itself moves in the HZ, a Trojan-like TP may move in a stable orbit around the Lagrangian equilibrium points L_4 or L_5.

In addition, combinations of all four groups may occur, when two or more planets are present.

Apart from the global studies of Menou and Tabachnik [2] and Jones et al. [3] concerning the stability of orbits of TPs in EPSs, there are many investigations on specific systems for the four mentioned groups: e.g., Érdi and Pál [4], Érdi et al. [5], Dvorak et al. [6], Dvorak et al. [7], Dvorak et al. [8], Jones et al. [9], Jones and Sleep [10], Laughlin and Chambers [11] and Asghari et al. [12].

In the following, we will present the kind of methods used to investigate the dynamics of fictitious TPs in EPSs (Section 2.2), briefly explain the problem of formation of terrestrial planets (Section 2.3), then discuss the results of studies of some examples of the **G2** and **G4** group (Section 2.4). We point at a catalogue of single-planet systems (Section 2.5), show an example of TPs in multiplanetary systems (Section 2.6) and finally discuss what we may expect of future discoveries on the basis of our present knowledge of EPSs.

[5)] This does not fit into the new definition of planets and dwarf planets of the IAU.

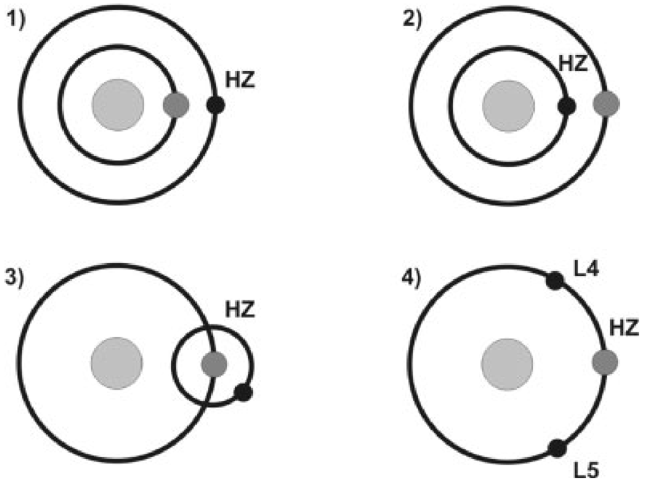


Fig. 2.1 Possible habitable regions for terrestrial planets in EPSs (**G1** to **G4**): the full black circles show the terrestrial planet, the light grey full circles show the gas-giant.

2.2
The Methods of Investigation

The question of the dynamical stability of hypothetical terrestrial planets can be treated in a system with only one giant planet in the framework of the well known and well studied elliptic restricted three-body problem (= ER3BP)[6]. In a system with more planets, the only reasonable way to test the orbital stability of additional planets seems to be performing numerical investigations. Different integrators for solving the equations of motions of the n-body problem are available, e.g., the Mercury6 package [13], the Lie-integration methods [14, 15] and many more.

- For the first group **G1**, where a hot Jupiter is present in the system with an almost circular orbit close to the host star, a possible TP in the HZ (around 1 AU) would not suffer from large perturbations which could cause the planet's orbit to be unstable. We will not discuss this situation further, it mainly is of interest to research treating the formation of TPs.
- For the second group **G2** we do have results of analytical studies for asteroids in mean motion resonances (= MMR) with Jupiter which is the dynamical model of the ER3BP. No analytical results for the stability regions of fictitious TPs were yet to be found in this model, where we have differ-

---

[6] Given the gravitational field of two massive primaries having Keplerian orbits around their common barycenter, one studies the orbit of a third body, which is regarded as massles and thus does not influence the primaries' motion.

ent mass ratios of the primaries and different eccentricities of the primaries' orbits. Interesting results have been derived for the averaged ER3BP for special MMR of periodic orbits [16], but no results for global studies of the dynamics of this group are available.

- For satellite TPs (**G3**), we can conclude from observations of the satellite systems around the GGs in the solar system that, for small eccentric planetocentric orbits with a semi-major axes $a < 1 \times 10^6$ km (e.g., Jupiter's Ganymede and Saturn's Titan), stable orbits do exist. Even large eccentricitic and retrograde orbits are stable; but for TPs in EPSs these studies should be done separately. In a recent paper Domingos et al. [17] investigated these orbits for many EPSs and concluded that most of such systems with a GG in the HZ could host TPs as satellites.
- For the 1:1 MMR (**G4**), the Trojan TPs, a lot of analytical work was done concerning the stability of the Lagrangian points in this model of the aforementioned ER3BP depending on the mass ratio of the primaries and the eccentricities of their orbits. A quite important contribution to this topic was the study of cases when the third body was not regarded as massless [18]. The results of a first-order stability analysis in the framework of the general three-body problem have been presented in the aforementioned book by Marchal [18]. With M the total mass and $m_1 \leq m_2 \leq m_3$ a mass parameter R was defined as $R = (m_2 + m_3)/M + m_2 m_3/m_1^2 + O(m_2^3 m_3/m_1^4)$. Using these results one can see in Fig. 2.2[7] that, in the case of a terrestrial-like planet with a relatively small mass (general three-body problem), compared to the two primary bodies there is no difference concerning the stability of the equilibrium points with the ER3BP. The fact is clear; when we set $m_3 = 0$ the parameter R reduces to the mass ratio $\mu = m_2/(m_1 + m_2)$.

  It is important to say that stability does not necessarily mean that one can find stable Trojan orbits in the vicinity of these equilibrium points, because to claim that, we would need to know the extension of the stable regions. Checking the observed eccentricities of the orbits of the planets in EPSs, the estimated (minimum) mass of the giant planet and the mass of a TP (comparable to the mass of our earth) it turns out that all planetary systems given by Jean Schneider[8] have stable equilateral Lagrangian points. Surprisingly enough this is also true for planets with large eccentric orbits of the primaries. Analytical studies concerning the extension of this region for Jupiter and the other giants have been undertaken to find an effective stability for the stability regions. It will be possible to adapt these results also to EPSs [19].

---

[7] Figure 13, p. 49 in Marchals's book.

[8] http://exoplanet.eu

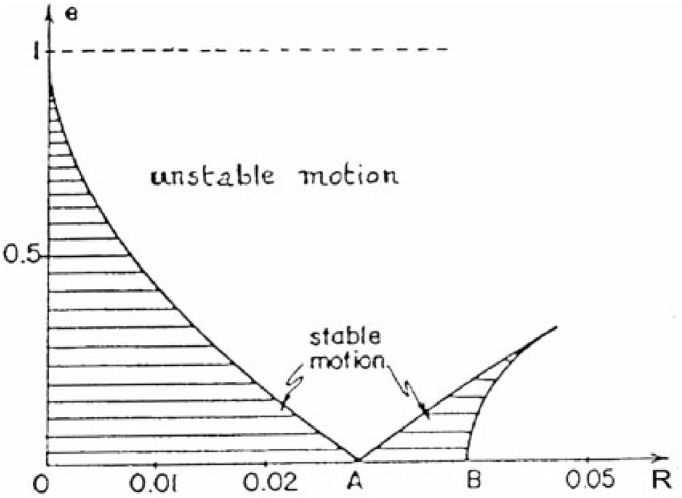


Fig. 2.2 Stability of the equilateral Lagrange points in the elliptic restricted three-body problem eccentricity versus mass ratio μ (after C. Marchal [18]).

Quite often, the analysis of the numerical results is carried out by making use of chaos indicators like the RLI [20], the MEGNO [21], the FLI [22] or the Lyapunov Characteristic Exponents (LCE) directly. According to Barnes and Raymond [23] "…direct N-body integrations are the best available method for determining stability". They argue that the finding of chaotic motion is not equivalent to unstable motion which is clearly shown in our solar system (see [24–26]). In particular the inner solar system planets are moving in chaotic orbits but they are stable at least for the time scale of the planetary system.

An efficient method in this sense is observing the eccentricities of the hypothetical TPs during the integration and to report the largest value. Another possibility is to watch out for close approaches (e.g., within the Hill's radius) to a large planet which cause a subsequent escape of the TP from its actual orbit. While the escape time (or stability time) is used by several authors, the check of the eccentricity additionally tells us a lot about the amount of radiation the planet received from its host star. A limit of $e < 0.2$ for a stable "habitable" orbit of a TP gives reasonable results because then the difference in distance between periastron (P) and apoastron (A) is 50 percent (e.g., for semi-major axis $a = 1$ AU $P = 0.8$, $A = 1.2$). The maximum eccentricity method (MEM) is thus a simple check for statistical quantities that does not require additional computation resources, just like using chaos indicators[9].

---

[9] One may ask why we call this a "method". The value of e_{max} of the orbit of a TP tells about the stability of its orbit in an EPS. The check for the largest eccentricity of an asteroid moving in the main belt (e.g., the 3:1 MMR) is also a reliable tool to detect an unstable orbit.

Nevertheless the computation of chaos indicators is very useful and can be seen as a complementary tool.

2.3
Basics of the Formation of Terrestrial Planets

Some articles concerning the possible formation of terrestrial planets [27, 28] cast doubt recently that the planetary systems found may have terrestrial habitable planets. Gas-giants are thought to form much faster than terrestrial planets; thus the authors made numerical integrations with a gas giant between 1.6 and 6 AU on orbits with eccentricities up to $e = 0.4$, computed the accretion possible under such conditions, and received results comparable to the moon and up to Mars-sized planets only. In such numerical simulations of several hundred n-body systems they found out how terrestrial water-rich planets could be formed. The results excluded the existence of such planets when a Jupiter-like planet is present at a distance closer than a=3.5 AU and also when the eccentricity of the orbit of the GG is significant. The authors concluded that: “the majority of the current sample of extrasolar giant planets appear unlikely to form habitable planets”. Although this sounds quite pessimistic, one can also say that the possibility of TPs around other stars is *not* so small. The same authors expressed this opinion themselves in the journal *Science* (Raymond et al. [29]) wherein they claim: “More than a third of the known systems of giant planets harbor earth-like planets”. In this article the authors took the possible migration of the giant planet caused by the torques with the massive disk of gas into account. So there are several mechanism to explain the existence of terrestrial planets in the HZ, inside (without migration of the GG) or outside a GG (hot Jupiter, after migration). A new investigation by Fogg and Nelson [30] shows, that even in hot Jupiter systems, terrestrial planets can form in the habitable zone during and after the migration of the GG, and that they may even have water on their surface!

So we do have reasonable theories about the formation of terrestrial planets for the groups **G1** and **G2**. Even the existence of earth-like Trojans, captured into a 1:1 resonance with a GG can be modeled, which could have been captured after (during) the migration of the large planets [11]10). Nevertheless the answer to the question whether TPs may form in EPSs is ambivalent, so much more theoretical work has to be done on this subject and further discoveries of EPSs are required.

---

10) Compare the recent articleMorbidelli et al. [31] for the formation of Jupiter Trojans in the solar system.

2.4
Stability Studies of Terrestrial Planets

A complete numerical study has been performed by Menou and Tabachnik [2] concerning the existence of stable orbits for TPs. They integrated 100 massless bodies for time scales up to 10^6 years, all randomly distributed in the habitable zone for each system known at that time. Their stability limit was defined in such a way that either the orbit became parabolic or hyperbolic, or the TP suffered from a close encounter within a distance of $3R_{Hill}$ 11) to the GG. Additionally EPSs were not regarded as possibly hosting "interesting" TPs when they left the HZ. According to their results, about half of the investigated 85 EPSs "are unlikely to harbor habitable terrestrial planets because of the strong perturbing influence of giant planet(s) in the vicinity of habitable zones." The authors estimate that about 25 percent may possibly be in a state of habitability as far as their orbits are concerned. The other TPs may have eccentricities too large to have low seasonal climate changes. In a more recent study Jones et al. [3] investigated the stable regions in the HZ of 152 EPS. They also used the Hill radius as their stability criterion. The respective conclusions are that 41 percent of the EPSs known to date are in a state of sustaining habitability.

Both studies ignore or underestimate the role of MMR, which can be rather important and may destroy stability altogether. If the MMR are dense, i.e., they overlap, a strong chaotic zone is present which may lead to a destabilization of a planet's orbit moving in such a region. We will come to this point when we discuss as an example the EPS Gl777A. Another aspect is a possible additional – still undetected – planet changing the picture, because then secular resonances between the large planets and the test planets arise inside the MMRs [32, 33].

2.4.1
The G2 Systems and Gliese 777A

By giving a few examples of systems where the GG orbits its host star in such a distance that a TP could move in the HZ (which would be well inside the GG's orbit) we will describe five EPS and one of them in a more detailed fashion. Here we would like to follow the results described in Asghari et al. [12].

The computations in the aforementioned article were performed with orbital elements known at that time; in Table 2.1 we show the new data derived from recent observations also. One can see the problem of determining stability regions: quite often, additional data from observers cause a change in the

---

11) This is defined as the radius ρ of a sphere around a secondary body with mass $\mu_2 = (m_2/(m_1+m_2))$ in the ER3BP, where the tidal forces and the mutual attraction of a massless body are such that the radial force vanishes. $\rho_{Hill} = (\mu_2/3)^{1/3}$.

Table 2.1 Parameters of the investigated systems. The upper line lists the values known in September 2003 (put on the webpage by J. Schneider); with these values the computations were carried out in the paper Asghari et al. [12]. The values in the lower line show the orbital parameters obtained from recent observations in http://exoplanet.eu/catalog.php (6th Feb. 2007).

Name	Mass	Semi-major axis [AU]	Eccentricity
HD 190360 (G6IV)	$0.90\ M_\odot$	–	–
	$0.96\ M_\odot$	–	–
HD 190360 b	$1.33\ M_J$	4.8	0.48 ± 0.2
	$1.502 \pm 0.130\ M_J$	3.92 ± 0.2	0.36 ± 0.03
HD 190360 c	$0.057 \pm 0.015\ M_J$	0.128 ± 0.002	0.01 ± 0.1
HD 72659 (G0V)	$0.95\ M_\odot$	–	–
HD 72659 b	$2.55\ M_J$	3.24	0.18
	$2.96\ M_J$	4.16	0.2
14 Her (K0V)	$1.00\ M_\odot$	–	–
	$0.9\ M_\odot$	–	–
14 Her b	$4.89\ M_J$	2.85	0.38
	$4.64 \pm 0.19\ M_J$	2.77 ± 0.05	0.369 ± 0.005
14 Her c	$2.1\ M_J$	6.9	0
HD 4208 (G5V)	$0.93\ M_\odot$	–	–
HD 4208 b	$0.8\ M_J$	1.67	0.05
47 UMa (G0V)	$1.03\ M_\odot$	–	–
47 UMa b	$2.54\ M_J$	2.09	0.061 ± 0.014
	$2.6 \pm 0.13\ M_J$	2.11 ± 0.04	0.049 ± 0.014
47 UMa c	$0.76\ M_J$	3.73	0.1 ± 0.1
	$1.34 \pm 0.22\ M_J$	7.73 ± 0.58	0 ± 0.12

orbital elements of the planets. Thus it is reasonable to use different orbital parameters within given error bars. The compilation of a whole catalogue for motions of TPs covering a large range of orbital eccentricities and mass ratios between the host star and the giant planet would be an even better idea (see Section 2.5).

A schematic view of Fig. 2.3 shows the dynamical structure of the EPS we are dealing with, namely of **Gl614**, **HD4208**, **HD72659** and **47 Uma** (a system with two giants). The results derived via numerical integration for a fine grid of initial conditions for the four systems may be summarized (for details see [12]) as follows (schemata from top to bottom):

- **Gl614**: It is unlikely that a TP will move there, the large mass in connection with the relatively large eccentric orbit of the GG causes large perturbations inside the HZ.
- **HD4208**: In this system there seems to be enough space for a TP. Orbits with low eccentricity can be stable.
- **HD72659**: This is quite a good candidate for hosting a TP with a low eccentric orbit in spite of the eccentricity of 0.18 for the GG.
- **47 Uma**: There is a good chance for TPs to move inside the HZ with relative small eccentricities between the strong acting MMR 3:1 and 4:1. The other MMR (5:2, 7:3 and 2:1) visible in Fig. 2.3 (bottom scheme) do not cause large perturbations.

A planet in Gl 777 A (= HD 190360) was discovered by Naef et al. [34]. This EPS consists of a wide binary (with $a = 3000$ AU) and therefore it is not necessary for dynamical studies of motions of TPs to take the more distant companion into account. The star is a G6 IV with $0.9 M_\odot$ which hosts a planet with a mass of 1.33 M_{Jup} and a semi-major axis of 4.8 AU (see Table 2.1). The large eccentricity ($e = 0.48$) limits the possible region of motion for an additional TP to $a < 2.4$ AU (= periastron). In Fig. 2.3 one can see the main characteristics and that low-order MMR are not acting in the HZ.

Its HZ corresponds to a region between 0.7 AU < a < 1.3 AU. The stability diagram (Fig. 2.4) shows, in the blue and green region, a large stable zone for a fictitious TP to move there with moderate eccentricities $e < 0.25$. Using the MEM, one can see some interesting features: strong vertical lines due to high-order resonances with large eccentric orbits in this resonances and a growing number of unstable orbits with larger semi-major axes of the terrestrial planet. All in all one can say that Gl 777 A is an EPS with a high possibility that planets will stay within the habitable zone for very long times with moderate eccentricities which are a necessary condition for a stable biosphere. With the new data (second line in Table 2.1) no significant changes of the stability regions are expected, because the smaller semi-major axis is compensated by a smaller orbital eccentricity of the GG.

In the global study of Menou and Tabachnik [2] they conclude that Gl 777 could host TPs within its HZ; in Jones et al. [3] the authors come to the same conclusions. In another investigation by Sandor et al. [20] it is mentioned that, with the new values of eccentricity the HZ is strongly chaotic and a region of overlapping of resonances. This result was interpreted using the RLI data for only some 30 revolutions of the primary bodies, but we would like to stress once more that a chaotic motion is not equivalent to an unstable one.

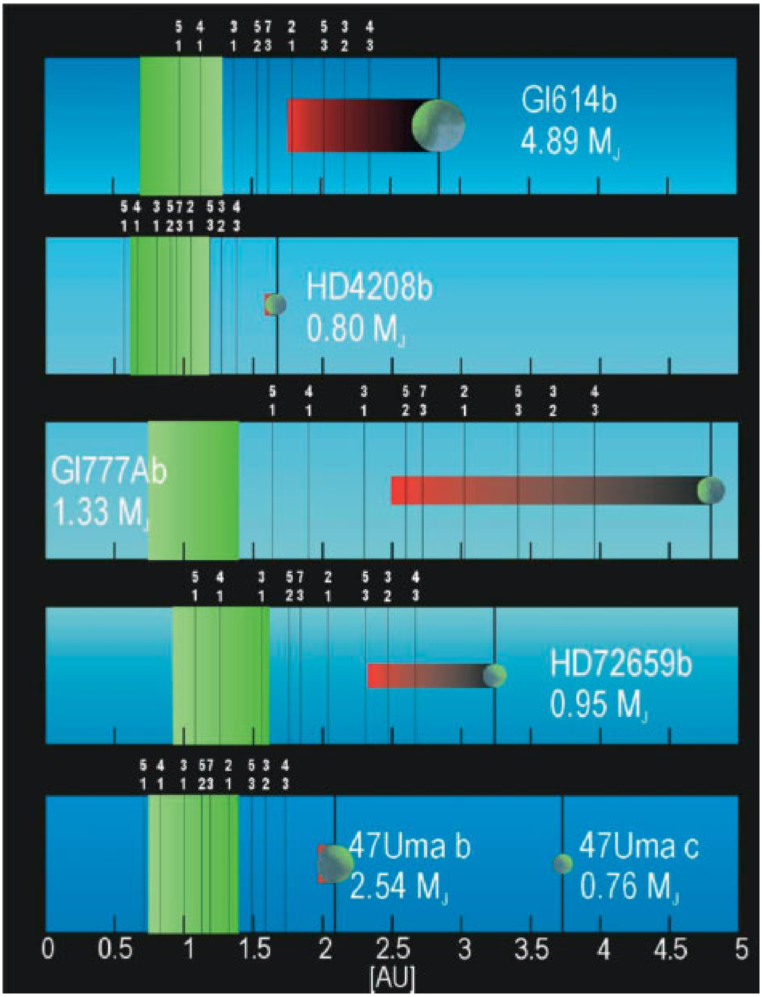


Fig. 2.3 Schematic view of five extrasolar planetary systems. From top to bottom: Gl614, HD4208, Gl777A, HD72659 and the system with two gas-giants 47Uma. x-axis is the distance from the planet to the host star, the red bands left to the planets show the region where the planets are moving towards the periastron position, the size of the planet is proportional to the minimum mass determined by the observers. The green region between the planet and the star marks the habitable zone. The dark verical lines with the numbers at the top of every scheme characterize the mean motion resonances of a hypothetical terrestrial planet with the GG (after F. Freistetter, unpublished).

2.4.2
Theoretical and Numerical Stability Investigation of the G4 and EPS HD108874

Several numerical studies are devoted to finding stability regions of type **G4** in EPSs as in Refs [4, 5, 35–38]; all of them are aiming to find the extension of such stable zones for fictitious TPs in semi-major axes and in libration around the equilateral equilibrium points.

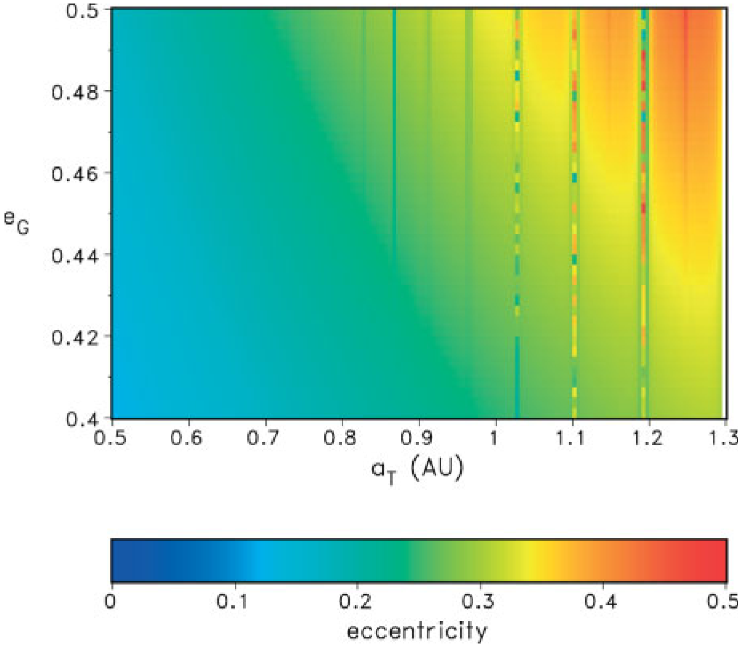
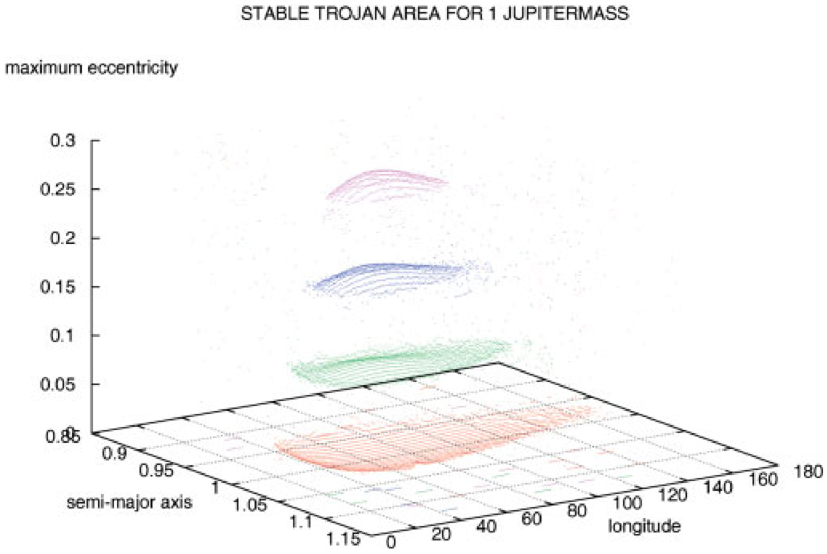


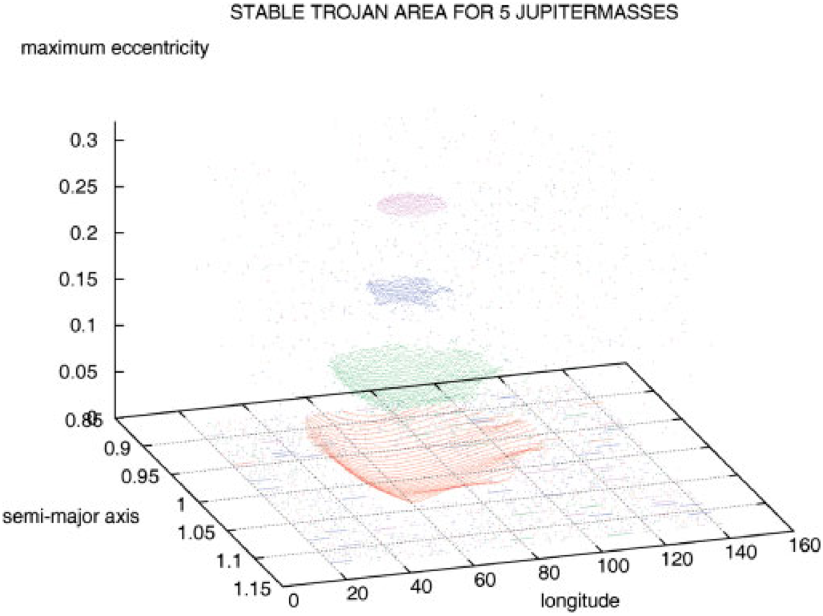
Fig. 2.4 Initial condition diagram of semi-major axis of the fictitious terrestrial planet versus the eccentricity of the GG of the extrasolar system Gl 777 A. The colors indicate the maximum eccentricity achieved during the integration of 1 million years.

From a theoretical point of view, the stability of the Lagrange points is well known in the ER3BP (see Section 2.2). The question of the extension is a more interesting one for TPs in EPSs and not yet answered in a satisfying way. Nevertheless, for special mass ratios, close to the earth–moon system there do exist numerical studies to establish such regions (e.g., [39, 40]). The interesting question for **G4** types is the following one: how does the extension of the stable region depend on the mass ratio of the primaries and of the eccentricity of the orbit of the GG? This means that a two-parameter study should be done. A numerical study (yet unpublished) provided the results which are shown in Figs. 2.5 and 2.6. It is quite obvious in Fig. 2.5 that – for a Jupiter mass GG (a) and a GG with five times the mass of Jupiter (b) – this stability region diminishes with a larger eccentricity of the primary's orbit.

In Fig. 2.6 we show the results of the two parameter study for 1, 3, 5 and 7 Jupiter masses and different eccentricities within $0.0 \leq e_{GG} \leq 0.5$ with $\delta e = 0.1$. It is visible that both parameters lead to a shrinking of the stable area around the equilibrium point, namely with increasing mass and increasing eccentricity. The criteria to establish stability was again the MEM, the time interval for the integrations was 10^6 periods of the primary's orbit.



(a)



(b)

Fig. 2.5 Stability region around the equilateral equilibrium point in the elliptic restricted three-body problem for four different eccentricities of the orbit of the GG (e = 0.0, 0.1, 0.2 and 0.3, lowest to highest plot) and for two different masses of the GG. The initial conditions in the x-axis are the angular distance to the point L_4 and the semi-major axis in the y-axis. Along the z-axis we plotted the maximum eccentricity of the "surviving" stable orbits.

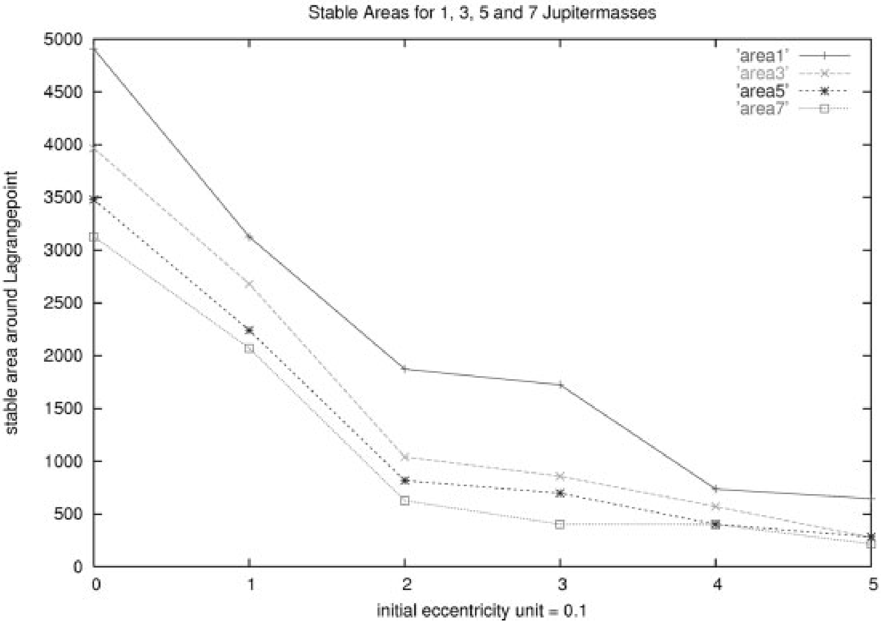


Fig. 2.6 Stable area (y-axes) around the equilateral equilibrium points in the ER3BP depending on the mass of the GG and the eccentricity of the primary's orbit (x-axes).

In a somewhat more detailed investigation, where the mass of the fictitious TP was also taken into account [41], these regions were determined via short time computations with the RLIs. These results agree with a recent publication [38] for a fine grid of initial conditions within $0.0 \leq e_{GG} \leq 0.3$ and a large scale to mass ratio μ of the primaries. The extension of the area was checked with the aid of the maximum eccentricity used as a purely numerical stability criterion: an orbit was regarded as unstable if $e_{max} > 0.5$. This limit for e_{max} is supported by many investigations of this problem. The results in Fig. 2.7 show the following: close to the $e_{primaries} = 0$ (the circular problem) and small mass ratios, the stable region is large (red and then orange), the edges of these large stable areas consist of smaller and smaller areas around the equilibrium point (yellow and then green). The large region outside (larger eccentricity and large mass ratio, blue color) depicts escaping orbits around the equilibrium points. The border between stable and unstable regions coincides quite well with the stability diagram (Fig. 2.2). The two unstable regions correspond to the mass ratios ($\mu = 0.015$ and 0.025) and are caused by the 3:1 and 2:1 resonances between the orbital period of the large planet and the libration period of the Trojan. The investigation was accomplished just in the ER3PB, because recent results have shown that, by giving the fictitious TP a mass of the order

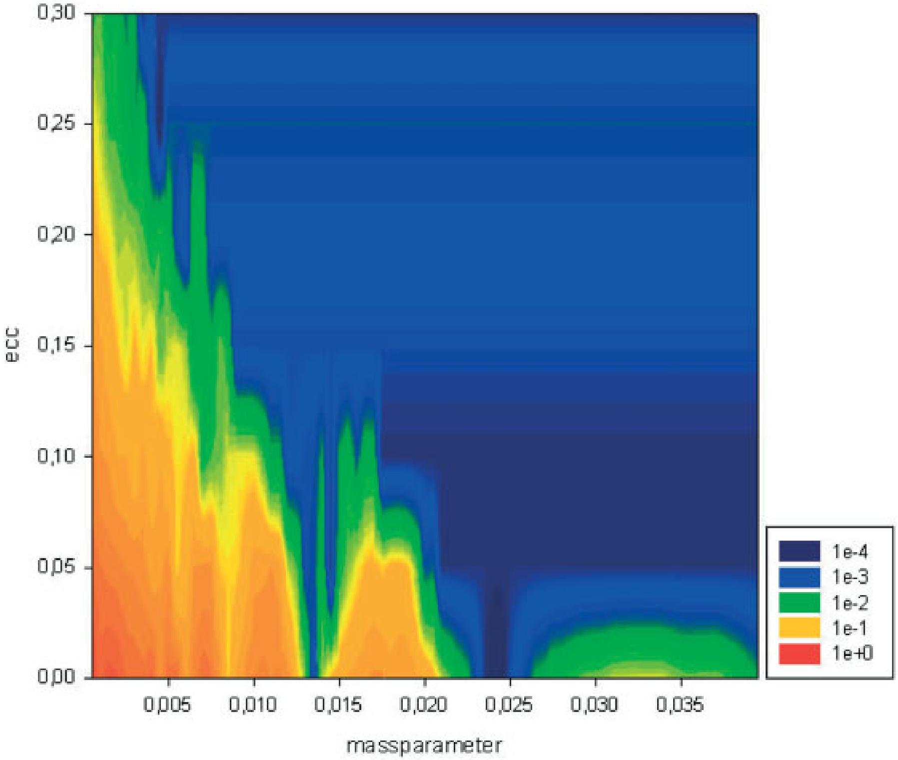


Fig. 2.7 Stability regions around an equilateral equilibrium point in the ER3BP. We varied the mass parameter μ (x-direction) and the eccentricity (y-direction) of the primary's orbit. The colors from red to blue indicate the extension of the region (e.g., green color on the edge means that the area diminished in size to just 1 percent of the red region!

of an earth mass – which is the most interesting case from an astrobiological point of view – the stability of the equilateral Lagrange points themselves does not change and the extension of the stable region is not reduced [42].

There is one very good candidate, when it comes to hosting a Trojan terrestrial planet in the habitable zone, namely HD 108874 [43]. The sun-like host star was known to have only one massive planet with approximately the mass of Jupiter and an orbit with a small eccentricity in the HZ. Recently a second planet was observed outside the orbit of the already known one [44]. In Table 2.2 we show the orbital parameters derived from observations with the given error bars.

Nine different models were tested by Schwarz et al. [42] to take the possible observational errors into account, especially those of the semi-major axis and

Table 2.2 Orbital elements of the EPS HD108874.

NAME	HD108874 b	HD108874 c
mass [M_j]	$1.36(\pm 0.13)$	$1.018(\pm 0.03)$
a [AU]	1.051 ± 0.02	2.68 ± 0.25
orbital period [days]	$395.4(\pm 2.5)$	$1605.8(\pm 88)$
eccentricity	$0.07(\pm 0.04)$	$0.25(\pm 0.07)$
ω [deg]	$248.4(\pm 36)$	$17.3(\pm 23)$

the eccentricity of the perturbing outer giant for the computation of orbits of fictitious Trojan planets. We used the nominal values, but also lower and upper bounds according to the given error bars in e and a:

- Models **M1x**, namely **M11, M12** and **M13** with $a = 2.43$ AU and three values for the eccentricity $e = 0.18, 0.25, 0.42$.
- Models **M2x**, namely **M21, M22** and **M23** with $a = 2.68$ AU and again $e = 0.18, 0.25, 0.42$.
- Models **M3x**, namely **M31, M32** and **M33** with $a = 2.93$ AU and $e = 0.18, 0.25, 0.42$.

The results of the integrations show that, in models **M1x** no orbit survived for times longer than 1 million years. For **M12** and **M13** this is due to the instability of the orbits of the two planets themselves; in fact, the same happens in **M23**. The orbits of the TPs for the largest initial eccentricity ($e = 0.32$) in models **M23** and **M33** are not stable.

The stable region for the Trojan planets is shown in Fig. 2.8a for **M21**, where one can see that this region around the equilibrium point L_4 (located at $a = 1.051$ and $\omega = 308°$) extends from $295° < \omega < 325°$ and 1.035 AU $< a <$ 1.075 AU. For **M22** in Fig. 2.8b we can see just a very small region of stable orbits which we suspect to disappear during a longer integration. In Fig. 2.9a for **M31** we observe a slightly larger stable region around the equilibrium point, which can be understood considering that the perturbing planet in this model is farther away from the inner planet and also far away from any low-order MMR. In Fig. 2.9b the stable region in model **M32** almost disappears for a larger eccentricity of the outer planet. In Table 2.3 we list the extensions in the semi-major axis and ω for different models.

It is evident, from these results, that the parameters of the system need to fulfill special restrictions when one can expect Trojan-like TPs to move in the HZ on stable low eccentric orbits.

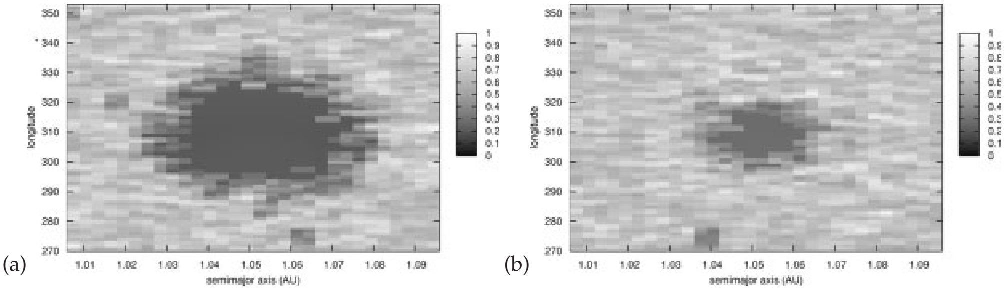

(a) (b)

Fig. 2.8 Stability regions around L_4 in model **M21** and **M22**. The initial semi-major axes (y-axis) are plotted versus the initial angular distance from the equilibrium point L_4. The dark regions around the point L_4 ($a = 1.051$ AU, $\omega = 308$) show that with such initial conditions the maximum eccentricity during the integration time was always $e_{max} < 0.2$.

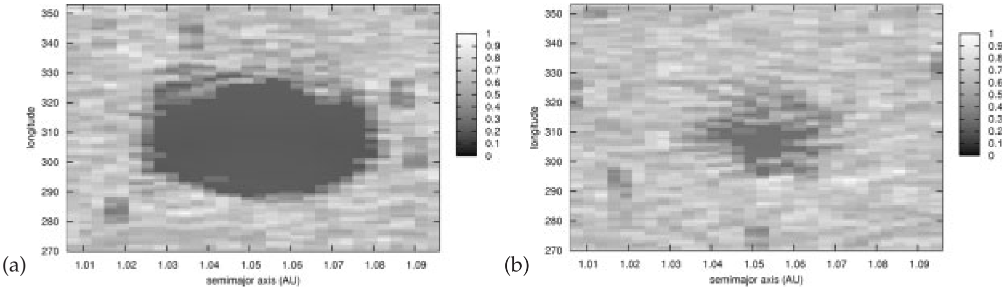

(a) (b)

Fig. 2.9 Captions as in Fig. 2.8 but for the models **M31** and **M32**.

Table 2.3 Extensions of stable regions for different models with respect to the semi-major axes and the angular distance to the Langange point L_4.

Stability parameter	M21	M22	M31	M32
a (AU)	2.68	2.68	2.93	2.93
e	0.18	0.25	0.18	0.25
$\delta\omega$	30°	14°	35°	10°
δa (AU)	0.04	0.015	0.045	0.01

2.5
A Global Approach: The Exocatalogue

Recently two groups involved with dynamical studies of fictitious terrestrial planets[12] compiled a catalogue of stable zones in "one planet EPSs" [45]. Every stability map shows the results of numerous numerical integrations of the equations of motion for fictitious TPs (massless!) in the ER3BP for a given mass ratio. The value of the RLI[13] is plotted in this initial condition diagram (the semi-major axes of the TP 0.1 AU $< a <$ 0.9 AU versus different eccentricities of the primaries' orbit $0 < e < 0.5$). The fine grid of $\delta a = 10^{-3}$ AU in semi-major axes and eccentricities $\delta e = 0.005$ unveils the "V-shape" structures of the resonances with increasing eccentricity of the GG. These MMR with the GG are marked on the upper line with numbers (like the 2:1 MMR at a = 0.65, which is known to be unstable in the main belt of asteroids). The scaling is chosen, so that "yellow" and "orange" may stand for regular motion, whereas "red, green and black" stands for chaotic orbits (see the scale given for each plot). Additionally the initial conditions for the fictitious test planet were varied in position (mean anomaly $0 < M < 2\pi$), the other elements were set to zero, which means that the orbit of the TP was initially circular; only the plane problem was treated.

The catalogue may simply be applied to any newly discovered EPS with one planet: in Fig. 2.10 we show four examples of systems, of which their orbital parameters are known to us. These four systems all have a mass ratio comparable to the one in our solar system where Jupiter is the dominating GG. It is plotted in the left lower corner of the figure and shows the extension of the habitable zone in semi-major axes. These values need to be multipled by the semi-major axes of Jupiter (5.2 AU) to get the estimated borders of the HZ of the Solar system. Consequently our planet earth is – fortunately – in a dynamically stable (yellow) region. In the same way we can handle the other three systems. Figure 2.11 represents a similar map for a different mass ratio, namely for a GG, five times more massive than Jupiter ($\mu = 0.005$). As explained above we can proceed to find out whether a system contains a habitable zone where TPs could move on nonchaotic orbits. To get the full information on how the catalogue is compiled we refer to the original paper and to the homepage http://astro.elte.hu/exocatalogue.

---

[12] A group at the Institute of Astronomy of the Eötvos University in Budapest (http://astro.elte.hu) and the AstroDynamicsGroup at the Institute of Astronomy at the University of Vienna (http://www.univie.ac.at/adg).

[13] The relative Liapunov Indicator (RLI) is a recently developed tool for detecting chaos in dynamical systems in relatively short time. A detailed technical explanation is given in [46].

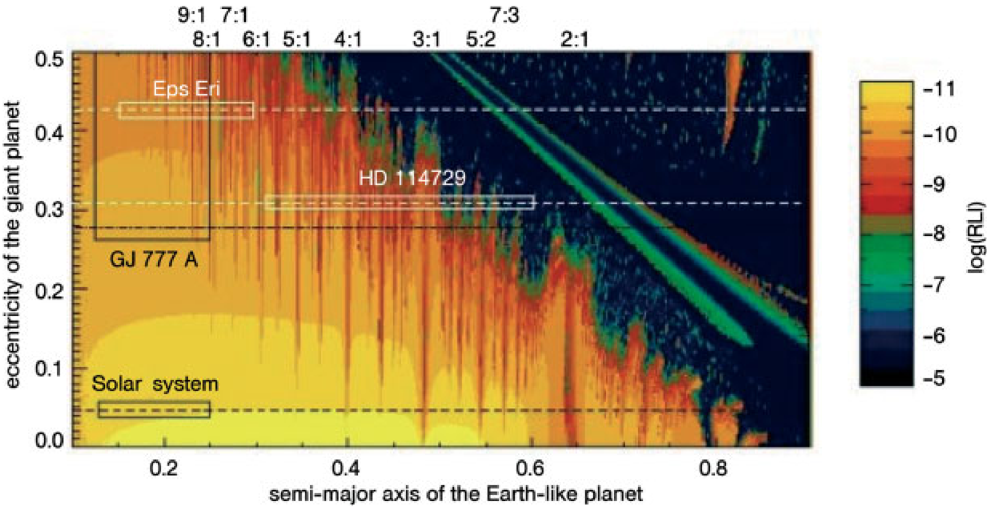


Fig. 2.10 Normalized Stability zones for the mass ratio $\mu = 0.001$ derived with the chaos indicator RLI. Dark regions correspond to chaotic orbits, the yellow and orange region contain regular orbits, dark red, green and black indicate chaotic orbits.

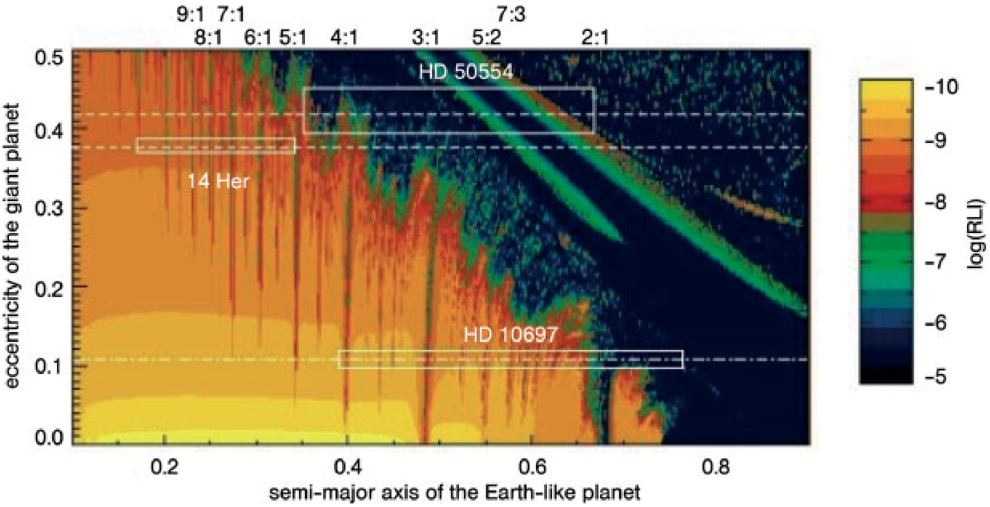


Fig. 2.11 Normalized Stability zones the mass ratio $\mu = 0.005$. Captions as in Fig. 2.10.

2.6
Terrestrial Planets in Multiplanetary Systems

In a series of papers [23, 27, 47] the stability of possible additional planets in multiplanetary systems is investigated via numerical integrations and the formation of such planets is discussed extensively. These systems can be classified [23] into three categories:

- **CI: resonant systems**, where the large planets are locked into a resonance.
- **CII: dynamically interacting systems** which are separated by less than a possible 10:1 MMR.
- **CIII: separated systems** which are beyond the 10:1 MMR between the two planets.

Numerical investigations of group **CI** and **CII** systems show that many of them are close to instability when one takes observational error bars for orbital elements into account (see Chapters 1 and 8 of this book). In this sense these two groups can be regarded as systems which do not allow additional planets to move in between them. Nevertheless TPs could still exist, either outside the two planets (when both are relatively close to the host star), inside (like the terrestrial planets in our solar system and the stable orbits of fictitious TPs in 47 Uma; see Section 2.4.1) or as TP in 1:1 resonance. These possibilities (groups **G1** to **G4**) are already mentioned in the introduction for single-star systems. Making the assumption that all systems may have "as many planets in stable orbits as possible", which is a somewhat sloppy definition of the *packed planetary systems* (= PPS) hypothesis, one can look for additional planets in any of the classes. In this sense the solar system is the typical example of such a PPS because in between the large planets (Jupiter to Neptune) there are only asteroids in unstable orbits (sometimes on very long time scales!).

2.6.1
The Changing Story of HD74156

In an abstract paper [48] Barnes and Raymond summarized the results of an investigation of several multiplanetary systems as follows: "Simulations of planetary system dynamics have shown that interacting systems are as tightly packed as possible. In this poster we explore the possibility that separated systems (specifically HD37124, HD74156, HD168443, HD12661, and 55Cnc) may contain additional companions, and hence are also interacting systems. First we fill the region between well separated planets with test particles and integrate for 10^7 years. We find that the systems HD168443, and HD12661 can hold no more planets. The other systems, however, contain broad zones of stability." This means that, according to this study, HD74156 contains "broad zones of stability".

In the aforementioned three papers cited in Section 2.6 the results of the numerical studies show that it is "unlikely" that the systems HD 168443 and HD 74156 host TPs. This result (no surviving orbits in the zone between the two massive planets) is in contradiction to their proper research, but it differs also from another investigation concerning HD 74156 [7], where TPs are found to be unstable.

In an extension to the dynamical study of HD 74156 [49] the authors put massive planets in the zone between the two planets, and found “that Saturn mass test planets are stable in many cases”. Their explanation of the different results for massless and massive planets is that, with a new massive planet, an apsidal alignment between the “new” planet and the outer planet leads to a stable orbit for 100 million years.

This example shows all the problems of determining the stable regions in two-planet EPSs which are much more difficult to handle than in the one-planet EPS: the uncertainties in the orbital parameters determined from observations lead to divergent results from one investigation to the other.

2.7
Conclusions

Planets around other stars have been a “hot topic” in astronomy for more than 10 years now, due to the fact, that this is a field of interest the human race shares as a whole. Although no planets like our earth have yet been observed, there is a huge number of TPs in our galaxy, as well as in the solar vicinity (most of the EPSs found so far are within 30 pc starting from the sun) – so no doubts about that within the scientific community any more. Our problem is how to find them. This question is addressed by more and more astronomers. Even the plans for an “Extremely Large Telescope” – a giant of 42 meters in diameter – which is under consideration by European scientists, list “the search for terrestrial planets around other stars” as a primary goal.14)

Up to now the questions of the existence of terrestrial planets in EPSs is primarily investigated via extensive numerical integrations for systems where one or more GG have been confirmed already. The attempts of general studies are only successful in systems only, where just one large planet is present. When another GG is moving within such a system, the number of parameters which are involved is too great. Just to cite some of them: the mass parameters star/GG1 and star/GG2, the semi-major axes of the two planets, their eccentricities... All this makes it impossible to find a general view of stable habitable zones for terrestrial planets in systems with more than one GG.

Many studies do exist for some systems like 55 Cancri with four known planets or upsilon Andromedae with three planets. Catalogues for one-planet systems can give a first hint of whether a newly discovered EPS may harbor terrestrial planets from a dynamical point of view. Ventures like this will be

---

14) e.g. http://www.astro-opticon.org/networking/elt.html

crucial for further observations (e.g., for CoRoT and TP) when one is "hunting" for terrestrial planets, for we are still looking forward to the statement "a second Earth has been found by astronomers!"

Acknowledgments

This study was also supported by the International Space Science Institute (ISSI) and benefits from the team "Evolution of Habitable Planets". EP-L wishes to acknowledge the support of the Austrian FWF.

References

1 Kasting, J.F., Whitmire, D.P., and Reynolds, R.T.: 1993 *Icarus*, **101**, 108–128.

2 Menou, K., & Tabachnik, S. 2003, *AJ*, **583**, 473–488.

3 Jones, B.W., Sleep, P.N., Underwood, D.R.: 2006, *ApJ* **649** 1010–1019.

4 Érdi, B., Pál, A.: 2003, Dynamics of resonant exoplanetary systems, In: *Proceedings of the 3rd Austrian-Hungarian Workshop on Trojans and related Topic* ed. by F. Freistetter, R. Dvorak. B.Érdi (Eötvös University Press) 3–10.

5 Érdi, B., Dvorak, R., Sándor, Zs., E. Pilat-Lohinger and B.Funk: 2004, *MNRAS* **351**, 1043–1048.

6 Dvorak, R., Pilat-Lohinger, E., Funk, B., and F. Freistetter: 2003, *A&A* **398**, L1–L4.

7 Dvorak, R., Pilat-Lohinger, E., Funk, B., and F. Freistetter: 2003, *A&A* **410**, L13–L16.

8 Dvorak, R., Pilat-Lohinger, E., Schwarz, R., and F. Freistetter: 2004, *A&A* **426**, L37–L40.

9 Jones, B.W., Sleep, P.N., Chamber, J.E.: 2001, *A&A*, **366**, 254 – 262.

10 Jones, B.W., Sleep, P.N. 2002, *A&A*, **393**, 1015–1026.

11 Laughlin, G., J.E. Chambers, J.E.: 2002, *AJ* **124**, 592–600.

12 Asghari, N., Broeg, C., Carone, L. et al.: 2004 *A&A* **426**, 353–365.

13 Chamber, J.E.: 1999, *MNRAS*, **304**, 793.

14 Hanslmeier, A. & Dvorak, R. 1984, *A&A*, **132**, 203–207.

15 Lichtenegger, H. 1984, *CMDA*, **34**, 357–368.

16 Psychoyos, D., Hadjidimetriou, J: 2005, *CMDA*, **92**, 135–156.

17 Domingos, R.C., Winter, O.C., and T. Yokoyama: 2006, *MNRAS* **373**, 1227–1234.

18 Marchal, C., 1990, *The Three-Body Problem*, Elsevier.

19 Lhotka, C., Efthimiopoulos, C., Dvorak, R. (in preparation).

20 Sándor, Z., Érdi, B., Széll, A., and B. Funk: 2004 *CMDA*, **90**, 127–138.

21 Cincotta, P.M. and Simó, C.: 2000, *A&AS.* **147**, 205–228.

22 Froeschlé, C., Lega, E., and Gonczi, R.: 1997, *CMDA*, **67**, 41–62.

23 Barnes, R. and S.N. Raymond: 2004, *ApJ*, **617**, 569–574.

24 Laskar, J.: 1994, *A&A* **287** L9–L12.

25 Laskar, J.: 1996, *CMDA* **64**, 115–162.

26 Lecar, M., Franklin, F.A., Holman, M.J., and N.J. Murray: 2001 Ann. Rev. of Astronomy and Astrophysics **39**, 581–631.

27 Raymond, S.N.: 2006, *ApJ*, **643**, L131–L134.

28 Raymond, S.N., Barnes, R., and N.A. Kaib.: 2006, *ApJ*, **644**, 1223–1231.

29 Raymond, S.N., Mandell, A.M., and Sigurdsson, S.: 2006*Science*, **313**,1413–1416.

30 Fogg, M. J., Nelson, R.P.:2007, *A&A* **461**, 1195–1208.

31 Morbidelli, A.; Levison, H. F.; Tsiganis, K.; Gomes, R.: 2005, *Nature*, **435** 462–465.

32 Morbidelli, A., Moons, M.: 1993, *Icarus* **102**, 316–332.

33 Morbidelli, A., Moons, M.: 1993, *Icarus* **103**, 99–108.

34 D. Naef, M. Mayor, S.G. Korzennik, D. Queloz, S. Udry, P. Nisenson, R.W. Noyes, T.M. Brown, J.L. Beuzit, C. Perrier, J.P. Sivan: 2003, *A&A* **410**, 1051–1054.

35 Érdi, B., and Sándor, Zs.: 2005, *CMDA* **92**, 113–121.

36 Schwarz, R., Pilat-Lohinger, E., Dvorak, R., Érdi, B., Sándor, Zs.: 2005, *Astrobiology Journal*, **5**, 579–586.

37 Schwarz, R. 2005b, *PHD thesis, University of Vienna*, online database: http://media.obvsg.at/dissd.

38 Schwarz, R.: 2006, In: Süli, Á., Freistetter, F., Pál, A. (eds.) Proceedings of the 4rd Austrian Hungarian Workshop on Trojans and related topics, Eötvös University Press, Budapest, 149–162.

39 Lohinger, E., Dvorak, R.: 1993, *A&A* **280**, 683–687.

40 McKenzie, R., Szebehely, V.: 1981, *Cel.Mech.*, **23**, 223–229.

41 Érdi, B., Fröhlich, G., Nagy, I., and Zsolt Sándor: 2006, Stability of hypothetical Trojan planets in exoplanetary systems, In: Süli, Á., Freistetter, F., Pál, A. (eds.) Proceedings of the 4rd Austrian Hungarian Workshop on Trojans and related topics, Eötvös University Press, Budapest, 149–162.

42 Schwarz, R., Dvorak, R., Pilat-Lohinger, E., Süli, A., Érdi, B.: 2007, *A&A* **462**, 1165–1170.

43 Butler, R., Marcy, G., Vogt, S. et al.: 2002, *ApJ*, **578**, 565.

44 Vogt, S.S., Butler, P.R., Marcy, W. G., Fischer A.D., Henry, W.G., Laughlin, G., Wright, T.J., Johnson, A.J.: 2005, *AJ*, **632**, 638.

45 Sándor, Z., Süli., A., Érdi, B., Pilat-Lohinger, E., and R. Dvorak: 2007, *MNRAS*, **375**, 1495-1502.

46 Sándor, Z., Balla, R., Téger, F, Ërdi, B.: 2001, *CMDA*, **79**, p. 29–40.

47 Barnes, R. and R. Greenberg: 2006, *ApJ*, **647**, L163–166.

48 Barnes, R., Raymond, S.N.: 2002, *AAS* **34** p. 1177.

49 Raymond, S.N., Barnes, R.: 2005, *ApJ* **619**, p. 549–557.

3
Mission Requirements: How to Search for Extrasolar Planets

Malcolm Fridlund and Lisa Kaltenegger

Abstract

In this chapter we will discuss the motivation, scientific requirements and technical implementation of space-based search for extrasolar planets. We discuss current and planned future missions, with a focus on designs that will be able to collect photons and therefore spectral information of the atmospheres of planets. The interferometric nulling technique, as well as simulations of the observations, are discussed in detail.

3.1
Introduction

Today, our generation is the first that possesses the technological ability to search for, and observe, planets orbiting other stars – so-called exoplanets. During the middle of the last century, a small number of scientists noted that this capability was emerging and suggested ways to implement it. Most notably, Struve [1] hypothesized that very large planets (Jupiter-sized or larger) could, if orbiting very close to a solar type star, be easily detected by the periodic change they would cause in the radial velocity (the stellar velocity along the line of sight) signature in their star's spectrum. Struve went on to point out that under these circumstances (a large planet close to its star), a percentage of the objects (of order 1%) would undergo eclipses which would also, in principle, be detectable using methods that were just becoming available around that time (e.g., modern photomultipliers as detectors). After these very astute predictions by Struve, almost 40 years were to pass before the hypothesis turned into observable fact when first Latham et al. [2] and then Mayor and Queloz [3] immediately followed by Butler and Marcy [4], discovered such bodies by detecting the radial velocity signature. By this time, the predictions by Struve had been more or less forgotten, and it came as somewhat of a surprise to find these very large planets close to their primary star. This "new" class of planets were then designated "hot Jupiters". Today they

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

are also sometimes called Pegasids after 51 Pegasi b, the first object detected in the study of Mayor and Queloz [3]. Since the method of discovery was by detecting only one component of the stellar velocity, and only a minimum mass could be determined for the first planets, some ambiguity existed if one was actually observing planets. For some time it was questioned wheter these peculiar objects orbiting very close to their primaries, which also often possessed a high degree of eccentricity, were not actually small stars seen in orbits almost face-on (e.g., [5]). Some kind of bias would have to be operating, through which one would select objects viewed pole-on, but one of the criteria for selection in the first objects was that they were slow rotators (in order to detect the small planetary signature more readily). This could possibly introduce such a bias.

Relatively soon after the first planets had been discovered, however, the first detection of a transit followed [6]. The planet in question orbits the star HD 209458 every 3.52 days at a distance of about 0.05 AU. Thus the transit lasts about 2.5 hours and, from this observation, the inclination is found to be 87.1 deg or nearly edge-on. Charbonneau et al. [6] derived a planetary mass of 0.63 M_{Jup} and a planetary radius of 1.27 R_{Jup}. This could be done since the orbital radius was well known from the radial velocity measurements, and the stellar radius was known to a reasonable accuracy from stellar evolution theory. The outcome of this observation of one transit was that the radial velocity measurements of a large number of potential planets immediately were considered to be bona fide exoplanets.

The continued work of what has eventually become a large number of scientists working in one of the most active fields of astronomy today has since then lead quickly to the understanding that "true" exoplanets have been discovered. Thus, the first stepping-stone on the path to finding true "exolife" has been registered. Since then, the field of exoplanets has developed with remarkable speed. Currently we know of more than 260 planets in more than 180 exosystems. What still eludes us is what has come to be regarded as the "holy grail" of Life Detection outside the solar system, namely "true" terrestrial (or "rocky") planetary bodies like our own earth found within their stars so-called "Habitable Zone". The Habitable Zone or HZ is defined as the region around a specific star where it is *in principle* possible to find the conditions that make life possible on our own Earth. Usually, it is taken to be the volume in space where water would be found in liquid form without postulating special conditions on the planetary surface like e.g., a very strong greenhouse effect. A related definition is the Continuously Habitable Zone, where one takes into account stellar evolution and the consequent change of the luminosity of the central source.

Although we have thus found plenty of planetary systems around solar-type stars in the vicinity of our sun, and, also taking into account that discoveries are continuing at an accelerated rate, what is currently lacking are any indication of *systems* like our own, e.g., giant planets in circular orbits found *outside* the HZ. The discovery of such systems would at least allow for the possibility of terrestrial bodies being in the HZ although current technologies do not allow for their detection (see also Raymond [7] on the formation of terrestrial planets in EGP systems). This situation is of course at least partly caused by an observational bias. The ground-based methods used so far – detection of the above mentioned radial velocity deflection caused in the stellar spectrum by a sub-stellar mass orbiting the primary, or the occultation of part of the stellar light when the planetary body passes between us and the star – has so far not had the sensitivity to either detect earth-sized objects, nor has the time been long enough to pick up planets similar to our Jupiter or Saturn (periods of 12 and 29 years). Progress is being made continuously, however, and objects with (minimum) masses of order 5–10 earth masses are now being picked up with some regularity. Note that the masses of most planets detected so far are minimum masses since the methods used mostly measures only one component of the velocity of the star (see below). Recently, even more exciting has been the technique of gravitational lensing [8]. In a first detection of a possibly "rocky" planet a 5.5 earth-mass object ("absolute mass") has been detected orbiting some astronomical units away from a M-dwarf star that is itself located several kiloparsecs away from the earth. Albeit that there are several sources of uncertainty in this observation, it shows the power of this method. Radial velocity recently detected two super Earths around an M star [9] at the edge of the HZ. This exciting discovery shows its ability to detect planets down to earth mass and makes it a prime candidate for first detection of worlds like our own.

So where do we go from here? The sky is literally the limit. The reasons for modern man to spend significant amounts of effort and money are multiple and important. It is of paramount importance today, as it has been all through history, for us to find man's place in the universe. Are we – as life forms – alone and unique in the Universe or is it inevitable that life arises when the conditions are right? In order to find this out we need to touch on several fundamental aspects of our understanding of nature. We need to find out if other worlds like our own exist; we must determine if conditions are similar to what has been the case on the earth historically and currently. We have to trace the evolution of exoplanetary systems and compare this with what we know about the solar system, and we need to actually look for the signatures of life – biomarkers – that can be traced over interstellar distances. This is a truly remarkable task.

While the ground-based techniques that have hitherto been used continue to be developed, it has, however, recently become more and more clear that to progress towards the direct detection and especially characterization of exoplanets, the size of our own world – with the ability to host life as we know it – will require telescopes and instruments to be deployed in outer space.

Recently, the scientific rationale given for exoplanetary missions has been put into the broader context of the European Space Agency's (ESA) new science plan for the period 2015–2025. This plan, designated Cosmic Vision, divides the major scientific questions to be addressed by European space science during the next few decades into four themes, the first of which is "What are the conditions for planet formation and the emergence of life?" Addressing the discovery and census of terrestrial planets around nearby stars, as well as a first determination of their physical parameters, including their habitability, is the challenging objective of Cosmic Vision theme 1. The same scientific issues appear also in other evaluations of highly prioritized issues for science in the near future. Most clearly, this has been seen in the decadal reviews commissioned by the space agency of the United States, the National Aeronautics and Space Administration (NASA). It thus appears very likely that significant effort will be spent over the next few decades in achieving these goals.

3.2
Formulation of the Problem and High-level Scientific Requirements

As far as we can tell, from our limited statistics of one, and our (still) very poor understanding of what life really is and how it forms, we need a rocky, planetary surface in order to expect the processes that have apparently taken place on our planet. It has, however, been hypothesized, e.g. by Carl Sagan, that life could originate, develop and continue to exist in the atmosphere of gas-giant planets like our Jupiter. This was described in Sagan's fascinating and epoch making TV series "Cosmos" from the beginning of the 1980s and in the accompanying book [10]. As far as we understand today none of the life forms we know on our planet could have done this, which is why the Galileo spacecraft was crashed into Jupiter's atmosphere in order to make it impossible for it to later contaminate the Jupiter moons, especially Europa, with any "hitchhiking" terrestrial life form. In order to make any progress in this area we need to understand correctly how planets form and evolve, and particularly how a solar system like our own came to be, and why so many of the systems found so far appear completely different from our own. This will mean that we will have to understand the physical processes involved. In turn, this will compel us to fully understand, at least empirically, the birth, evolution and death of stars in more detail, as well as having a much more thorough understanding of the fundamental physical processes involved.

In Fridlund [11], the high-level scientific requirements were described in the context of space missions, as being the answer to the following questions:

- Are we alone in the universe?
- How unique is the earth as a planet?
- How unique is life in the universe?

A feasibility study carried out by ESA between 1997 and 2000, in order to investigate if it was currently technically possible to address these questions led, after a number of detailed technical investigations, to what is currently being represented by the mission concept of a "nulling interferometer" [12] (see also below), which is designated "Darwin" and is being developed to carry out the following tasks:

1. Search a large number of nearby stars for terrestrial planets to find their frequency and location.
2. Detect planets within the so-called "Habitable Zone". In the Darwin studies, the HZ was considered only in terms of black body temperature. No provision was made to take into account atmospheric pressure, etc.
3. Determine the planet's orbital characteristics (period, eccentricity, inclination, etc.).
4. Observe the spectrum of the planet. Detection of the presence of an atmosphere, effective temperature and diameter of the planet (through the albedo).
5. Determine the composition of the atmosphere, namely, the presence of water, ozone/oxygen in an earth-type planet, mainly inert gases in a Mars/Venus type planet and hydrogen/methane atmospheres in Jupiter-type planets or "primordial" earth-like planets.

These high-level scientific requirements are translated into specific observational requirements that can be converted into mission requirements. The process is detailed in [13], but can be summarized briefly as:

1. Minimum number of single, solar type (F-K main sequence) stars to be surveyed for terrestrial exoplanets in the HZ during primary mission is to be 165 to 500. M-dwarfs, the most common stars in the galaxy are also considered.
    - 165 under the added condition that significant amounts of dust (10 times the level in the solar system) are present in every object.
    - 500 (this is essentially a complete sample of single F, G and K stars found out to 25 pc; added to this a number of M dwarfs) with the conditions of similar levels of dust as found in the solar system.
2. Completeness of survey (probability that one has not missed a planet in the HZ for a specific star) to be better than 90%.

The reasons for these mission-specific requirements are, of course, so that a negative answer (i.e., the *nondetection* of any earth-like world) is going to be meaningful in a scientific sense.

3. Spectral signatures to be observed for each planet in any detected exoplanetary system, e.g., the so-called bio-markers (see Kaltenegger this volume).

These are very demanding requirements and, as a consequence, the Darwin mission concept (see below) is arguably the most ambitious scientific space experiment ever contemplated. It is therefore apparent that a step-wise approach with simpler systems answering (fundamental) parts of the issues (such as, e.g., the frequency of earth-sized planets, etc.) could be implemented. This would not only save resources and funds, but it would also alleviate the risk when finally implementing the more complex systems.

3.3
Comparative Planetology

Comparative Planetology is currently rapidly acquiring the status of a new sub-discipline of Astronomy and Astrophysics. Although originating from within this subject, it is now becoming truly cross-disciplinary incorporating elements of biology (astrobiology), geophysics, chemistry, meteorology, etc., and thus also attracts interest from researchers in all these disciplines. Many of these individual scientists are of course attracted by the possibility that, through comparison, we will make advances in areas of research relating to our own planet and its existence that have been impossible or are very difficult to carry out here on Earth. An important example would be the origin of life itself on our planet. We do not know, either when, how or where this occurred on our planet. The evolving state of the planet's crust has obliterated direct evidence and, consequently, we do not know if life arose spontaneously and easily (as would be indicated by an early occurrence), or under very special circumstances (as may be indicated by the extremely long development time for complex life forms). Observing a large number of exo-planetary systems of a young age and finding (or not finding) signs of, e.g., chemical disequilibrium in their atmospheres, would allow us to draw some conclusions in this area.

When we use the term Comparative Planetology, we have to realize that this label has hitherto only been used for comparisons between objects within our own solar system, e.g., Venus and the earth. It is, however, now beginning to be used in the much broader sense of the comparison of different solar systems (and the components of the planet's comets and dust) with each other. This is where one benefits from the study of different systems, having different histories and evolution, (original) physical conditions, different formation processes, etc. This then allows making full use of empirical science to understand our own system more completely. When we compare planets within the solar system with each other, one makes a cross-section through time at the current status of objects which have had the same origin and have experienced the same evolution, and where the major differences between bodies are probably caused only or mostly by their different location within the system (the boundary conditions of the formation process, as it were). Expanding our knowledge to other stars and their accompanying planets eliminates these limitations. This is already evidenced by the amount of information acquired over the last 10 years since the first bona fide exoplanet was discovered [3]. Always keeping in mind that we are still limited by ground-based observational bias.

We have for a long time (essentially since the days of Copernicus) made a fundamental assumption that we soon will be able to prove or disprove. We have assumed that the earth is an average planet orbiting around an average star. But we do not really know this to be true. After searching for a true solar analogue for more than 50 years, with continuously more refined methods, and failing to find an exact copy, we realize that our sun is only average in the same way, as for instance, people are average. When we take the "finger prints" of stars, although we find very similar stars, we cannot find an identical solar analogue. Thus we must draw the conclusion that the sun is "special"as people are "special". What we still do not know is how important these small differences in chemistry and evolution are when investigating, e.g., how common small rocky planets like our own are, or if these differences would impact on either the formation process of different types of planets or on life's ability to arise on a planet with a slightly different chemical make-up or even on different kinds of planet's ability to continue to host life over a long period.

3.4
Methods and the Need to go into Space

The prime methods that have been used so far (from the ground) are:
1. The radial velocity method.
2. Occultation's of a star by a planetary body, transits.
3. Gravitational lensing.
4. Astrometry.

All of these methods have delivered tremendous results as can most clearly be seen from an inspection of *The Extrasolar Planets Encyclopaedia* web pages (http://exoplanets.eu) maintained by Jean Schneider at the Meudon Observatory in Paris. Nevertheless, each method has severe limitations imposed by being implemented from the surface of planet earth.

The first method; the radial velocity method, measures the deflection of the spectral lines in a stellar spectrum, caused by the gravitational tug imposed upon its surface by an orbiting body. It is of course a function of both the mass of the orbiting body and its distance from the stellar surface, and therefore the method is naturally biased towards (as pointed out by Struve in the above quoted paper from 1952) massive planets orbiting very close to the star. *It is interesting to notice that Struve also pointed out that there were no a priori reasons NOT to expect such planets, just because they are nonexistent in our own solar system.* As we want to detect planets like our own, we realize that an earth-mass body, orbiting 1 astronomical unit away from a G2V star (our sun), will cause a deflection (amplitude) in the radial velocity curve of 0.1 $\mathrm{m\,s^{-1}}$ over a period of one year. This is significantly smaller than the amplitude change caused by the five-minute acoustical oscillations in the solar atmosphere (so-called p-modes), which have lifetimes of roughly half a year. We also have the noise introduced by the solar activity, which has amplitudes that are similar or larger. If we search for an earth within the zone where we expect to find life, around smaller, less luminous, solar-type stars (K-dwarf stars), the situation is alleviated regarding the first two of these problems (larger planetary amplitude, p-modes likely to have lifetimes more different from orbital period, since orbital period is shorter and p-modes have longer lifetimes) but not in the third area (activity) where it may be actually somewhat worse. Taken together with the technical difficulties it is clear that, while not impossible, it will be hard to detect an earth analogue from the ground using this method. It will require dedicated large telescopes with long observing runs (years). Further, since we only measure one component (along the line of sight) of the velocity, the mass of the planet that we determine will be a *minimum* mass. Some other method (see below) can be used in special cases to determine the other component and therefore lead to an exact mass (only depending on the estimate of the stellar mass).

The method second in importance (so far); the transit method, has already delivered results in a number of cases (currently, September 2007, about 25 stars). This is also a method that shows strong promise of being important in space-based applications – see the CoRoT and Kepler mission descriptions below. Here we detect the drop in luminosity of the star as the planet passes between us and the star, and draw conclusions about both star and planet from the shape of the light curve. The problem here is of course that it is also biased towards large planets orbiting very close to a (small) star – a situation not found in our own solar system. A Jupiter-sized planet passing between us and a solar-type star will cause a drop in luminosity of about 1%, while an earth-sized body only causes a drop of about 10^{-4}. Further, transits last for some hours, which means that the shorter the period (the first detected transits of an exoplanet repeats every 3.5 days) the easier it is to detect it. Then we have to take into account the random orientation of exoplanetary orbital planes. This results in about a 1–2% chance of a transit happening for planets close in, while it is significantly smaller for planets orbiting in the Habitable Zone of a solar-type star. This means that the method needs to be applied either to very large samples of stars, simultaneously, or that one has very strong reasons to suspect that the planetary orbital plane is to cross the line of sight. This was the case for the first detected transit, that of HD 209458b [6]. Using wide-field telescopes, either from the ground (where an additional limitation is the earth's rotation which requires networks of telescopes) or from space, one can observe large numbers of stars at the same time. It is then a powerful method mostly limited by the disturbances in photometric precision induced by our atmosphere.

The advances that have been made through the detection of the gravitational lensing effect caused by large, and so far, in one case of a possible small, "rocky" planet of maybe 5.5 times the mass of our own Earth [8] are very promising. This method is no doubt going to be significantly developed over the next decade. It is a potentially very powerful method for determining statistics of exoplanets. The drawback is that, for this method to work, one needs both a lensed object as well as a lensing system. In order to have any significant chance of detecting planets of all types, both sets of objects need to be large, which means that they will also be very distant. We then have difficulty in assigning types and luminosities to both the lensed object and the lenser. The planetary mass will depend on both of these parameters. In the last case quoted above [8], we assume that the blended image of the lensed object close to the galactic centre and the lenser, at about half this distance, are of F and M type, respectively. We will not know this for many years as the star slowly moves out of the line of sight of the lensed object.

The fourth method; the astrometric method, where one measures the proper motion changes of a star as it travels across the sky, and interprets deviations from its predicted motion, has so far been less successful. A number of multi-epoch observations carried out during the last century – mainly with long focus refractors – reported the discovery of large planets around some nearby stars. One can mention 70 Ophiuchi, where Reuyl and Holmberg [14] found a 10 Jupiter-mass body orbiting the star. None of these objects have been confirmed today.

Finally, astrometric observations, utilizing mainly the fine guidance sensors on the Hubble space telescope, have been used to determine the deflection in the plane of the sky in a few cases. In the case of the planet Gliese 876b, the deflection is about 25 milli-arcseconds, which, taken together with a well determined parallax, leads to a planetary mass of 1.89 ± 0.34 Jupiter masses, the largest uncertainty being the assumption about the stellar mass [15]. Dedicated space missions, with the capability of determining proper motions of a few micro-arcseconds, will make systematic surveys for planets of sizes down to maybe 10 earth masses possible within the foreseeable future (see below).

Finally, we will say something about the capability of Extremely Large Telescopes (ELT's) to detect exoplanets directly. The problem here is to have a wave front arriving on the imaging detector of such a quality, that one can detect contrast differences of between 10^5 and 10^{10}. This should be possible within a distance from the optical centroid of the star that would be 0.5 arcseconds for Jupiter and 0.1 arcseconds for the earth, if viewed at a distance of 10 pc. The requirement that one needs to be able to detect the planet among the residual of the airy disk at these angular distances is daunting enough without adding the noise sources provided by the atmosphere and either the segmentation of very large telescope mirrors or, in the case of monolithic mirrors, irregularities in the mirror surface. Referring the reader to the paper by Chelli [16], we here only conclude that a telescope with 100 m diameter, and with a correction of the wave front arriving on the detector to a precision of 2–3 orders of magnitude better than what is currently achievable, is required to detect the earth, orbiting in the HZ around an early G-type star. The limiting distance for a search would then be 10 pc [16] to 18 pc [Gilmozzi, private communication]. Within 18 pc we find a total of 39 single G-type stars [17]. Most of these belong to the later sub-classes of the main-sequence G type, and thus have a smaller HZ which is more difficult to resolve. All other solar-type stars within the 18 pc limit will have their HZ at much smaller angular distances from the central object, and thus be significantly harder to detect. Add to this the problem that, when searching for a planet similar to our own hosting life as we know it, the spectral signatures to search for are also in our atmosphere – but enormously stronger.

While thus not excluding the possibility of the detection of planets similar to our earth from the ground, we may safely conclude that this is a very difficult task, comparable with the most complicated of space missions. Further, it appears that the scientific case is going to be limited. This being said it is, on the other hand also quite clear that the large (30 m–40 m) class ELT's currently being planned for operation beyond the year 2020 will make significant contributions to the studies of gas-giants – particularly those of the kind found in our own solar system. These objects will be bright and well separated from their primaries, and spectroscopy will not be a problem since the features to study are very different from those found in our atmosphere. Radial velocity observations and other indirect detections of small rocky planets will also be feasible.

3.5
Space Missions

The key problem in detecting planets orbiting other stars is the dual one of spatial resolution and dynamic range. The star will outshine the planet (if it is of the size of the earth) by a factor of 10^{10}–10^{11}, in the visual, and 10^{6}–10^{7} in the mid-infrared. The maximum separation between a star and its planet in the HZ will be approximately 0.1 arcsec and that is for a very limited sample of stars. It is clear that, in order to achieve these requirements from the ground, improvements in adaptive optics will have be made at the level of two to three orders of magnitude over that currently achievable. Further, the telescope itself would have to be at least one order of magnitude larger in diameter, and several orders of magnitude larger in size and weight of the mounting. Finally, the study of earth-like planets would be impaired by our own atmosphere and the thermal background of the environment.

Thus, if we want to address the specific scientific issues outlined in previous sections, we will need to go into space. This has become more and more apparent over the last 10 years through a large number of very ambitious studies and projects developed in Europe, The United States and Canada. As a consequence, a number of projects/missions of increasing level of complexity and scope have been and are being developed. The Canadian mission MOST has been in orbit for several years, CoRoT since Dec. 2006, others – Kepler, GAIA – are under construction or approaching launch, and the "flagships" – Darwin and/or TPF are being developed both scientifically and technically. We now briefly describe these missions one by one.

3.5.1
MOST – The First Step

MOST (Microvariability and Oscillations of Stars, Walker et al. [18]) is a suitcase sized mini telescope launched into orbit by the Canadian Space Agency (CSA) on a Russian Rockot (ex-ICBM) into a low polar orbit around the earth on June 30, 2003. It deploys a 15 cm telescope feeding two 1024 by 1024 pixel CCDs. It measures very high precision variability in white light. A number of target stars have already been searched for exoplanetary signals, with negative results so far. Although it is a small telescope and the maximum dwell-time on any target star is only 58 days, nevertheless, it has been possible to exclude, e.g., 2–5 earth radii planets in several orbits around the known occulting star HD 209458 (http://www.astro.ubc.ca/MOST/milestones.html/PressReleaseA see also above). MOST only observes its targets one by one, and thus does not have the multiplexing advantage of observing large numbers (up to hundreds of thousands) of stars like CoRoT and Kepler (see below).

3.5.2
CoRoT – The First True Exoplanetary Mission

While MOST is designed as an asteroseismological mission, that can also search for exoplanetary transits under some circumstances, CoRoT (Convection, Rotation and planetary Transits, Boisnard and Auvergne [19]) is designed from the beginning with such transits as one of the primary objectives, and with most of the observing time allocated to this goal. A mini-satellite from the PROTEUS series of small spacecraft initiated by the French space agency, CNES, it is part of a true international mission with contributions from the European Space Agency (ESA), Brazil, Belgium, Spain, Austria and Germany. It consist of an afocal telescope with a 45 cm main mirror and a field of view of 2.8 by 2.8 square degrees. There are 4 CCDs consisting of 2048 by 2048 pixels each. Two CCDs each are allocated to the two main objectives: stellar seismology and the search for terrestrial exoplanets. In the latter program, CoRoT is directed towards the same field of view for 150 consecutive days with up to 12 000 targets brighter than magnitude 15.5. A prism produces a short spectrum of each target and the data of the (pre-selected) transit targets are individually read out and transmitted to the ground. Having a short spectrum of each target allow us to discriminate between (flares of) activity and a true planetary transit. The size of the planet that can be discerned depends on how close it is primary star. In principle, if the star is in an orbit with a period shorter than a few days, planets the size of the earth can be detected. For longer periods, of more than a few tens of days, radii between 1.5 and 2 times our own planet are observable.

CoRoT was launched from Russia in late 2006. The orbit is a polar circular orbit of 896 km altitude. The nominal mission is $2^1/_2$ years with possibilities of extending the mission to more than double that. CoRoT will of course easily detect hot Jupiters. CoRoT also requires a major program of follow-up observations with ground based instruments. Among others, the HARPS instrument on the 3.6 m reflector of the European Southern Observatory (ESO) at La Silla in Chile, normally used for searches for radial velocity signatures of exoplanets in nearby stars [20], as well as ESA's 1m telescope in Tenerife, Spain, which is normally used for laser communication experiments with satellites, are going to be used. Add to this telescopes in France and Germany and the supporting program can be seen to be very ambitious and will also produce science in itself. With the radial velocities determined by these means, and taken together with the inclination of the planetary orbital plane from the actual measured light curve we will be able to determine the true mass (and not only the minimum mass which we get when we only have one part of the data like the radial velocity curve). Taken together with the diameter (which we again get from the shape of the light curve together with the diameter of the star – the latter parameter either measured with interferometers or by modeling), we can obtain the planet's average density. Further follow-up observations with large telescopes can provide information about the atmosphere of giant planets through spectra or colors.

A large program of supporting observations is also required in order to learn as much as possible about each host star (and the total number of stars during an extended mission could reach 100 000), with either spectroscopic observations or Johnson and Strömgren photometry or both.

3.5.3
Kepler – The Exploration Continues

CoRoT will be followed in 2009, or later, by the National Aeronautics and Space Administration's (NASA's) of the United States, Kepler mission. This is essentially a larger version of the former mission with a 0.95 m Schmidt telescope. This mission will be deployed in a so-called drift-away orbit, traveling along the earth's orbit while slowly drifting away (after half a year in orbit the spacecraft is several tenths of astronomical units away). This provides very benevolent circumstances for the mission, and allows its essentially uninterrupted, direction towards a dedicated star field for a period of no less than four years. Under these conditions, it now becomes possible to detect earth-sized planets orbiting within the Habitable Zones (1+ year orbital period for an early G-type star, about half a year for a late K-type object) of their primaries. Sometime around 2011–2012, these objects will be accessible for a detailed follow-up from the ground, along the same guidelines as mentioned above for CoRoT. We must remember, however, that we are dealing with objects many hundreds of parsecs away, and the direct detection of these planets will be highly unlikely – at least with presently envisaged techniques. Nevertheless, we will learn much from these missions, and one of the most important pieces of information will be the value of η_{Earth}, which designates the percentage of solar type stars that have planets like our own [21]. A first glimpse of this number will have been obtained by CoRoT, and will be very important when designing missions with the most ambitious goal of them all – that of directly detecting an earth-like planet at interstellar distances, and analyzing the light from it spectroscopically in order to determine its characteristics.

3.5.4
The Role of the Herschel Mission in the Search for Other Earths

When ESA's Far-InfraRed (FIR) space mission Herschel launches in 2008, it will – with its 3.5 m primary mirror – be for years to come, the world's largest space telescope. As such, it will play an important role in the fulfillment of the objectives outlined above.

Herschel has the unique capability of detecting cold dust of the temperatures found in the solar system's Kuiper–Edgeworth (KE) belt. For really nearby targets, within 15–20 pc, it will be possible, for the first time, to detect systems with amounts of dust similar to those found in the outer parts of our system. Distributed between about 10 AU and 100 AU, this dust component is probably being provided by collisions between cometary objects within this zone. Without a continuous re-supply through repeated collisions, and/or larger planetary bodies that lock the dust into resonances where it cannot be ejected through, e.g., radiative pressure or falling into the sun through the Poynting–Roberson effect, this dust component would disappear in a relatively short time. The detection of a dust component similar to our own would thus indicate the presence of a KE belt and/or large outer planets, and would be a clear indicator of which systems should be searched first when the capability of direct planetary detection becomes available (see below).

3.5.5
GAIA – The First Global Survey Instrument and the Cosmic Census

GAIA [22] is an ambitious mission, under development by ESA, and with the goal of charting a three-dimensional map of our galaxy. Building on the highly successful HIPPARCOS mission, GAIA will provide never before realized positional and radial velocity measurements of more than one billion stars in the Milky Way galaxy as well as stars in galaxies located in the Local Group, i.e., the cluster of galaxies of which the Milky Way and, e.g., the Andromeda galaxy are members. Launched in 2011 or 2012, and operating for five years, each of its three telescopes will observe the billion stars about 100 times, measuring its position with down to about 20 microarcseconds angular accuracy. At the same time a spectrometer, operating in a narrow band in the visual wavelength range, will obtain about 40 spectra of each object, and thus provide the velocity along the line of sight (see above).

Placed in an orbit around the sun–earth L2 libration point, it will be undisturbed by the presence of either the earth or the moon, and over one year can observe the whole sky many times.

GAIA will play an important role for exoplanetary research in two ways. First, it will provide a database of relatively large (about 15 earth mass or larger) planets around all detectable stars out to a distance of several hundred parsecs. Based on the knowledge of the exoplanets found since 1995, we can expect literally thousands of new objects. This will allow statistical investigations several orders of magnitude more accurately than at present.

In addition, GAIA will detect all unknown nearby (within 25 pc) stars of all spectral types. It is clear from the work on the target lists for missions like Darwin (see below) that we have a very poor knowledge about late K-type and M-dwarf stars in the intended sample. These stars are intrinsically weak, and therefore do not exist in the parallax databases (e.g., HIPPARCOS). Since GAIA will operate down to 23 magnitudes and weaker in the R-band, the parallaxes of several hundred microarcseconds will be detectable only 1½ year into the mission.

3.5.6
SIM – Planet Quest

SIM (Space Interferometer Mission) is NASA's version of GAIA. Using interferometric methods (GAIA has three separate telescopes spinning across the sky, measuring the distances between individual stars globally), the relative distance between objects are measured extremely accurately (perhaps a factor of 10 better than GAIA or close to 1 microarcsecond at best). Both parallaxes and proper motions can be measured in this way. The price you pay is that you only measure a fraction of the number of objects that GAIA will, and you have limited sensitivity, as well as significantly increased complexity and cost. On the other hand it is, in principle, possible to see the deflection in the motion across the sky caused by earth-sized planets, for a handful of the very closest stars [23].

Further, SIM will be the first interferometric system in space and, as such, will be a precursor to Darwin and TPF (see below) – the missions dedicated to direct observation of planets like our own. The technology of optical interferometry is in its early phases here on earth, and it would be extremely valuable to test at least elements of this technology in space before one launches gigantic systems like Darwin with all it entails.

SIM will also study a number of important galactic and extragalactic targets, where the highly accurate measurements will provide new insight into the physical mechanisms. The added name "Planet Quest", however, demonstrates the order of priority of the scientific targets.

3.6
Darwin and the Terrestrial Planet Finder(s) – Other Worlds with Life as we Know it

The direct detection of earth-like exoplanets orbiting nearby stars, and the characterization of such planets – particularly concerning their evolution, their atmospheres and their ability to host life as we know it – and thereby to understand how planets form and life emerges – is the ultimate goal of Theme 1 in the Cosmic Vision plan. The same is true for the long-term plans of the US National Aeronautics and Space Administration (NASA) as evidenced by the program suggested by their decadal committees. Thus both of the world's major space agencies have, as one of their most important goals, to place our solar system into context.

These space missions, capable of meeting the challenges posed by these goals are Darwin [11, 24, 25] (ESA) and the Terrestrial Planet Finder (NASA) [21, 26].

As mentioned above, the direct detection of a planet of the size of the earth, orbiting its parent star in the HZ, constitutes a challenging problem, since the signal detected from the planet is between about 10^{10-11} (visual wavelength range) and 10^{6-7} (mid-IR spectral range) times fainter than the signal received from the nearby star. Selecting the appropriate spectral region in which to attempt detection is governed by this contrast problem *and* the selection of a region in which the characterization of the planet and its habitability is optimum. This problem has recently been addressed by a number of researchers, e.g., Selsis [27] and Traub [28]. The European mission Darwin has selected the spectral region between 6 μm and 20 μm, a region that contains (among others) the CO_2, H_2O, CH_4, and the O_3 spectral features found in the terrestrial atmosphere. The presence or absence of these spectral features would indicate similarities or differences with respect to the atmospheres of known telluric planets such as Venus, Earth and Mars.

As outlined below, Darwin, which is the scenario for ESA's project, is based on interferometry, i.e., connecting a number of smaller telescopes spread over a larger area and thus achieving the same spatial resolution as would a telescope of the diameter equivalent to the longest baseline of the interferometer. NASA, has also investigated this possibility in their TPF (Terrestrial Planet Finder) concept. The results of their study are very similar to what ESA has found, and a close collaboration between the two agencies have resulted (see below). NASA has, however, also investigated another concept, namely coronagraphy. More specifically, this study, TPF-C, has focused on an apodized telescope with an elliptical mirror that measures 8 m × 3.5 m. Several different coronagraphic masks have been investigated. The main advantage of the coronagraph over the interferometer is of course that we are dealing with only one spacecraft. With the interferometer, we are depending on the development of autonomous precision formation flying in space, something hitherto untested. Formation flying is being aggressively pursued in Europe, however, with a first launch of the so-called PRISMA mission in 2008 (see www.prismasatellites.se). This is a Swedish–French–German mission utilizing technology developed for the Darwin program by ESA. This will likely be followed by more representative technology precursors in the beginning of the next decade (e.g., the French SIMBOL-X – see http://www.dapnia.cea.fr/Phocea/Vie_des_labos/Ast/).

Concerning the coronagraph (TPF-C), the difficulties are mainly caused by the very high precision of the final wave front that is required. An rms value of 0.3 Angstrom is needed, something about two orders of magnitude beyond current technology. Even with such a high precision, the search space is limited. Only about 30 of the closest stars can be searched in full (searching for bio-markers) for earth-like planets in their HZ. A further about 80 stars can be partly searched. Another problem is that the coronagraph represents a "dead end", i.e., it is unlikely that the technology will be used to construct larger telescopes in space. Contrariwise, it is believed that eventually very large interferometers, both visual and operating in the infrared, will have to be constructed in order to fulfill the scientific objectives of tomorrow.

3.6.1
Nulling Interferometry

For the detection and study of earth-like planets, ESA selected a so-called "Nulling Interferometer" in 1996. This instrument, operating by the principle of destructive interferometry, and in the mid-IR wavelength range, for detailed study and possible implementation as the Darwin mission. It utilizes free flying telescopes, i.e., each telescope on a separate satellite and thus no connected structures (Fridlund [11]). It thus implements the new technology of nulling (or destructive on-axis) interferometry, first suggested for this purpose by Bracewell [12] and Bracewell and McPhie [29].

The basic concept of such an interferometer is to sample the incoming wave front from the star and its planet(s) with several (≥ 2) telescopes that individually do not resolve the system. By applying suitable phase shifts between the different telescopes in this interferometer array, destructive interference is achieved on the optical axis of the system in the combined beam. At the same time, constructive interference is realized a (very) short distance away from the optical axis. Through the appropriate choice of configurations and distances one can, for the specific case, place areas of constructive interference on regions representative of the HZ and so achieve the required contrast at this location.

The first practical demonstration of "nulling", carried out from the ground, was undertaken by Hinz et al. [30]. Using the multiple mirror telescope on Mount Hopkins, Arizona, these authors were able to cancel out the image of the star, Orionis. The ability of the interferometer to suppress the entire Airy pattern was demonstrated and in this case the nulled image had a peak in intensity of 4.0% and a total integrated flux of 6.0% of the constructive image. Since, light suppression in excess of 10^5 has been achieved in laboratories both in Europe and the US.

The simplest nulling interferometer is the so-called Bracewell array [12], which consists of two telescopes whose output combines with a π phase shift. The beams from each of the telescopes are made to interfere destructively (the process known as nulling) on the optical axis. Thus light from a star placed on the optical axis of each of the telescopes and thus of the interferometer can be centered on a deep null, and effectively extinguished. Then, as mentioned above, light from a source which is separated from the optical axis by a small angular distance, θ, will arrive with a small external phase delay, $\theta B \cos\phi/\lambda$, where B is the distance between the two telescopes and θ is the azimuthal sky coordinate of the source.

In the absence of perturbations, the interferometric transmission of the Bracewell interferometer with two circular entrance apertures of diameter D for a point source at the sky coordinates (θ, ϕ) is given by the following equation:

$$T_B W(\theta, \phi, \lambda) = 2x \frac{J_1(\pi\theta D/\lambda)}{(\pi\theta D/\lambda)^2} \sin^2\left(\pi \frac{\theta B}{\lambda} \cos(\phi)\right) \tag{3.1}$$

where J_1 is the Bessel function. Equation (3.1) shows that the interferometer response is (for small angles) proportional to $(\theta B/\lambda)^2$. Because of the shape of the destructive interference fringe, as a function of position, we refer to this as a θ^2 null.

The use of a single Bracewell as a nulling interferometer for planet detection is limited by the difficulty of rapidly modulating the planet signal against disturbances such as stellar leakage and the background. For a system at 10 pc observed at 10 μm the local zodiacal cloud signal will be the biggest noise source. Second in importance is the stellar leakage followed by the signal of the exozodiacal dust disk (if present). Here we have assumed an exozodiacal dust disk similar to our local zodiacal cloud. The planetary signal will be hidden in that background and fast modulation is required in order to detect a potential planet. Modulation could be achieved by rotation of an interferometric array. The problem with such an approach will be that such a modulation will be necessarily slow because of limitations in operating the spacecraft. The modulated planet signal could thus easily be corrupted by introducing drifts leading to $1/f$ noise. An alternative modulation technique is internal modulation, which is based on combining the outputs of various sub-arrays e.g., two Bracewell arrays, and implementing a variable phase shift [31].

Configurations based on a number of Bracewell pairs, arranged in certain patterns will have destructive responses with a θ^4 or θ^6 null and thus a wider and a better suppression of excursions from the on-axis flux of, e.g., the resolved stellar disk. This thus leads to a lower stellar leakage. While an ideal interferometer with a π phase shift would provide 100% extinction at an infinitesimal bandwidth, in reality, the nulling "band" will have a thickness, and a depth and shape depending on B, θ and the configuration (including the number of telescopes making up the arrangement). As seen by this reasoning, due to the size of the (resolved) stellar disk a residual stellar signal S_{leak}

will pass through the central transmission pattern. This leak is calculated according to Eq. (3.2), and the shape of the null for two different configurations (representative of θ^2 and θ^4 null profiles) can be seen in Fig. 3.1.

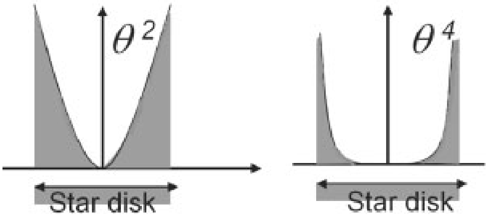


Fig. 3.1 Different shape of the central null and rejection of light from the stellar disk for the θ^2 and θ^2 configurations (from Kaltenegger et al. [32]).

The sensitivity pattern plotted onto the sky, the transmission map, is most easily visualized by imagining the transmission pattern in Fig. 3.2 to be rotated by 2 π out of the plane of the paper.

$$S_{\text{leak}}(\lambda) = A_{\text{tot}} \int_{\theta} \int_{\phi} T(\lambda, \theta, \phi) B_{\text{star}} \, d\theta \, d\phi \tag{3.2}$$

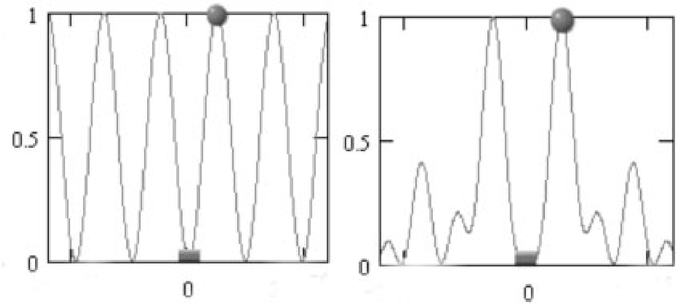


Fig. 3.2 Transmission pattern of a configuration with a θ^2 [12] and a θ^4 (the Bowtie configuration as suggested by Absil [33]) geometry null. A radius is displayed from the center (photocenter of the star) out to a radius of 5 AU. Both cases are assumed to be observed at a distance from the earth of 10 pc. The stellar disk is magnified and a planet is shown at 1 AU (gray dot) (from Kaltenegger et al. [32]).

In practice, the central null of the transmission map will be degraded by amplitude and optical path differences between the interferometer arms that will lead to a level of noise which sets a limit for the achievable null depth given. Averaging over the detector readout time leads to Eq. (3.3).

$$\frac{\langle F_{\text{leak}} \rangle}{F_S} \approx \frac{\alpha^2}{4} + \frac{4\Delta\phi\alpha}{3\pi} + \delta_A^2 + \frac{\Delta\phi^2}{4} \tag{3.3}$$

Here $\alpha = \pi\theta_S B/\lambda$. The $\Delta\phi$ term describes the wavelength -dependent dispersion effects while σ_A is the root-mean-square fluctuation of the relative amplitude between the two arms of the interferometer after spatial filtering. The leakage term and the resulting null degradation depend strongly on the baseline used. Optimizing the baseline of the interferometer for each stellar target system will minimize the leakage, while a fixed baseline such as would be used for an interferometer implemented on a structure, would lead to a high level of leakage for most of the target systems. In the input catalogue for searches around nearby stars, distances vary between 3 pc and 25 pc [17]. It would therefore be necessary, in the case of a structurally mounted interferometer, to implement a mechanism that allows for a variation of optimized baselines. Clearly this optimization of the leakage strongly supports an interferometer deployed on free-flying spacecrafts.

The stellar leakage is, however, not the dominant noise source for stars with distances larger than about 10 pc. Instead, local and exo-zodi will dominate. This means that only the integration time for the closest stars are negatively affected by a reduction of the null width from a θ^4 to θ^2 profile [34–36]). The number of single, accessible solar-type stars within 10 pc is relatively small, a total of 12 stars excluding 46 M-dwarfs of the, in total, more than 1135 candidates in the target star catalogue out to 25 pc. Of these 628 are considered single and 241 of the latter are classified as M-dwarfs [17]. The majority of targets are thus found at a distance where the local zodiacal cloud is instead the dominant noise factor. The original Darwin mission concept [11] minimized stellar rejection in order to optimize the search for earth-like planets orbiting the very closest target stars. This led to a baseline configuration with six free-flying telescope spacecraft and a central beam combiner, as originally suggested by Jean-Marie Mariotti, and later modified and developed by Olivier Absil [33].

The fact that we are now considering configurations optimized for the much larger sample of stars at distances greater than 10 pc does not mean that the nearby stars have been discarded. Since, in the case of the nearby objects, the stellar leakage is greater it will take a longer time to acquire the required signal-to-noise. The planets will have an intrinsically higher flux, which in most cases will compensate for the increased noise.

Because of this realization, recent candidate Darwin and TPF configurations tend to use three or four telescopes (Fig. 3.3) significantly reducing the complexity and cost of the mission. Three telescopes is the minimum number of telescopes needed for an interferometer mission that uses rapid signal modu-

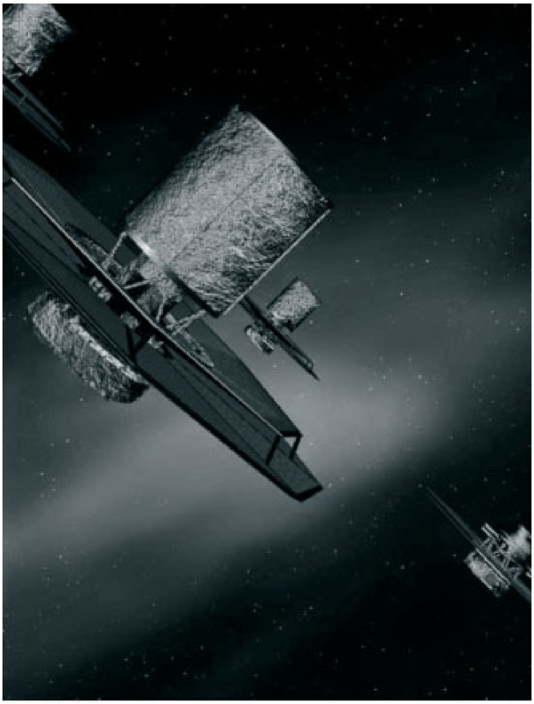


Fig. 3.3 Artist's impression of a four-telescope Darwin mission with the spacecraft carrying the telescopes configured in a X-shaped pattern and with a beam combiner spacecraft flying in the middle of the array. The telescopes in this example are equipped with 3.15 m mirrors based on the technology used for ESA's Herschel mission. Courtesy of ESA. Illustration by Medialab.

lation with sub-interferometers [31] to detect a planet in the high background noise.

One example is the so-called stretched X-array, which consists of four telescopes in an X-shape with a beam combiner in the centre. By "stretching" the X in one direction we create several sub-interferometers, of which the output of each can be combined or subtracted from each other, thus creating a modulation in the transmission pattern in the sky, which effectively removes the background (which is significantly stronger than the planetary signal). Recent work suggests that requirements on the shape of the null can be relaxed when one takes into account unavoidable noise contributions introduced by instrumental errors [34, 35]. Thus configurations with fewer telescopes (three or four as compared with the original five or six) are now investigated at system level as candidates for the Darwin mission reducing the complexity and cost of the mission.

3.6.2
Background and Foreground Flux

The detected signal-to-noise depends on the contribution of the separate background and foreground noise terms, to the interferometer output signal. These noise terms can be divided into:

- Instrumental noise, such as thermal background from the telescopes and their components, and other constant and variable noise sources caused by imperfections in the system.
- Noise sources caused by other astronomical objects.

Excluding the possibility that a background (e.g., a galaxy) or foreground (e.g., a solar system) IR source is found to be falling within the field of view of the synthesized beam, these noise sources consist mainly of the solar system zodiacal light caused by interplanetary dust and the exozodiacal emission caused by similar dust that can be found in the target systems. The local zodiacal cloud (LZ) provides the foreground through which Darwin will have to observe. Because this zodiacal foreground is diffuse, it cannot be cancelled by the nulling interferometry. Thermal emission of the exozodiacal (EZ) dust will be the strongest source of noise photons at the shorter wavelengths. Currently, the range of EZ expected from target systems is unknown. It is only very recently, that observations with NASA's Spitzer infrared observatory have detected Kuiper–Edgeworth belt (K-E belt) dust around normal main sequence stars [37, 38]. This dust, although similar in composition is located much farther out both in the solar system and in the target systems. In the solar system the K–E belt dust originates outside Saturn where it is presumably created by collisions between the aggregate of bodies orbiting there. Kept into place temporarily by resonances it will, on timescales of thousands of years, disappear through either the Poynting–Robertson effect or through photon pressure. Any observed dust must thus be continuously replenished, and would imply a significant repository of cometary or asteroidal bodies. In one case (so far) Spitzer sees warm dust, which thus originates at most a few AU from the central star, and no cold (K–E belt) dust around a late K-type star [39]. In this case the amount of dust is in excess of 1000 times that found in our solar system. Spitzer, is anyway limited to observe systems with about 100 times the amount of dust (both in the EZ and in the K–E regions) than that found in our solar system.

The level of EZ dust will determine the efficiency with which the missions like Darwin/TPF will be able to discern an earth-like planet. As we do not know in what fraction of stars the dust levels will be prohibitively high, for the calculations carried out during the development, one has assumed a range within which it will be possible to detect terrestrial planets. The assumption for the "standard case" is 1 EZ (being equal to the solar system value), and a "worst case scenario" of 10 EZ. It would be possible, in principle, to detect the earth against a background of 30 EZ, but this would come at the price of very long integration times. The LZ and EZ emission, within the mid-IR wavelength band, diminishes with the distance from the sun and from the central star, respectively. This is due to both decreased temperature and decreased dust density. Note that large coherent structures, such as clumps behind planets, can masquerade as planets. Also "wakes" caused by the passing of a (forming) planet through the dust is, in principle, detectable. These structures could also serve as markers for the presence of planets, if their location with respect to a planet were well understood. Exozodiacal clouds are thus not expected to be uniform, and a planet must be detected against a nonflat field of corrugations. In our own cloud, these "clumps" are assumed to have roughly less than 0.1% of the amplitude of the total cloud brightness and are thus no source of confusion.

The stellar leakage is the dominating astronomical noise source for the shorter wavelengths within the mid-IR wavelength band, while the thermal background level due to the emission of the optics, can approach that of the zodiacal dust for temperatures above 40 K at the longer wavelengths. This then provides a strong requirement that the temperature of the optics of the interferometer is less than 40 K. However, the LZ dust disk remains the biggest noise factor, for most of the wavelength band.

The analytical model using the EZ in our calculation builds on the one by Kelsall et al. [40] which describes our own LZ cloud. This model is based on observations of the solar system made by the Diffuse InfraRed Background Experiment (DIRBE) aboard the COBE satellite (COsmic Background Experiment). The zodiacal cloud in our solar system extends outward at least to the asteroid belt, thus roughly 3.5 AU from the sun, and inwards to the solar corona, at a few solar radii. As an inner cutoff for the exozodiacal cloud we use a sublimation temperature of 1500 K. The flux and temperature is then:

$$B_{\mathrm{EZ}} = \int B(T(r))\rho_0 r^{-\alpha} e^{-\beta \frac{(}{z}\gamma)^{\gamma}} dl \tag{3.4}$$

with $\alpha = 1.39 \quad \beta = 3.26 \quad \gamma = 1.02 \quad \rho_0 = 1.1410^{-7}$

$$T(r) = T_0 r^{-}\delta \left(\frac{L}{L_{\mathrm{sun}}} \right)^{\xi} \tag{3.5}$$

with $\delta = 0.42 \quad T_0 = 286 \quad \xi = 0.234$

This LZ disk model is then used as input for the exosystems emission, scaling the number density of the dust as a free parameter, and changing the values of the stellar luminosity according to the spectral type.

3.6.3
Model of an Exosolar System

In order to calculate the throughput of a nulling interferometer we need a model of the exosystem and the noise sources. We can assume a target system consisting of one or multiple planets, an exozodiacal dust disk and a host star. Obviously also the local zodiacal dust disk and the thermal radiation of the instrument at 40 K have to be accounted for. Following Kaltenegger et al. [32], the nulled output signal $F(t)$ for an extended source of brightness distribution $B_{\text{disk}}(\theta, \phi)$ that is transmitted through the interferometer pattern, can be determined using Eq. 3.4.

$$F(t) = A_{\text{tot}} \int_{\lambda} \int_{\phi} \int_{\theta} B_{\text{disk}}(\theta, \phi) T(\theta, \phi, \lambda, t) \theta \, d\theta \, d\phi \, d\lambda \tag{3.6}$$

where A_{tot} is the total collecting area of the interferometer, and T is the normalized response of the interferometer, the so-called transmission. The instrumental transmission is wavelength dependent and therefore the baseline of the array has to be optimized such that the response at the assumed planet position, $T_{\text{planet}}(\lambda, t)$, is maximized over as wide a part of the wavelength range as possible. In an alternative scenario the array can be reconfigured during the observation in order to provide a high transmission of the planetary signal over several selected wavebands in turn. When sub-interferometers are used to modulate the planetary signal, the time-averaged interferometer response is called the Modulation $M(\theta, \phi, \lambda, t)$ and can be used instead of the $T(\theta, \phi, \lambda, t)$ in the calculations. The star and the planet are assumed to be represented by uniform disks of angular radius θ_S and θ_{planet} respectively and use Planck's equation to obtain the brightness distribution B_{star} and B_{planet} as a function of wavelength. $T_{\text{planet}}(\lambda, t)$ denotes the response of the array at the planet's position over wavelength and time with a baseline optimized for its detection. When the planet search is performed by rotating the array, the mean value of the transmission at that radial distance within the field-of-view can be used in a first approximation to determine the signal strength. The wavelength dependence of the interferometer's response allows the extraction of basic color information for the detected planet. This technique would rapidly distinguish between giant planets and terrestrial ("rocky") planets during the detection phase of the mission. We thus have the following set of equations contributing to the model [32].

$$F_{\text{planet}}(t) = A_{\text{tot}} \int_{\lambda} B_{\text{planet}}(\lambda) T_{\text{planet}}(\lambda, t) \pi \theta_{\text{planet}}^2 \, d\lambda \tag{3.7}$$

$$F_{\text{leak}}(t) = A_{\text{tot}} \int_{\lambda} \int_{\phi} \int_{\theta} B_{\text{star}}(\theta, \phi, \lambda) T(\theta, \phi, \lambda, t) \theta \, d\theta \, d\phi \, d\lambda + \langle F_{\text{leak}} \rangle \tag{3.8}$$

$$F_{\text{EZ}}(t) = A_{\text{tot}} \int_{\lambda} \int_{\phi} \int_{\theta} B_{\text{EZ}}(\theta, \phi, \lambda) T(\theta, \phi, \lambda, t) \theta \, d\theta \, d\phi \, d\lambda \tag{3.9}$$

$$F_{\mathrm{LZ}}(t) = A_{\mathrm{tot}} \int_{\lambda} B_{\mathrm{LZ}}(\theta, \phi, \lambda) \int_{\phi} \int_{S\Omega} T(\theta, \phi, \lambda, t)\theta \, d\theta \, d\phi \, d\lambda \tag{3.10}$$

$$F_{\mathrm{thermal}}(t) = A_{\mathrm{tot}} \int_{\lambda} B_{\mathrm{thermal}}(\theta, \phi, \lambda) \int_{\phi} \int_{S\Omega} T(\theta, \phi, \lambda, t)\theta \, d\theta \, d\phi \, d\lambda \tag{3.11}$$

Only the flux of the exozodiacal dust disk F_{EZ}, the star F_{leak} and the planet F_{planet} are transmitted through the interferometer pattern. The flux from the local zodiacal dust disk F_{LZ} as well as the thermal background flux, F_{thermal} do not interfere destructively.

Spatial filtering of the high-frequency wavefront errors by monomode fibers or waveguides is envisioned for Darwin [36] and other types of nulling interferometers. This will limit the field-of-view of the synthesized beam. The geometric extent of the beam seen by such a fiber or waveguide is given by $S\Omega \approx \lambda^2$. Here S is the geometrical cross section of the optical fiber, Ω is the solid angle of the input beam and the product is known as the geometrical entendue. The IR thermal background as well as the LZ contribution is proportional to $S\Omega$ while the signals in the field of view of the interferometer like the planet and the EZ are proportional to the interferometer's collecting area. Note that the change of the coupling of the light from different off-axis angles into the fiber over the field-of- view must be approximated by a constant factor when calculating the throughput.

3.7
The Future – Mission Accomplished?

Both ESA and NASA have performed an ambitious technology development program with some collaboration (on both scientific issues such as the definition of the scientific case, and on more mission-oriented issues such as the optimum configuration). In these programs, essentially every one of the technological challenge has been addressed – at least to such a level of development that the ultimate success appears possible: achromatic phase shifters (to control the required path lengths in the interferometer to the very high precision required over the complete spectral band); optical fibers operating within the required band (to perform the required spatial filtering relaxing the requirements on the optical surfaces with 2–3 orders of magnitude); and integrated optics ("optics on a chip", i.e., mixing and splitting of signals on an integrated chip instead of with bulk optics, and with no moving parts – something with a potential to reduce mass and complexity).

Darwin and TPF-I have been developed in parallel over the last ten years. Already in 1998, an agreement has been made between ESA and NASA that information should be exchanged with the ultimate goal of a joining of the two missions. The input star catalogues for the two mission concepts are essentially identical. This is of course not surprising since the number of nearby stars is relatively limited (about 1000 single stars to a distance of 25 pc). Being very complex and thus probably also expensive, it makes very good sense to join the two missions since very little (except safety from catastrophic failure) could be gained by having two identical systems observing the same, relatively few, targets.

The results of space missions like those described above, especially Darwin/TPF perhaps like systems will no doubt impact on a number of scientific disciplines, and change them beyond recognition. The successful results will allow us to address questions within these subjects about not only how common our system is, but also will indicate its history and how our planet (and the life on it) have evolved, and in what direction our planet and its geo/biosphere will continue to evolve. Among the 1000 closest stars, we have a nice randomized distribution of ages of stellar systems spanning the range from approximately 10 million to about 10 billion years.

In particular, the successful detection of "true" earth analogues will of course influence our understanding of how life itself arose and evolved on our planet. Empirical evidence, long ago destroyed by our planets dynamic surface, can be observed on other worlds – ultimately in a time-lapse movie fashion.

Eventually, more detailed observations will be required. It is clear that instruments more complex and ambitious than have been discussed in this chapter will not be seriously considered before we know if a "cousin" or more exist in the vicinity of the sun. The fascinating aspect of our times is that this "first step" is being contemplated right now. Success in this endeavor will lead to the design of larger systems (such as "hypertelescopes") currently only in their first experimental phase [41], but involving maybe literally dozens of 10 m class space telescopes flying in formation with inter-satellite distances of several hundred km, and with the focal plane up to perhaps 5000 km away from the telescope units. Such an instrument could take "true" pictures of the surface of the earth at distances of 10–20 pc, with resolutions of 10–50 elements and using white light. It would then become possible to draw proper maps with features like clouds, land, sea, mountains and vegetation. It is possible that, within the foreseeable future, it will be possible to analyze the kind of life that exists outside of the solar system by, e.g., studying the so-called "red edge" of vegetation at about 720 nm wavelength and also to study the evolution of an earth over geological time [42]. Alternatively, polarization studies of terrestrial exoplanets could provide information about the chirality of molecules on their surfaces. It would demand gigantic space installations like the hypertelescopes described above, and the costs would be truly enormous – but it is possible. We would thus make another leap forward in our understanding of our own place in the cosmos.

We do truly live in exciting times.

References

1 Struve, O.: 1952, *The Observatory*, **72**, 199.

2 Latham, D.W., Stefanik, R.P., Mazeh, T. Mayor M., and Burki, G.: 1989, *Nature*, **339**, 38.

3 Mayor M. and Queloz, D.: 1995, *Nature*, **378**, 355.

4 Butler, P. and Marcy, G.: 1996, *ApJ L*, **464**, L153.

5 Gray, D.F.: 1997, *Nature*, **385**, 795.

6 Charbonneau, D., Brown, T.M., Latham, D.W., and Mayor, M.: 2000, *ApJ*, **529**, L45.

7 Raymond, S.N., Mandell, A.M., and Sigurdsson, S.: 2006, *Science*, **313**, 1413.

8 Beaulieu, J.-P., Bennet, D., et al.: 2006, *Nature*, **439**, 437.

9 Udry, S, Bonfils, X, Delfosse, X, Forveille, T, Mayor, M, Perrier, C, Bouchy, F, Lovis, C., Pepe, F., Queloz, D, and Bertaux, J.-L.: 2007, *A&A*, **469**, L43–L47.

10 Sagan, C.: 1980, *COSMOS*, Random House Inc., ISBN 0-394-50294-9.

11 Fridlund, M.: 2000, DARWIN The InfraRed Space Interferometer(Leiden: ESA), *ESA-SCI(2000)*, **12**, 47.

12 Bracewell, R.N.: 1978, *Nature*, **274**, 780.

13 Fridlund, M. and Kaltenegger, L., et al.: 2007 *A&A*, in preparation.

14 Reuyl, D. and Holmberg, E.:1943, *ApJ*, **97**, 41.

15 Benedict, G.F., et al.: 2002, *ApJ*, **581**, L115.

16 Chelli, A.: 2005, *A&A*, **441**, 1205.

17 Kaltenegger, L., Eiroa, C., Stankov and A.-M., Fridlund, M.: 2007, *A&A*, submitted.

18 Walker, G., Matthews, J., Kuschnig, R., Johnson, R., et al.: 2003, *PASP*, **115**, 1023.

19 Boisnard, L. and Auvergne, M.: 2007, The CoRoT Mission – Pre-Launch Status, *ESA SP*, **1306**,(in press).

20 Pepe, F., Mayor, M., Rupprecht, G., et al.: 2002, *The Messenger*, **110**, 9.

21 Beichman, C. A.: 2000, Darwin and astronomy, *ESA SP*, **451**, 239.

22 Perryman, M.A.C.: 2002, *ApSS*, **280**, 1.

23 Ford, E. B. and Tremaine, S.: 2003, *PASP*, **115**, 117.

24 Leger, A., Mariotti, J.M., Mennesson, B., Ollivier, M., et al.: 1996, *Icarus*, **123**, 249.

25 Leger, A., Mariotti, J.M., Mennesson, B., Ollivier, M., et al.: 1996, *Astrophysics Space Sciences*, **241**, 135.

26 Traub, W.A., Levine, M., Shaklan, S. et al.: 2006, *Proc. SPIE*, **6268**, 26.

27 Selsis, F.: 2002, *Proc. First European Workshop on Exo-Astrobiology*, *ESA SP*, **518**, Noordwijk, The Netherlands, 365.

28 Traub, W.A.: 2003, *Proc. Towards Other Earths: DARWIN/TPF and the Search for Extrasolar Terrestrial Planets*, *ESA SP*, **539**, Noordwijk, The Netherlands) 231.

29 Bracewell, R.N. and McPhie, R.H.: 1979, *Icarus*, **38**, 136.

30 Hinz, P.M., Angel, J.R.P., Hoffman, W.F., McCarth et al.: 1998, *Nature*, **395**, 251.

31 Absil, O., Karlsson, A., and Kaltenegger, L.: 2003, *Proc. SPIE*, **4852**, 431.

32 Kaltenegger, L., Fridlund, M., Karlsson, A.: 2006, *Advance in Space Research*, **arXiv:astro-ph/0606762**.

33 Absil, O.: 2001, *Proc. Nulling interferometry with IRSI-Darwin*, Liege International Astrophysical Colloquia, **36**, 79.

34 Dubovitsky, S. and Lay, O.P.: 2004, *Proc. SPIE*, **5491**, 284.

35 Kaltenegger, L. and Karlsson, A.: 2004, *Proc. SPIE*, **5491**, 275.

36 Karlsson, A., Wallner, O., Perdigues Armengol, J.M., and Absil, O.: 2004, *Proc. SPIE*, **5491**, 798.

37 Kim, J.S., Hines, D.C., Backman, D.E., Hillenbrand, L.A., et al.: 2005, *ApJ*, **632**, 659.

38 Bryden, G., Beichman, C.A., Trilling, D.E., and Rieke, G.H., et al.: 2006, *ApJ*, **636**, 1098.

39 Beichman, C.A., Bryden, G., Gautier, T.N., Stapelfeldt, K.R., et al.: 2005, *ApJ*, **626**, 1061.

40 Kelsall, T., Weiland, J.L., Franz, B.A., Reach, W.T., et al.: 1998, *ApJ*, **508**, 44.

41 Le Coroller, H. et al.: 2004, *A&A*, **426**, 721.

42 Kaltenegger, L., Traub, W.A., and Jucks, K.: 2006, *ApJ*, **arXiv:astro-ph/0609398**.

4
Biomarkers Set in Context

Lisa Kaltenegger and Franck Selsis

Abstract

In a famous paper, Sagan et al. [1] analyzed a spectrum of the earth taken by the Galileo probe, searching for signatures of life. They concluded that the large amount of O_2 and the simultaneous presence of CH_4 traces are strongly suggestive of biology. The detection of a widespread red-absorbing pigment with no likely mineral origin, supports the hypothesis of biophotosynthesis. The search for signs of life on possibly very different planets implies that we need to gather as much information as possible in order to understand how the observed atmosphere works physically and chemically. The earth–sun intensity ratio is about 10^{-7} in the thermal infrared ($\sim$ 10 μm), and about 10^{-10} in the visible ($\sim$ 0.5 μm). The interferometric systems suggested for Darwin and the Terrestrial Planet Finder Interferometer (TPF-I) mission operates in the mid-IR (5–20 μm), the coronagraph suggested for Terrestrial Planet Finder Coronagraph (TPF-C) in the visible (0.5–1 μm). For the former it is thus the thermal emission emanating from the planet that is detected and analyzed while for the later the reflected stellar flux is measured. The spectrum of the planet can contain signatures of atmospheric species that are important for habitability, like CO_2 and H_2O, or result from biological activity (O_2, O_3, CH_4, and N_2O). Both spectral ranges will indicate similarities or differences in the atmospheres of terrestrial planets and are discussed in detail and set into context with the physical characteristics of a planet in this chapter.

4.1
Introduction

Over 260 giant exoplanets have already been detected, and hundreds, perhaps thousands more, are anticipated in the coming years. The detection and characterization of these exoplanets will begin to fill in a gap in the astrophysical description of the universe; the chain of events between the first stages of star formation and the evident mature planetary systems. The nature of these

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

planets, including their orbits, masses, sizes, constituents, and the likelihood that life could develop on them, can be probed by a combination of observations and modeling. So far, our main detection methods from the ground; Radial Velocity and Transit Search, are biased towards big planets orbiting close to their parent star because they are easier to detect. Those planets produce a bigger, more frequent signal than small planets further away from their parent star. Even so, the number of smaller Extrasolar Giant Planets (EGP) found, indicates a trend towards smaller masses. The present stage of exoplanet observations can be characterized as one in which information is being gathered principally by indirect means, whereby the photons that we measure are from the star itself, or a background star, or a mixture of the star and planet. Indirect techniques include radial velocity, micro-lensing, transits, and astrometry. These indirect observations are of great value, giving us measures of the planet mass, orbital elements, and (for transits) the sizes as well as indications of the constituents of the extreme upper atmospheres, like the detection of sodium in the upper atmosphere of HD209458b [2]. Recent detection of super earths by Udry et al. [3], Rivera et al. [4] and Beaulieu et al. [5] imply that earth-mass planets might be common [6]. Current and future space missions like CoRoT (CNES, [7]) and Kepler (NASA), [8] will give us statistics on the number, size, period and orbital distance of planets, extending to terrestrial planets on the lower mass-range end. In the next stage of exoplanet observations, we hope to have direct observations, in which most of the measured photons are reflected or emitted by the planet itself. A fundamental part of the problem of directly detecting the planet with its feeble light in the glare of the strong parental stellar flux, is the huge contrast (see Fridlund, this volume). Direct techniques include coronagraphic imaging at visible wavelengths, and interferometric imaging in the thermal infrared. With direct photons in the visible and thermal infrared wavelength band, and depending on the signal-to-noise ratio, we can characterize a planet in terms of its size, albedo, its atmospheric gas constituents, total atmospheric column density, clouds, surface properties, land and ocean areas and general habitability. As discussed in Traub et al. [9], full characterization requires the synergy of both direct and indirect measurements. Direct detection of photons from giant exoplanets can be implemented using current space-based telescopes like HST and Spitzer. Such studies have led to the detection of infrared emission from several transiting hot Jupiters (see, e.g., [10–12]) where the planetary signal is the difference between the flux from a star plus planet versus the flux from the star alone. Photons from earth-like planets in the habitable zone (HZ) (see [13]) around their parent star are beyond the capabilities of these telescopes and require future missions like Darwin and Terrestrial Planet Finder (TPF). In this chapter we discuss the biomarkers at different wavelengths and the potential of each signatures in Section 4.2. In Section 4.2 to 4.8 we focus on what makes a habitable planet using earth as our example, and discuss surface, clouds and biosignature evolution over geological time. Section 4.9 concentrates on planets around different stars, Section 4.10 focuses on abiotic sources of potential biosignatures and Section 4.11 and 4.12 show how to interpret biosignatures by setting measurements in context with the physical characteristics of a planet. Section 4.13 summarizes the chapter.

4.2
Biomarkers

Biomarkers (or biosignature) is used here to mean detectable species, or set of species, whose presence at significant abundance strongly suggests a biological origin [14]. This is, for instance, the case for the couple $CH_4 + O_2$. *Bio-indicators* are indicative of biological processes but can also be produced abiotically in significant quantities. Our search for signs of life is based on the assumption that extraterrestrial life shares fundamental characteristics with life on earth, in that it requires liquid water as a solvent and has a carbon-based chemistry [14, 15]. Life on the basis of a different chemistry is not considered here because the vast possible life-forms produce signatures in their atmosphere that are so far unknown. Therefore, we assume that extraterrestrial life is similar to life on earth in its use of the same input and output gases, that it exists out of thermodynamic equilibrium, and that it has analogs to bacteria, plants, and animals on earth [16].

The first step of exploration of terrestrial extrasolar planets will be a space mission that can detect and record-low resolution spectra (see Fig. 4.1) of extrasolar planets like Darwin and Terrestrial Planet Finder (TPF). O_2, O_3, CH_4 are good biomarker candidates that can be detected by a low-resolution (Resolution < 50) spectrograph. There are good biogeochemical and thermodynamic reasons for believing that these gases should be ubiquitous byproducts of carbon-based biochemistry, even if the details of alien biochemistry are significantly different from the biochemistry on earth. Note that life can also exist without producing either O_2 or CH_4. Even so, we need to understand the abiotic sources of biomarkers better, so that we can identify when it might constitute a false positive for life detection, when abiotic sources could produce high quantities of a species we understand as a biomarker on earth. The theoretical modeling research goals are to explore the plausible range of habitable planets and to improve our understanding of the detectable ways in which life modifies a planet on a global scale.

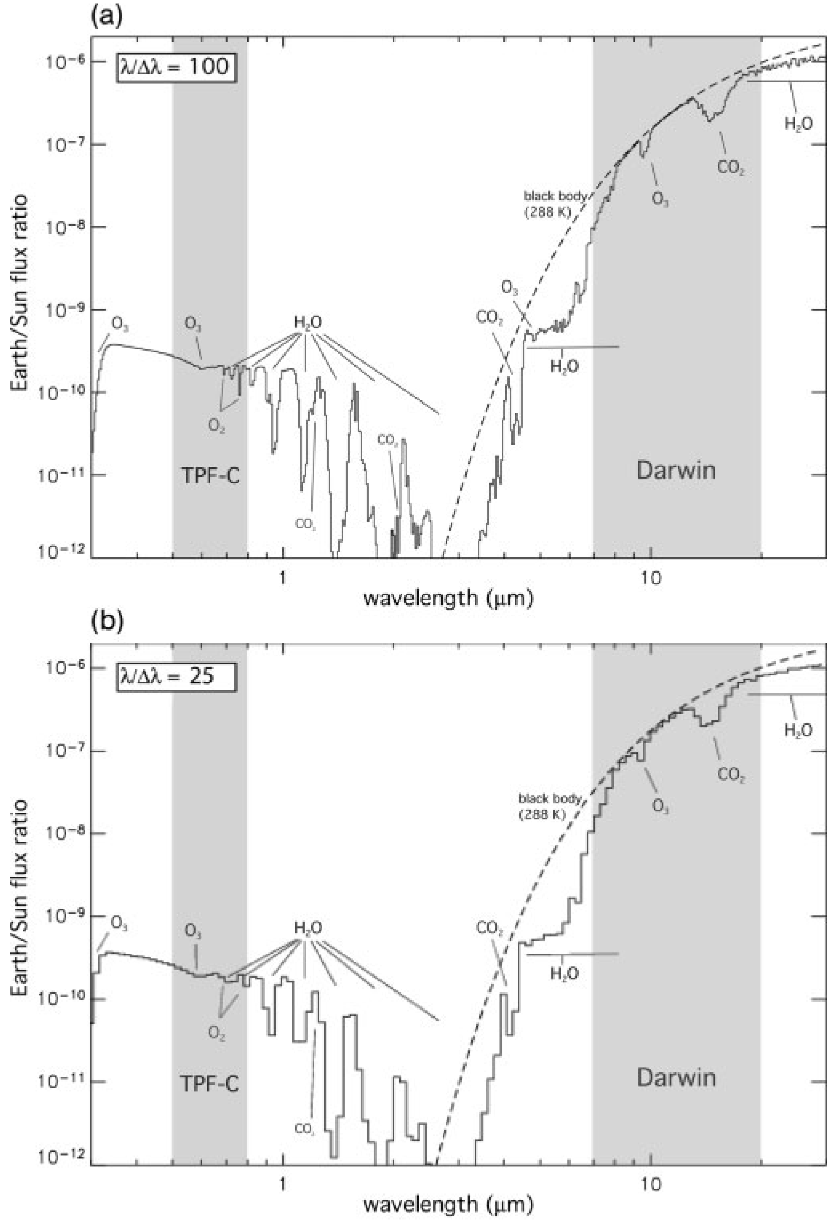


Fig. 4.1 Synthetic spectra of the earth from UV to IR shown in two different resolutions (R=100 and 25) representing the proposed resolution for TPF-C (R=70) and Darwin TPF-I (R=25). The intensity is given as a fraction of solar intensity. The atmospheric features as well as the spectroscopic range of Darwin/TPF-I and TPF-C is indicated [17].

4.3
Biomarker Signatures in Different Wavelength Ranges

With arbitrarily high signal-to-noise and spatial and spectral resolution, it is relatively straightforward to remotely ascertain that earth is a habitable planet, replete with oceans, a greenhouse atmosphere, global geochemical cycles, and life. The interpretation of observations of other planets with limited signal-to-noise ratio (SNR) and spectral resolution, as well as absolutely no spatial resolution, as envisioned for the first generation missions like TPF-C and Darwin/TPF-I, will be far more challenging.

To search for signs of life with low resolution and limited information implies that we need to gather as much information as possible in order to understand what we will see. The following step-by-step approach can be taken to set the system in context. After detection, we will focus on main properties of the planetary system, its orbital elements as well as the presence of an atmosphere using the light curve of the planet and/or a crude estimate of the planetary nature using very low-resolution information (three or four channels) [18]. Then a higher resolution spectra will be used to identify the compounds of the planetary atmosphere and to establish the temperature and radius of the observed exoplanet. In that context, we can then test if we have an abiotic explanation of all compounds seen in the atmosphere of such a planet. If we do not, we can work with the exciting biotic hypothesis. Second-generation space mission will then investigate those targets in more detail to improve our understanding of those environments. The results of a first-generation mission will most likely result in an amazing scope of diverse planets that will set planet formation, evolution as well as our planet in an overall context.

The thermal infrared concepts, the Darwin and the Terrestrial Planet Finder Interferometer (TPF-I), and the visible wavelength concepts, the Terrestrial Planet Finder Coronagraph (TPF-C), are designed to detect terrestrial exoplanets, and to measure the color and spectra of terrestrial planets, giant planets, and zodiacal dust disks around nearby stars (see, e.g., [19–25]). These missions have the explicit purpose of detecting other earth-like worlds, analyzing their characteristics, determining the composition of their atmospheres, investigating their capability to sustain life as we know it, and searching for signs of life. They are a first step in characterizing a vast number of unknown fascinating planetary worlds. Their respective resolution is envisioned to about 25 for Darwin and TPF-I and about 70 for TPF-C. These missions also have the capacity to investigate the physical properties and composition of a broader diversity of planets, to understand the formation of planets and interpret potential biosignatures.

The range of characteristics of planets is likely to exceed our experience with the planets and satellites in our own solar system. Earth-like planets orbiting stars of different spectral type might evolve differently [26–28]. Models of such planets need to consider the changing atmosphere structure, as well as the interior structure of the planet (see e.g., [29, 30]). One crucial factor in interpreting planetary spectra is the point in the evolution of the atmosphere when its biomarkers become detectable. Studies of individual constituent effects on the spectrum and resolution estimates were previously discussed for earth in Des Marais et al. [14]. These calculations were for a current atmospheric temperature structure, but different abundances of chemical species. Spectra of the earth explore temperature sensitivity (a hot house and cold scenario) and different singled out stages of its evolution (e.g., [18, 26, 27, 31–33]) as well as the evolution of the expected spectra of earth [33] and produce a variety of spectral fingerprints for our own planet. Those spectra will be used as part of a big grid to characterize any exoplanets found. This also influences the design requirements for a spectrometer [33] to detect habitability.

4.4
Potential Biomarkers

Oxygen in high abundance is a promising bio-indicator. Oxygenic photosynthesis, which is a byproduct of molecular oxygen extracted from water, allows terrestrial plants and photosynthetic bacteria (cyanobacteria) to use abundant H_2O, instead of having to rely on scarce supplies of electron donor to reduce CO_2, like H_2 and H_2S. With oxygenic photosynthesis, the production of the biomass becomes limited only by nutriments and no longer by energy (light in this case) nor by the abundance of electron donors. Oxygenic photosynthesis at a planetary scale results in the storage of large amounts of radiative energy as chemical energy, in the form of organic matter. For this reason, oxygenic photosynthesis had a tremendous impact on biogeochemical cycles on earth and eventually resulted in the global transformation of the earth's environment. Reduced gases and oxygen have to be produced concurrently to be detectable in the atmosphere, as they react rapidly with each other. Thus a detectable concentration of O_2 and/or O_3 and of a reduced gas like CH_4 can be considered as a signature of biological activity. The spectrum of the earth has exhibited a strong infrared signature of ozone for more than 2 billion years, and a strong visible signature of O_2 for a period of time between 2 and 0.8 billion years (depending on the required depth of the band for detection and also the actual evolution of the O_2 level). This difference is due to the fact that a saturated ozone band appears already at very low levels of O_2 (10^{-4} ppm) while the oxygen line remains unsaturated at values below 1 Present Atmospheric Level (PAL). Note that the non-detection of O_2 or O_3 on an exoplanet cannot be interpreted as the absence of life.

N_2O is produced in abundance by life but only in trace amounts by natural processes. Nearly all of earth's N_2O is produced by the activities of anaerobic denitrifying bacteria. N_2O would be hard to detect in the earth's atmosphere with low resolution, as its abundance is low at the surface (0.3 ppmv) and falls off rapidly in the stratosphere. On a low-O_2 early earth, its abundance would be even smaller because it photolyzes rapidly in the near ultraviolet. As signs of life in themselves H_2O and CO_2 are secondary in importance because, although they are not indicators of its presence, they are raw materials for life and thus necessary for planetary habitability.

In the visible reflected spectrum of earth, the detectable signatures of biological activity in low resolution are water vapour and molecular oxygen (mainly the 0.76 μm band), in the near-IR CO_2 and CH_4 are detectable at concentrations significantly higher than on current earth.

In the mid-IR thermal emission spectra of earth the combined detection of 9.6 μm O_3 band, the 15 μm CO_2 band and the 6.3 μm H_2O band or its rotational band that extends from 12 μm out into the microwave region [38] indicating habitability. The 9.6 μm O_3 band is a very nonlinear indicator of O_2: First, for the present atmosphere, low resolution spectra of this band show little change with the O_3 abundance because it is strongly saturated. Second, the apparent depth of this band remains nearly constant as O_2 increases from 0.01 to 1 PAL [26]. The primary reason for this is that the stratospheric warming decreases with the abundance of ozone, making the O_3 band deeper. The depth of the saturated O_3 band is indeed determined by the temperature difference between the surface-clouds continuum and the ozone layer. CH_4 is not readily identified using low-resolution spectroscopy for present-day earth, but the methane feature at 7.66 μm in the IR is easily detectable at higher abundances (see, e.g., 100× abundance, Fig 4.5 epoch 4 below), provided of course that the spectrum contains the whole band and a high enough SNR. Taken together with molecular oxygen, abundant CH_4 can indicate biological processes (see also [1, 16, 26]). Depending on the degree of oxidation of a planet's crust and upper mantle nonbiological mechanisms can also produce large amounts of CH_4 under certain circumstances.

N_2O is produced by life but only in negligible amounts by abiotic processes. There are potentially three weak N_2O features in the thermal infrared at 7.75 μm and 8.52 μm, and 16.89 μm. Methane and nitrous oxide have features nearly overlapping in the 7 μm region, and additionally, both lie in the red wing of the 6 μm water band. Although its abundance is less than 1 ppm in the earth's atmosphere, the 7.75 μm shows up in a medium resolution infrared spectrum. Spectral features of N_2O would become more apparent in atmospheres with more N_2O and/or less H_2O vapor. On a low-O_2 early earth, its abundance would be even smaller because it photolyzes rapidly in the near-ultraviolet. Segura et al. (2003) have calculated the level of N_2O for different

O_2 levels and found that, although N_2O is a reduced species, its levels decreases with O_2. This is due to the fact that a decrease in O_2 produces an increase of H_2O photolysis, resulting in the production of more hydroxyl radicals (OH) responsible for the destruction of N_2O.

There are other molecules that could, under some circumstances, act as excellent biomarkers, e.g., the manufactured chloro-fluorocarbons (CCl_2F_2 and CCL_3F) in our current atmosphere in the thermal infrared waveband, but their abundances are too low to be spectroscopically observed at low resolution.

4.5
A Habitable Planet

The circumstellar Habitable Zone (HZ) is defined as the region around a star within which starlight is sufficiently intense to maintain liquid water at the surface of the planet, without initiating runaway greenhouse conditions vaporizing the whole water reservoir and, as a second effect, inducing the photodissociation of water vapor and the loss of hydrogen to space (Fig. 4.2), see [13, 39] for a detailed discussion. On an earth-like planet where the carbonate–silicate cycle is at work, the level of CO_2 in the atmosphere depends on the orbital distance: CO_2 is a trace gas close to the inner edge of the HZ but a major compound in the outer part of the HZ. Earth-like planets close to the inner edge are expected to have a water-rich atmosphere or to have lost their water reservoir to space. As the HZ is defined for surface conditions only, chimiolithotrophic life, a metabolism which does not depend on the stellar light, can still exist outside the HZ, thriving in the interior of the planet where liquid water is available. Such metabolisms (at least the ones we know on earth) do not produce O_2 and rely on very limited sources of energy (compared to stellar light) and electron donors (compared to H_2O on earth). They mainly catalyze reactions that would occur at a slower rate in purely abiotic conditions and they are thus not expected to modify a whole planetary environment in a detectable way.

As we said, partial pressure of CO_2 and H_2O at the surface of an earth-like habitable planet is a function of the orbital distance a, within the HZ and for the present solar luminosity. We assume here that the planet contains one terrestrial ocean of superficial water and that the carbonate–silicate cycle is at work, controlling the CO_2 level in equilibrium with a surface temperature at about 290 K, for $a > 1$ AU. For $a > 1.3$ AU, CO_2 condenses in the atmosphere producing CO_2 clouds that can affect the temperature–CO_2 coupling significantly. For $a < 0.93$ AU, H_2O becomes a major atmospheric compound and is rapidly lost to space after UV photolysis. Thus, a Venus-like fate (no H_2O remaining and a massive CO_2 build-up) is likely in this inner part of the HZ. This is one of the first theories we can test with a first-generation space

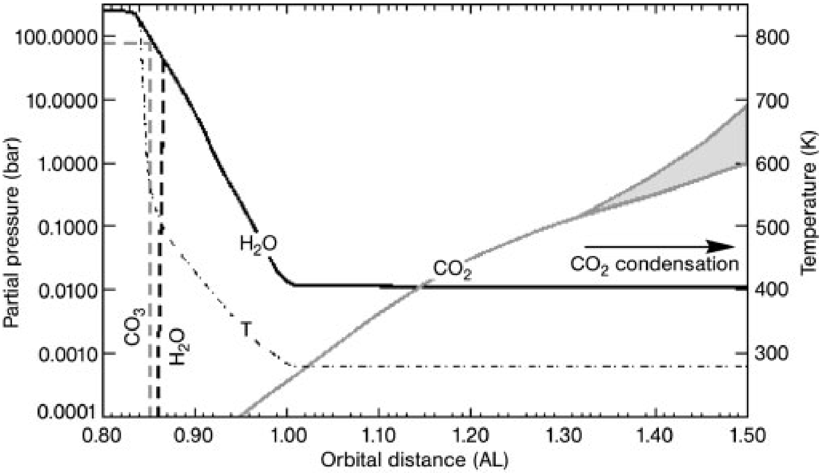


Fig. 4.2 The mean surface temperature (TS) and partial pressure of CO_2 and H_2O as a function of the orbital distance on a habitable planet within the habitable zone (data adapted from Kasting et al. [13] and Forget and Pierehumbert [40]) (partial pressure (left y-axis) and TS (right y-axis)).

mission. However, the limits of the HZ are known qualitatively, more than quantitatively. This uncertainty is mainly due to the complex role of clouds but also three-dimensional climatic effects not yet included in the modeling. Thus, planets slightly outside the computed HZ could still be habitable, while planets at habitable orbital distance may not be habitable because of their size or chemical composition.

4.6
Oxygen and Ozone Production on Earth

Owen (1980) suggested searching for O_2 as a tracer of life. In the particular case of earth, O_2 is fully produced by the biosphere. Less than 1 ppm comes from abiotic processes [41]. Cyanobacteria and plants are responsible for this production by using the solar photons to extract hydrogen from water and using it to produce organic molecules from CO_2. This metabolism is called oxygenic photosynthesis. The reverse reaction, using O_2 to oxidize the organics produced by photosynthesis, can occur abiotically when organics are exposed to free oxygen, or biologically by eukaryotes breathing O_2 and consuming organics. Because of this balance, the net release of O_2 in the atmosphere is due to the burial of organics in sediments (see Fig. 4.3). Each reduced carbon buried, releases a free O_2 molecule into the atmosphere. This net release rate is

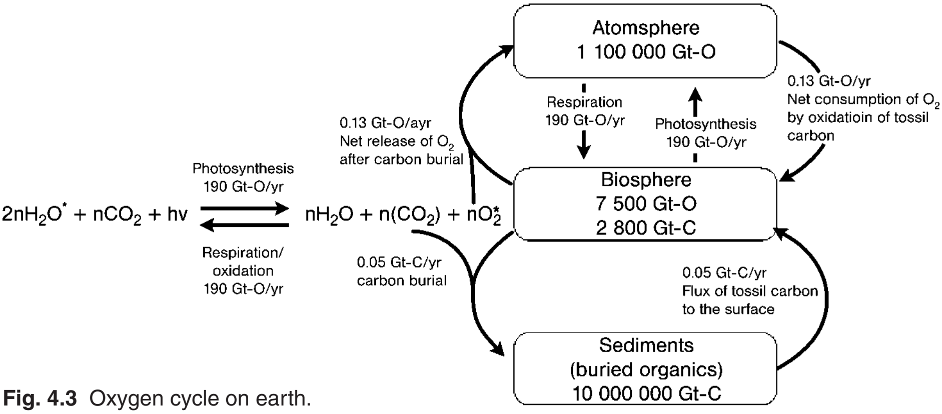


Fig. 4.3 Oxygen cycle on earth.

also balanced by weathering of fossilized carbon when exposed to the surface. The oxidation of reduced volcanic gasses such as H_2 and H_2S also accounts for a significant fraction of the oxygen losses. The atmospheric oxygen is recycled through respiration and photosynthesis in less than 10 000 yrs. In the case of a total extinction of the earth's biosphere, the atmospheric O_2 would disappear in a few million years.

Ozone is produced in the atmosphere by a unique chemical reaction: $O + O_2 + M \rightarrow O_3 + M$, where M is any compound. This reaction is not very efficient as it requires, at the same time, a high enough pressure (because it is a three-body reaction), and oxygen atoms that are produced at lower pressures where photolysis of O_2 by UV can occur. Ozone can be efficiently destroyed by a large number of reactions, dominated, in the earth's atmosphere, by catalytic cycles involving trace species such as hydrogenous compounds (H, OH, HO_2), nitrogen oxides (NO_X) and chlorine compounds (ClO_X). These species have various origins and their amount depends on the nature and intensity of the bio-productivity, the thermal profile of the atmosphere, human pollution, and many other parameters. Without these compounds, an atmosphere made of N_2 and O_2 only would contain 10 times more O_3. The column density of O_3 in the atmosphere depends weakly on the abundance of O_2, the mean opacity of the 9.6 μm band remaining > 1 for O_2 abundance as low as 10^{-3} PAL [19, 26].

The abundance of O_3 and its observability also vary with the spectral distribution of the incoming stellar radiation. For atmospheric compositions similar to that of the earth, numerical simulations show that O_3 increases with the UV flux [26, 27]. The depth of the O_3 feature depends, however, on the difference beween the brightness temperature of the continuum (given by the temperature of the surface and/or the clouds) and the temperature of the ozone layer (where the opacity of O_3 is 1). Because of this complex coupling, less ozone means less warming and still a deep feature, while more ozone produces more warming and a shallow feature.

4.7
Cloud Features

Clouds are an important component of exoplanet spectra because in the visible/NIR range, their reflection is high and relatively flat with wavelength, while in the infrared they lower the emitted flux and hide the lower convective region of the atmosphere which produces most of the spectral features (due to the strong gradient of temperature associated with the convection). Clouds hide the atmospheric molecular species below them, weakening the spectral lines in both the thermal infrared and visible. In the thermal infrared, clouds emit at temperatures that are generally colder than the surface, while in the visible the clouds themselves have different spectrally-dependent albedos that further influence the overall shape of the spectrum. Figure 4.4 shows the visible and thermal infrared spectral emission of the earth for three cloud conditions (a) and the model-to-data comparison (b) [33].

4.8
Biomarkers and their Evolution over Geological Times on Earth

The spectrum of the earth has not been static throughout the past 4.5 Ga (Ga = 10^9 years ago). This is due to the variations in the molecular abundances, the temperature structure, and the surface morphology over time. Kasting and Catling [44] and Kasting [45] established a scenario for the earth's atmosphere evolution. At about 2.3 Ga oxygen and ozone became abundant, affecting the atmospheric absorption component of the spectrum. At about 2 Ga, a green phytoplankton signal developed in the oceans and at about 0.44 Ga, an extensive land plant cover followed, generating the red chlorophyll edge in the reflection spectrum. The composition of the surface (especially in the visible), the atmospheric composition, and temperature-pressure profile can all have a significant influence on the detectability of a signal. Note that we assume that the cloud cover over these epochs is the same as the cloud cover today. However, changes in this distribution could significantly change the overall spectra in both wavelength regions.

Figure 4.5a shows theoretical visible and mid-infrared spectra of the earth at six epochs during its geological evolution and (b) the required resolution to match the main spectral features and biomarkers of atmospheric compounds over geological time for earth [33]. The epochs are chosen to represent major developmental stages of the earth, and life on earth. Earth's atmosphere has experienced dramatic evolution over 4.5 billion years, and other planets may exhibit similar or greater evolution, at similar or different rates. The climate model to create a schematic atmospheric model of our earth over geological timescales is based on a combination of results from work by Kasting

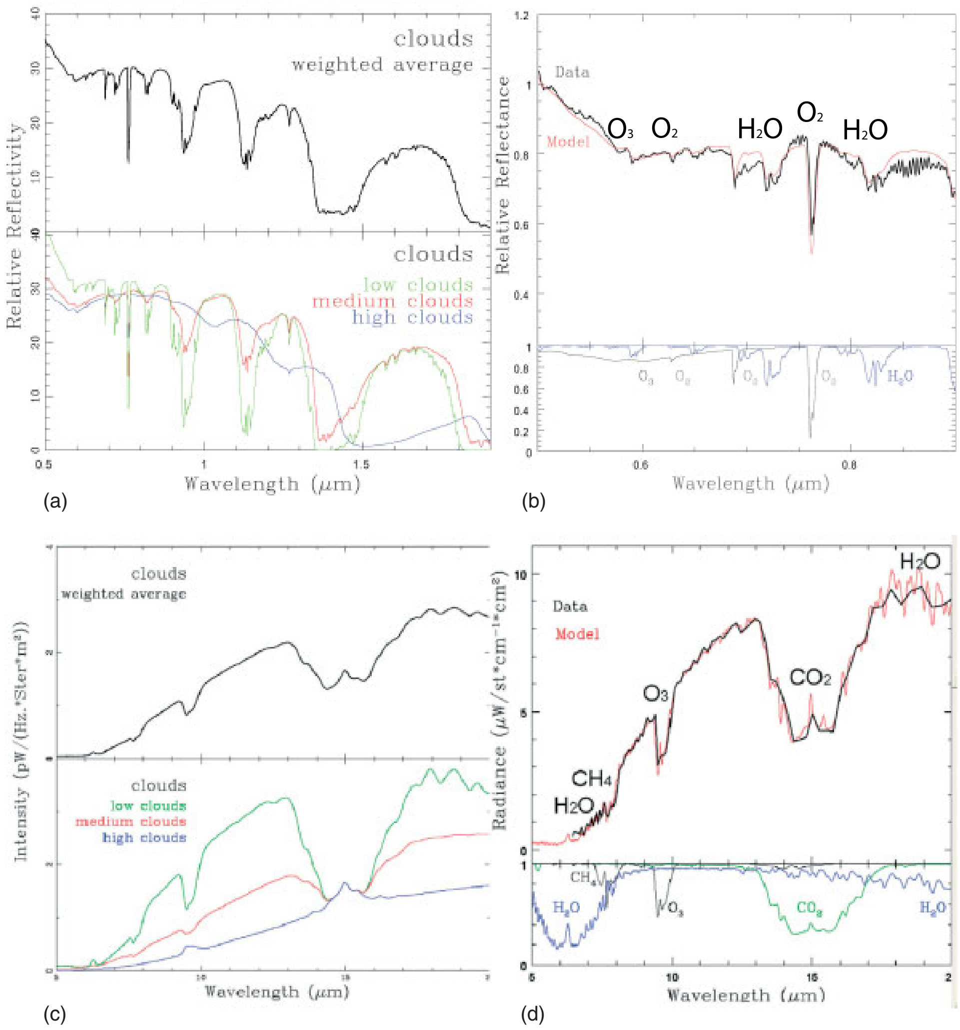


Fig. 4.4 (a, c) spectra of present-day earth with 100% cumulus cloud coverage at 1 km, 100% cumulus cloud coverage at 6 km, 100% cirrus cloud coverage at 12 km, and spectra of a mixture of clouds resembling the present earth in the visible (a, b) and in thermal infrared (c, d). Model Data comparison (b, d) data in black [42, 43] model in gray [33].

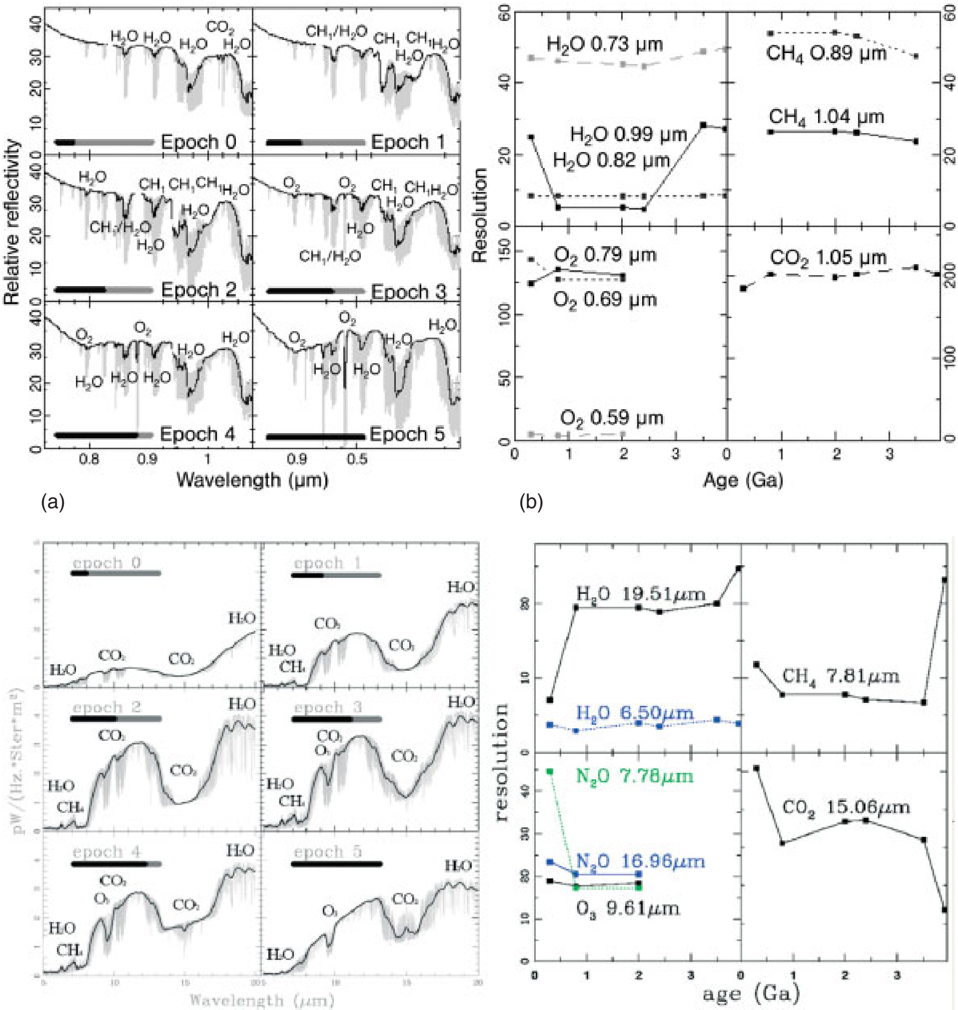
and Catling [44], Kasting [45], Pavlov et al. [46], Segura et al. [26] and Traub and Jucks [18]. The model atmosphere evolves from a CO_2-rich atmosphere (3.9 Ga = epoch 0) to a CO_2/CH_4-rich atmosphere (epoch 3) to a present- day atmosphere (epoch 5 = present-day earth). It shows epochs that reflect significant changes in the chemical composition of the atmosphere. The oxygen and ozone absorption features could have been used to indicate the presence of biological activity on earth at any time during the past 50% of the age of the solar system. The dark lines show a resolution of 70 and 25, as proposed for TPF-C in the visible and Darwin/TPF-I in the IR respectively.

Different signatures in the atmosphere are clearly visible over earth's evolution and observable with low resolution. Only in the last epoch (5), vegetation modifies the reflected spectrum of earth by introducing a sharp increase of the reflectivity between 700–750 nm, the red edge. Its detection is extremely challenging in a spatially unresolved global planetary spectrum, integrated over a long exposure [17, 33–37, 47]. The spectral resolution required for optimal detection of habitability and biosignatures has to match those features on our own planet for the dataset we have over its evolution.

4.9
Planets around Different Stars

For an earth-like planet in the HZ around a given star, the thermal flux will, to first order, be constant for a given planetary size, while the reflected stellar flux will scale with the brightness of the star. The suppression of the primary's thermal emission will, on the other hand, be progressively easier for later and later spectral types. The contrast ratio is larger by a factor of about 2 for FV stars, while it is smaller by a factor of 3 for KV stars and 11 for MV stars, respectively, in the N band compared to sun–earth [48]. Surprisingly enough, it may thus be easier for the IR interferometer concept to detect a habitable earth around an M-Dwarf than around something more akin to our own sun. This is true for interferometric systems like Darwin and TPF-I that can be adapted to each individual target system, since the HZ moves closer and closer to the star for later and later stellar types. The baseline of the interferometers have to increase to resolve M star planetary systems at larger distances, a constraint that is taken into account for the M-target systems in the target star catalogue [48].

Using a numerical code that simulates the photochemistry of a wide range of planetary atmospheres, Selsis [27] and Segura et al. [26, 28] have simulated a replica of our planet orbiting different types of star: an F-type star (more massive and hotter than the sun) and a K-type star (smaller and cooler than the sun). The models assume the same background composition of the atmosphere as well as the strength of biogenic sources. The orbital distance was chosen in order to give the planet the earth's effective temperature (by receiving the same energetic flux): 1.8 and 0.5 AU respectively for the F- and the K-type star. Scaling the energetic flux allows us to consider habitable planets irradiated by a nonsolar spectrum: the contribution of the UV range (150–400 nm, the most important for photochemistry) is higher for the F-type star



(a) (b)

(c) (d)

Fig. 4.5 The visible (a, b) and mid-IR (c, d) spectral features on an earth-like planet change considerably over its evolution from a CO_2 rich (epoch 0) to a CO_2/CH_4-rich atmosphere (epoch 3) to a present-day atmosphere (epoch 5) (blacklines shows spectral resolution of 70 for the visile and 25 for the mid-IR). (b, d) Required resolution to match the main spectral features and biomarkers of atmospheric compounds over geological time for earth [33].

and lower for the K-type star (this is no longer true in the EUV range, below 150 nm, where low-mass stars like K-type stars are very active).

The results of modeling, illustrated in Fig. 4.6, show the changes in detectability and shape of spectral features due to ozone, carbon dioxide, and methane for the “same” planet around stars of different spectral type. These

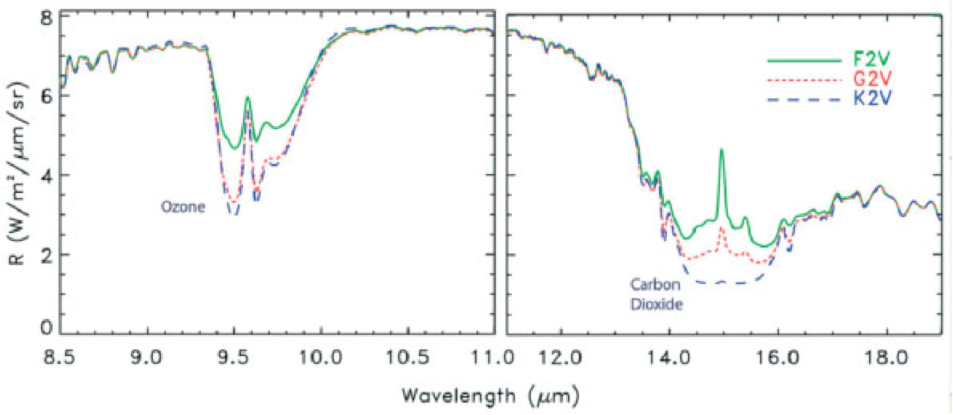


Fig. 4.6 These spectra show the appearance of earth-like planets orbiting within the habitable zones of stars of F2V, G2V and K2V stellar type. In each case, a weakly-coupled radiative/photochemical atmospheric model was used to determine the equilibrium atmospheric composition and vertical structure for a planet with earth's modern atmospheric composition, and radiatively forced by the UV to the far-infrared spectrum of a star of each spectral type [28].

changes are due to an interplay between the star's spectrum, the photochemistry of ozone, and coupled changes in the thermal structure of the planet's atmosphere. These models were run for host stars of F, G, K [26, 27] and M spectral type [28] and show that in high resolution the detectable features around, e.g., a K star are deeper than the features around an F host star.

4.10
Abiotic Sources

We need to address the abiotic sources of biomarkers, so that we can identify when it might constitute a "false positive" for life. CH_4 is an abundant constituent of the cold planetary atmospheres in the outer solar system. On earth, it is produced abiotically in hydrothermal systems where H_2 (produced from the oxidation of Fe by water) reacts with CO_2 in a certain range of pressures and temperatures. In the absence of atmospheric oxygen, abiotic methane could build up to detectable levels. Therefore, the sole detection of CH_4 cannot be attributed unambiguously to life.

O_2 also has abiotic sources, the first one is the photolysis of CO_2, followed by recombination of O atoms to form O_2 ($O + O + M \rightarrow O_2 + M$), a second one is the photolysis of H_2O followed by escape of hydrogen to space. The first source is a steady state maintained by the stellar UV radiation, but with a constant elemental composition of the atmosphere, while the second one is a net source of oxygen. In order to reach detectable levels of O_2 (in the reflected spectrum), the photolysis of CO_2 has to occur in the absence of outgassing of reduced species and in the absence of liquid water, because of the rainout of oxidized species. Normally, the detection of the water vapor bands simultaneously with the O_2 band can rule out this abiotic mechanism (Segura et al., in preparation). In the infrared, this process cannot produce a detectable O_3 feature [38]. The loss of hydrogen to space can result in massive oxygen leftovers: 240 bars of oxygen could build up after the loss of the hydrogen contained in the earth's ocean. However, the case of Venus tells us that such oxygen leftover has a limited lifetime in the atmosphere (because of the oxidation of the crust and the loss of oxygen to space): we do not find O_2 in the Venusian atmosphere despite the massive loss of water experienced in the early history of the planet. Also, such evaporation-induced build-up of O_2 should occur only close to the star (see Fig. 4.2) and affect small planets with low gravity more dramatically. For small planets ($< 0.5\ M_{\text{earth}}$) close to the inner edge of the habitable zone (< 0.93 AU from the sun), there is a risk of abiotic oxygen detection, but this risk becomes negligible for bigger planets further away from their star.

4.11
Biomarkers Detection Set in Context

The search for signs of life implies that one needs to gather as much information as possible in order to understand how the observed atmosphere physically and chemically works. Knowledge of the temperature and planetary radius is crucial for the general understanding of the physical and chemical processes occurring on the planet (tectonics, hydrogen loss to space). In theory, spectroscopy can provide detailed information on the thermal profile of a planetary atmosphere. This however requires a spectral resolution and a sensitivity that are well beyond the performance of a first generation spacecraft like Darwin/TPF-I and TPF-C. Thus we will concentrate on the initially available observations here like flux variations of a planet throughout its orbit. Smaller coronagraphs or interferometers can probe similar characteristics for extrasolar giant planets.

4.11.1
Temperature and Radius of the Planets

One can calculate the stellar energy of the parent star that is received at the planet's measured orbital distance. This gives only very little information on the surface temperature of the planet, which depends on its albedo and the efficiency of the greenhouse effect. However, with a low-resolution spectrum of the thermal emission, the mean brightness temperature and the radius of the planet can be obtained, in first approximation, by fitting the envelope of the thermal emission by a Planck function. The ability to associate a brightness temperature to the spectrum by doing this fit relies on the existence and identification of spectral windows probing the surface or the same atmospheric levels. For an earth-like planet there are some atmospheric windows that can be used in most of the cases, especially between 8 and 11 μm as seen in Figs. 4.1 and 4.5. This window would, however, become opaque at high H_2O partial pressure (e.g., the inner part of the HZ where a lot of water is vaporized) and at high CO_2 pressure (e.g., a very young earth (see Fig. 4.5) or the outer part of the HZ). A much better estimate of the radius and of the temperature can be obtained by comparing the spectrum, including its features, with a grid of synthetic spectra (Paillet [17] and Kaltenegger and Selsis, in preparation).

4.12
Orbital Flux Variations

The orbital flux variation in the IR, measured in the detection phases, can provide some important information. The thermal light curve (i.e., the integrated infrared emission measured at different positions in the orbit) exhibits variations due to the phase (whether the observer sees mainly the day side or the night side) and to the season (if the planet has a nonzero obliquity). Important phase-related variations are due to a high day/night temperature contrast and imply a tenuous, or no, atmosphere and the absence of a stable liquid ocean. Therefore, habitable planets can be distinguished from airless or Mars-like planets by the amplitude of the observed variations [49, 50], see Fig. 4.7. Note that also Venus-like atmosphere would exhibit extremely low amplitudes and can only be distinguished by spectroscopy from habitable planets.

The mean value of T_b estimated over an orbit can be used to estimate the albedo of the planet, A, through the balance between the incoming stellar radiation and the outgoing IR emission:

$$F_{\text{star}}(1 - A) = 4 < T_b >^4_{\text{orbit}} \tag{4.1}$$

In the visible ranges, the reflected flux allows us to measure the product AR^2, where R is the planetary radius (a small, but reflecting, planet could be as bright as a big, but dark, planet). The first generation of optical instruments will be very far from the angular resolution required to directly measure an exoplanet radius. Currently, such a measurement can only be performed when the planet transits in front of its parent star, using an accurate photometric technique. If the same target is observed in both visible (TPF-C) and IR (Darwin/TPF-I) ranges, the albedo can be obtained in the visible once the radius is inferred from the IR spectrum. The measurement of the albedo in the visible can then be compared to the albedo estimated from the IR light curve and the results can be iterated. The possibility of obtaining spectral

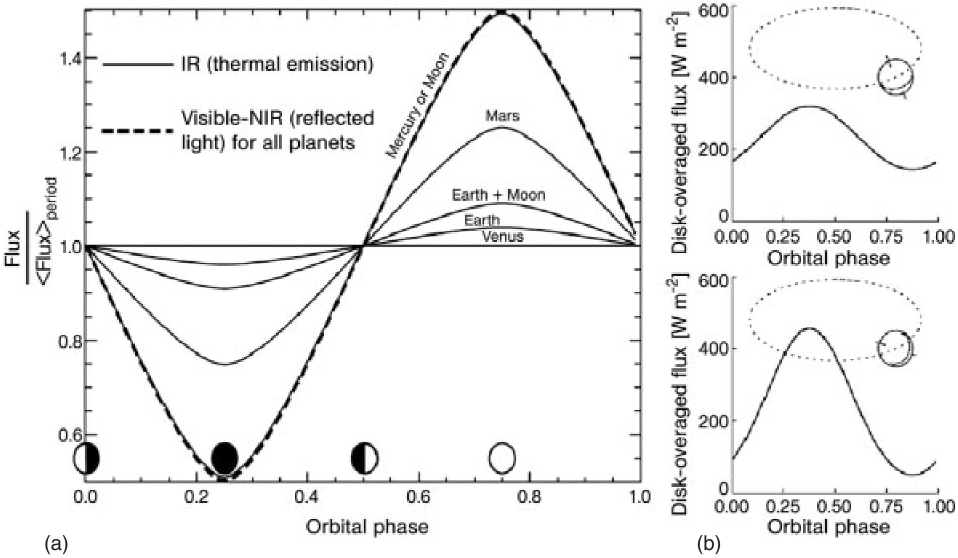


Fig. 4.7 Orbital light curve for black body planets in a circular orbit with and without an atmosphere in the thermal infrared [49] (a) and the influence of a different rotation axis on the shape of the light curve [50] (b).

information from both wavelengths will allow a more detailed characterization of individual planets and it also allows exploration of a wide domain of planet diversity. Observations in both complementary wavelength bands can confirm the presence of atmospheric compounds. Some important species like CO_2 and N_2O appear only in the IR range, while only the reflected spectrum can give information on the nature of the surface. Information on the cloud and surface characteristics (ocean, ice, rocks) can be obtained once the absolute level of the albedo is known, which requires the knowledge of the radius. At extremely high temporal resolution and signal-to-noise ratio additional modulations of the amplitude variations of the visual reflected light over a daily rotation of a planet could indicate surface features [51]. As shown in the same paper, clouds destroy that relation, because of their own individual rotation pattern.

The outgoing short-wave and long-wave radiation, combined with their variations along the orbit, can constrain albedo, greenhouse gases and would allow to explore climate systems at work on the observed worlds.

4.13
Summary

Any information which we collect on habitability, is only important in a context that allows us to interpret what we find. The search for signs of life implies gathering as much information as possible in order to understand how the observed atmosphere works physically and chemically. Knowledge of the temperature and planetary radius is crucial for the general understanding of the physical and chemical processes occurring on the planet. These parameters, as well as an indication of habitability, can be determined with low resolution spectroscopy and low photon flux, as assumed for first-generation space missions.

Our search for signs of life is based on the assumption that extraterrestrial life shares fundamental characteristics with life on earth, in that it requires liquid water as a solvent and has a carbon-based chemistry. We adopt this conservative approach to rule out false positives, completely aware that we will miss labeling some planets as habitable, where life has developed based on other chemistry or where the environment does not lead to a build-up of oxygen in the atmosphere. An amazing feature of future space-based missions like Darwin and TPF, is that they will have the capability, for the first time, to carry out comparative planetology on a wide variety of planets, with atmospheres and climatologies far outside our current understanding, as well as finding planets similar to our own and probing them for habitable conditions.

References

1 Sagan, C., Thompson, W. R., Carlson, R., Gurnett, D., and Hord, C.: 1993, *Nature*, **365**, 715.

2 Charbonneau, D., Brown, T. M., Noyes, R. W., and Gilliland, R. L.: 2002, *ApJ*, **568**, 377.

3 Udry, S, Bonfils, X, Delfosse, X, Forveille, T, Mayor, M, Perrier, C, Bouchy, F, Lovis, C., Pepe, F., Queloz, D, and Bertaux, J.-L.: 2007, *A&A*, **469**, L43–L47.

4 Rivera, E., et al.: 2005, *ApJ*, **634**, 1.

5 Beaulieu, J.-P. et al.: 2006, *Nature*, **439**, 437.

6 Gould A. et al.: 2006, *ApJ*, **644**.

7 Rouan, D., Baglin, A., Copet, E., Schneider, J., Barge, P., Deleuil, M., Vuillemin A., and Leger, A.: 1998, *Earth, Moon, Planets*, 79.

8 Borucki, W. J., Koch, D. G., Dunham, E. W., and Jenkins, J. M.: 1997, Planets Beyond the solar system and the Next Generation of Space Missions, *ASP Conf. Ser.*, **119**, 153.

9 Traub, W.A., Ridgway, S.T., Beichman, C.A., Johnston, K.J., Kasting, J., and Shao, M.: 2006, *Proc. IAUC 200*, Cambridge: Cambridge University Press.

10 Charbonneau, D.; Allen, L. E.; Megeath, S. T.; Torres, G.; Alonso, R.; Brown, T. M.; Gilliland, R. L.; Latham, D. W.; Mandushev, G.; O'Donovan, F. T., and Sozzetti, A.: 2005, *ApJ*, **626**, 523.

11 Deming, D., Seager, S., Richardson, L. J., Harrington, J.: 20205, *Nature*, **435**, 740.

12 Harrington, J., et al.: 2006, *Science*, **314**, 623.

13 Kasting, J.F., Whitmire, D.P., and Reynolds, H.: 1993, *Icarus*, **101**, 108.

14 Des Marais, D. J., Harwit, M. O., Jucks, K. W., Kasting, J. F., Lin, D. N. C., Lunine, J. I., Schneider, J., Seager, S., Traub, W. A., Woolf, N. J.: 2002, *Astrobiology*, **2**, 153.

15 Owen, T.: 1980, *Strategies for Search for Life in the Universe*, Dordrecht, The Netherlands, 177.

16 Lovelock J.E., *Proc.R. Soc. Lond. B* , **189**, 167.

17 Paillet, J.: 2006, Spectral characterization of terrestrial exoplanets, *phD thesis*.

18 Traub, W.A. and Jucks, K.: 2002, *AGU Monograph Series*, **130**, 369.

19 Leger, A., Pirre, M., and Marceau, F. J.: 1993, *A&A*, **277**, 309.

20 Beichman, C. A., Woolf, N. J., and Lindensmith, C. A.: 1999, The Terrestrial Planet Finder (TPF): a NASA Origins Program to Search for Habitable Planets / the TPF Science Working Group, Washington DC: NASA JPL, *JPL publication*, **99–3**.

21 Beichman, C.A., Fridlund, M., Traub, W.A., Stapelfeldt, K.R., Quirrenbach, A., and Seager, S.: 2006, *Protostars and Planets V*, Tuscon: University of Arizona Press.

22 Fridlund, M.: 2000, DARWIN The InfraRed Space Interferometer, (Leiden: ESA), ESA-SCI(2000), **12**, 47.

23 Kaltenegger, L.: 2004, Search for Extraterrestrial Planets: The DARWIN mission. Target Stars and Array Architectures, *phD thesis astro-ph/0504497*.

24 Kaltenegger, L. and Fridlund, M.: 2005, *Advances in Space Research*, **36**, 1114.

25 Borde, P.J., Traub, W.A.: 2006, *ApJ*, **638**, 488.

26 Segura, A., Krelove, K., Kasting, J. F., Sommerlatt, D., Meadows V., Crisp D., Cohen, M., and Mlawer, E.: 2003, *Astrobiology*, **3**, 689.

27 Selsis, F.: 2000, Darwin and astronomy: the infrared space interferometer, *ESA SP*, **451**, 133.

28 Segura, A., Kasting, J.F., Meadows, V., Cohen, M., Scalo, J., Crisp, D., Butler, R. A. H., and Tinetti, G.: 2005, *Astrobiology*, 706.

29 Valencia, D., et al.: 2006, *Icarus*, **181**.

30 Sotin C. et al.: 2007, (**in prep**).

31 Schindler, T. L. and Kasting, J. F.: 2000, *Icarus*, **145**, 262.

32 Pavlov, A.A., Kasting, J.F., Brown, L.L., Rages, K.A., Freedman R., and Greenhouse R.: 2000, *J. Geophys. Res.*, **105**, 981.

33 Kaltenegger, L., Traub, W.A., and Jucks, K.: 2006, *ApJ*.

34 Arnold, L., Gillet, S., Lardiere, O., Riaud, P., and Schneider, J.: 2002, *A&A*, **392**, 231.

35 Montanes-Rodriguez P., Palle E., and Goode P.R.: 2006, *ApJ*.

36 Montanes-Rodriguez P., et al.: 2007, *ApJ*, **629**.

37 Seager, S. and Ford, E.B.: 2002, Astrophysics of Life, Cambridge: Cambridge University Press, 57.

38 Selsis, F., Despois, D., Parisot, J.-P.: 2002, *A&A*, **388**, 985.

39 Kasting, J.F.: 1997, *Origins of Life*, 291.

40 Forget, P. and Pierehumbert, H.: 1997, *Science*, **278**.

41 Walker, J.C.G.: 1977, *Evolution of the Atmosphere*, Macmillan, New York.

42 Christensen, P.R., Pearl, J.C.: 1997, *J. Geoph. Res.*, **102**, 875.

43 Woolf, N. J., Smith, P. S., Traub, W. A., and Jucks, K. W.: 2002, *ApJ*, **574**, 430.

44 Kasting, J.F. and Catling, D.: 2003, *Ann. Rev. Astron. Astrophys.*, **41**, 429.

45 Kasting, J.F.: 2004, *Scientific American*, **80**.

46 Pavlov, A.A., Hurtgen, M.T., Kasting, J. F., and Arthur, M.A.: 2003, *Geology*, **31**, 87.

47 Seager, S., Turner, E. L., Schafer, J., and Ford, E. B.: 2005, *Astrobiology*.

48 Kaltenegger, L., Eiroa, C., Stankov, A.-M., and Fridlund, M.: 2007, *A&A*, **submitted**.

49 Selsis, F.: 2003, *Proc. ESA-SP539, ESA, Netherlands*.

50 Gaidos, E. and Williams, D. M.: 2004, *New Astronomy*, **10**, 67.

51 Ford, E. Seager, S., and Turner, E.L.: 2001, *Nature*, **412**, 885.

5
The Formation of Resonant Planetary Systems

Wilhelm Kley and Zsolt Sándor

Abstract

Among the detected extrasolar planetary systems with multiple planets, a high fraction (about 30%) is engaged in a mean-motion resonance. Here the orbital periods of the involved planets are commensurable and their ratio can be described by two small integers. In particular 2:1 commensurabilities of the periods are very common in extrasolar planetary systems. In the solar system, despite several satellite systems, the only planetary mean-motion resonance is between Neptune and Pluto which are locked in a 3:2 resonance. From the analysis of the solar system resonances it is known that satellites and also planets can be driven into a resonant configuration through the operation of a dissipative mechanism that is able to change the energy of the orbits, i.e., the corresponding semi-major axis of the objects. In the context of planet formation of massive planets, the interaction of the protoplanets with the ambient gaseous protoplanetary disk is able to provide a differential migration between two planets and, upon reaching commensurability, resonant capture may occur. In this review we briefly summarize the observational evidence for mean-motion resonances in the solar system and in extrasolar planetary systems, and then describe in more detail possible formation scenarios and the particular conditions that may have led to the presently observed states.

5.1
The Solar System

After the former planet Pluto lost its planetary status in August 2006 none of the remaining eight planets in the solar system are engaged in a mutual mean-motion resonance (MMR). The so-called Plutinos, of which Pluto is the name-giving major member, are located in 3:2 resonance with Neptune. However, the Jovian planets all lie close to a MMR when considered in pairs. Well known is the *great inequality* of Jupiter and Saturn whose period ratio lies near the 5:2 commensurability. The corresponding ratios in Saturn and Uranus lie

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

near 3:1, and in Uranus and Neptune near 2:1. In the solar system orbital MMR are found in the asteroid belt where exact commensurabilities with Jupiter occur, and in the Kuiper belt with Neptune [1].

Additionally, in the satellite systems of Jupiter and Saturn several resonances are found. For the MMR in satellite systems it has been established that they most likely have been formed and stabilized in the resonance through tidal torques [2, 3]. Here the tidal interaction between the planet and a satellite will yield a slowing down of the spin-rate of the planet and an enlargement of the satellite's orbit. As a consequence, the outward moving satellite will capture its next neighbor in a resonance. However, for the asteroid belt and the Kuiper belt this tidal mechanism will not work and other dissipative mechanisms have to be found [4]. Malhotra has shown that the enhancement of Plutinos in the 3:2 resonance with Neptune is a consequence of Neptune's outward motion due to its interaction with the planetesimal disk in the late stage of the solar system's formation [5, 6]

Whether the likelihood of the near-resonant configuration of the major planets in the solar system is greater than anything produced by chance, has been a matter of debate in the past. By statistical arguments Roy et al. [7] conclude that the present configuration is highly unlikely to be random. Some earlier planet-formation scenarios have tried to explain the origin and particular spacing of the planets (Titius–Bode), for example, by a special vortical structure of the protoplanetary nebula [8]. Other explanations of the origin of the spacing of the planets in the solar system have invoked resonance trapping of smaller particles in the outer 2:1 resonance by an already existing planet. Here, these particles are driven by gas drag towards the planet to become eventually captured in the resonance where they will collide and grow to larger objects. According to Patterson [9] this scenario might have operated for the formation of additional planets outside the, already existing, Jupiter or Venus. Later Beaugé et al. [10] showed, through N-body simulations, the feasibility of growing Saturn in the 5:2 orbital resonance with Jupiter via this process. In these cases the resonances and subsequent capture occur with small particles and a massive planet, and the dynamics of such systems can be analyzed in the restricted three-body problem.

Whether two massive gaseous planets can be formed in this way is highly questionable. Currently, the two main formation scenarios, i.e., the gravitational instability model [11] or the core instability model [12] both allow the condensation of planets at arbitrary positions as long as the (instability) conditions are satisfied. This implies that the planets must have been brought at a later time into these near resonant conditions through a dissipative mechanism. After a description of the properties of the extrasolar systems in MMR, we will continue this discussion in Section 5.3.

5.2
Extrasolar Systems in Mean-motion Resonance

In this review we restrict ourselves to the extrasolar planetary systems in a configuration involving MMR. In such a configuration the orbital periods P (or the mean motions n) of the planets are in a ratio of two small integers. In the following we will denote the inner planet (of the resonant configuration under consideration) with a subscript "1" and the outer planet with subscript "2". The presence of a MMR is indicated if at least one of the so-called resonant angles librates such that they do not cover the full $(0 : 2\pi)$ domain (see Murray and Dermott [13] for details). For a general $p : q$ commensurability the resonant angles are defined in the planar case by

$$\Phi_{p,q,k} = p\lambda_2 - q\lambda_1 - p\varpi_2 + q\varpi_1 + k(\varpi_2 - \varpi_1) \tag{5.1}$$

where $\lambda_1, \lambda_2, \varpi_1$ and ϖ_2 denote the mean longitudes and the longitude of periapse for the inner (1) and outer (2) planet. The positive integers p and q satisfy $p > q$ and there are $p - q + 1$ possible k values with $q \leq k \leq p$. Of the $p - q + 1$ resonant angles at most two are linearly independent, i.e., for a resonant configuration either all angles librate or only one [14]. In the following we shall primarily concentrate on the 2:1 MMR which appears to be the configuration for at least four exoplanetary systems. In this case ($p = 2$ and $q = 1$) we have only two resonant angles ($k = 1, 2$) which we denote by

$$\Theta_i = 2\lambda_2 - \lambda_1 - \varpi_i \tag{5.2}$$

with $i \in \{1, 2\}$, and typically we replace the second angle Θ_2 by

$$\Delta\varpi = \varpi_2 - \varpi_1 = \Theta_2 - \Theta_1 \tag{5.3}$$

If the second resonant angle $\Delta\varpi$ is in libration, i.e., the periapses of the two orbits are basically aligned, we name the configuration to be in *apsidal co-rotation*.

A summary of the properties of all exoplanetary systems has been given recently by Butler et al. [15]. Among the 20 multi-planet systems detected up to now, about 30% are known to reside in a MMR. Table 5.1 gives an overview of the main orbital characteristics of the presently known resonant systems. The majority of them (four) are in a 2:1 MMR and the three others are in 3:1, 4:1 and 5:1, although the orbital data of the latter will have to be determined more accurately in the future. The best investigated system is GJ 876 which is also the first planetary system where a MMR between two planets has been discovered [16]. Due to the short orbital periods of the two outer planets ($\approx$ 30 and 60 days) many orbital periods have been observed, and the inferred orbital parameters are the most accurate of all the systems quoted in Table 5.1. The large-mass ratios in GJ 876 lead to a strongly interacting system, such that resonant interactions can be seen straight away, and the fitting procedure has

Table 5.1 Dynamical properties of planetary systems in mean-motion resonance (MMR). In the table we have only included those planets (denoted chronologically by 'b', 'c', ...) involved in the MMR, even though some systems (eg., GJ 876) may contain additional planets. Here P denotes the orbital periods, $M \sin i$ the mass of the planets, a the semi-major axis, e the eccentricity, ϖ the position of periastron, and M_* the stellar mass. The planetary mass and the distance a are derived quantities assuming the quoted stellar mass. The references quoted refer to the latest publication from which the listed data have been taken.

System	No	P [d]	$M \sin i$ $[M_{\mathrm{Jup}}]$	a [AU]	e	ϖ [deg]	M_* $[M_\odot]$	Reference
GJ 876	c	30.57	0.56	0.13	0.24	159	0.32	Laughlin et al. [17]
(2:1)	b	60.13	1.89	0.21	0.04	163		
HD 82943	b	219.4	1.85	0.75	0.38	124	1.15	Mayor et al. [18]
(2:1)	c	435.1	1.84	1.18	0.18	237		
55 Cnc	b	14.67	0.78	0.11	0.02	131	0.95	McArthur et al. [19]
(3:1)	c	43.93	0.21	0.24	0.44	244		
HD 128311	b	464	1.56	1.11	0.38	80	0.80	Vogt et al. [20]
(2:1)	c	910	3.08	1.73	0.21	21		
HD 202206	b	255	17.42	0.83	0.43	161	1.15	Correia et al. [21]
(5:1)	c	1383	2.43	2.54	0.26	55		
HD 108874	b	394	1.36	0.07	1.06	250	1.00	Butler et al. [15]
(4:1)	c	1600	1.02	0.25	2.68	20		
HD 73526	b	188	2.90	0.66	0.19	203	1.08	Tinney et al. [22]
(2:1)	c	377	2.50	1.05	0.14	13		

to be based on dynamical fits [23]. This system is clearly engaged in a 2:1 MMR with apsidal co-rotation where Θ_1 and ϖ are both librating around zero with small amplitudes (7 and 34°), see Lee et al. [24].

The orbital elements of the system HD 82943 are not so well determined, partly due to the longer periods and the Keplerian fits, and consequently the published orbital elements [18] lead to dynamically unstable solutions [25, 26]. However, the system nevertheless appears to be in a 2:1 resonance where either both resonant angles are librating around zero or one may be circulating [26].

The two recently discovered systems, HD 128311 and HD 73526, are dynamically very similar to each other. In both systems the planets are engaged in 2:1 mean motion resonances with a large temporal variation of at least one of the planetary eccentricities. In the first system apsidal co-rotation is broken with only θ_1 librating and $\Delta\varpi$ circulating [27]. In the latter case (HD 73526) a similar configuration has been suggested by Tinney et al. [22]. However, our dynamical fits also allow for apsidal co-rotation, see Section 5.5.2.

The system 55 Cnc has an additional two planets at larger and shorter distances which do not influence the intermediate planets engulfed in the 3:1 resonance. Additionally, in the system 47 UMa a 5:2 MMR has been suggested [28] and a stability analysis has been performed [29].

In the following we will concentrate on formation scenarios for systems in a 2:1 MMR and will comment briefly on higher order resonances as well.

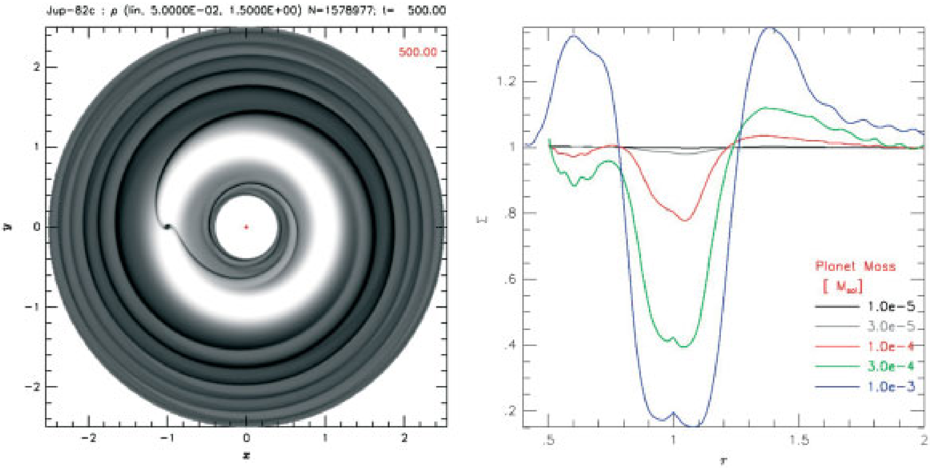
5.3
Planet–Disk Interaction

Possible mechanisms responsible to change the semi-major axis have to alter the orbital energy of the planets and hence must be dissipative, which requires a sort of frictional force acting on the planet. Based on the studies of satellite disk interaction with application to the Saturnian ring system [30], it was soon realized that a very similar processes must have operated between the gaseous protoplanetary disk and the embedded young protoplanet [31, 32]. Similarly, dissipative effects may occur for a planet embedded in a planetesimal disk [33]. In this report we will restrict ourselves to the interaction of a forming planet with its gaseous protoplanetary disk.

Here, the embedded planet disturbs the surrounding disk due to its gravitational interaction. The disturbances created in the disk (sound waves) superimpose and create, for small planetary masses, two wakes, or spiral waves in the ambient disk, the outer one trailing and the inner one leading [34]. Upon increasing the planetary mass the wake amplitudes increase and they transform into shock waves which lead to dissipation. As these are stationary in a coordinate system co-rotating with respect to the planet, the outer (inner) spiral is faster (slower) than the ambient disk which is consequently accelerated (decelerated) by the spiral shock waves. The result is that material is pushed away from the planetary radius towards smaller and larger radii which leads to the opening of a gap.

We performed two-dimensional (2D) hydrodynamical simulations for this gap-opening process of embedded planets on circular orbits in an initially axisymmetric viscous disk, with constant surface density, $\Sigma = 1$. For the simulations we solve for the 2D Navier–Stokes equations in cylindrical coordinates ($r - \varphi$). The general setup and equations are described in detail in Kley [35] and de Val-Borro et al. [36], and the numerical method is described in Kley [37, 38]. The computational domain covers the radial range from $r_{min} = 0.4$ up to $r_{max} = 2.5$ and in azimuth $\varphi_{min} = 0, \varphi_{max} = 2\pi$, which is tiled by $N_r = 256$ and $N_\varphi = 768$ equidistant gridcells. When placed in an axisymmetric disk the planet will create the spiral arms after a few orbital periods and the gap-opening process itself proceeds then on a longer (viscous) timescale. The back-reaction of the gas onto the planet is neglected in these simulations, i.e., the planet is not allowed to move, and it does not accrete any material from its surroundings.

The density distribution after 500 orbits of an embedded one-Jupiter-mass planet is shown in Fig. 5.1. Clearly seen are the spiral shocks created by the



(a)

(b)

Fig. 5.1 The results of two-dimensional hydrodynamical simulations with one embedded planet. (a) Gray scale plot of the two-dimensional density distribution for a Jupiter mass planet after 500 Orbits. The x and y-axes denote distances in units of $a_{Jup} = 5.2$ AU. The gray scale is linear in the density and the maximum and minimum are given in the header. White refers to low density and black to high density. (b) Azimuthally averaged density profile for different planet masses after 300 orbits.

planet (here two on each side), and the deep gap at the planetary orbit. In part (a) the azimuthally averaged surface density is shown as a function of planetary mass, quoted in units of the stellar mass. To open a significant deep gap with a depth of at least half of the unperturbed density a Saturn-mass planet ($1/3\ M_{Jup}$) is required. The exact depth and width of the gap region depends on a detailed balance of gravitational, viscous and pressure torques [39]. Increasing the planet mass deepens the gap, while increased viscosity and pressure tend to close the gap.

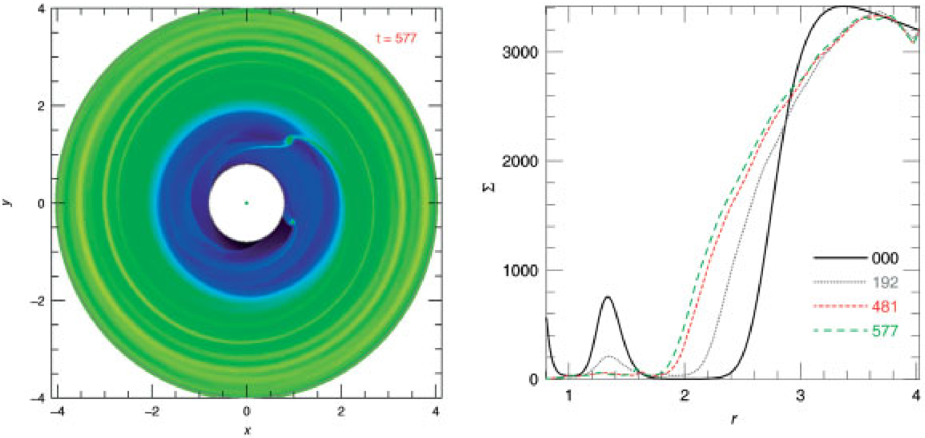
The nonaxisymmetric density enhancements in such a disk, the spiral arms, will pull gravitationally on the planet or, phrased differently, exert torques on it. As the sense of rotation of the planet in Fig. 5.1 is positive, or counter-clockwise, it is clear that the outer spiral pulls the planet back and slows it down, while the inner spiral will accelerate it. Hence, the outer disk will push the planet inwards and the inner, outwards. The detailed balance of the two contributions will determine the net force acting on the planet and determine its direction of motion through the disk. As a result the semi-major axis of the planet will change and lead to a radial migration of the planet through the disk [32]. Quite generally, in the case of the spiral arms the outer arm will pull more strongly and the planet will migrate inwards. It has been shown that, under typical conditions in the protoplanetary nebula, the timescale for a Jupiter-type planet to migrate from about 5 AU all the way towards the star is only about 10^5 years [40]. In addition, material orbiting at the same radius as the planet (so-called co-orbital material) will also exert torques on the planet. However, these will not play a role for sufficiently massive planets with a deep gap. The masses of the observed extrasolar planets in the 2:1 MMR are sufficiently large such that we expect that they created well cleared gaps and only the Lindblad torques (of the spiral arms) will be important in the evolution of the system. In any event, in a well-resolved hydrodynamic calculation all contributions will be taken care of automatically.

5.4
Resonant Capture

Having determined the acting agent, the protoplanetary disk, that might be responsible for moving planets around in their forming environment, we are now in a position to study in detail the evolution process of a pair of embedded planets in an accretion disk. We will analyze the capture process through full hydrodynamic simulations, forced N-body calculations and theoretical estimates.

5.4.1
Hydrodynamical Studies

In the case of two embedded planets in the disk there no longer exists a preferred reference frame in which the generated disturbances could be stationary. This will lead to a highly variable flow field in the disk, and two massive Jupiter-type planets will open a wide joint gap in the disk in which both planets will orbit [41, 42]. To demonstrate this effect we show here the results of numerical hydrodynamical simulations of two embedded planets in a disk starting at 1 and 2 AU on circular orbits, respectively. The masses of the two planets are 2.4 and 2.55 M_{Jup} orbiting a star of 1.08 $M_{\odot}$, where the parameters have been chosen to roughly match the system HD 73526. To allow for a sizable evolution during the simulations we have used a viscosity with a relatively large value of $\alpha = 0.01$ which may be on the large side for protoplanetary disks. In Fig. 5.2 we present a gray scale plot of the two-dimensional density distribution while in part (b), the azimuthally averaged density is shown at different times of the evolution. Clearly, between the planets a joint deep and wide gap is created after about 200–300 orbital periods of the planets. Due to the subsequent evolution *and* the inner outflow boundary condition the material inside of the two planets (in the inner disk) is cleared rapidly and both planets orbit in an empty cavity. As the inner planet has no disk material surrounding it, no torques are acting on it and there is no migration. The outer planet, on the other hand, is driven inward by the outer



(a) (b)

Fig. 5.2 (a) Gray scale plot of the density distribution of two embedded planets in a protoplanetary disk after 577 yrs, starting with an initially axisymmetric configuration. (b) Azimuthally averaged density distribution (dimensionless units) at four different times (in yrs).

disk and approaches the inner planet. We have reached the situation where the relative migration of the two planets is convergent.

When the outer planet approaches the 2:1 commensurability, its gravitational interaction is exciting periodically large perturbations in the orbit of the inner planet, which leads to an increase of the eccentricity and eventually to a capture in the 2:1 mean-motion resonance. This is illustrated in Fig. 5.3, where we show the change in semi-major axis and eccentricity of the two planets. The resonant capture at $t \approx 500$ yrs leads to an increase in the eccentricities of the two planets. For comparable planet masses the eccentricity of the outer planet will be smaller than that of the inner planet. At the end of this simulation the inner planet has reached $e = 0.4$ and the outer planet $e = 0.12$, values which are still increasing, due to the continued driving of the outer disk material. In this scenario, the final eccentricity of the outer planet (and subsequently of the inner) will be determined by the damping action of the ambient disk. In the case of a growing eccentricity the outer planet will periodically enter the disk material. The action of the disk will be to circularize the planet and damp the eccentricity. An equilibrium value for the eccentricity of the outer planet has been given by Snellgrove et al. [43] which depends for given planetary masses only on the ratio of migration timescale τ_{mig} and the circularization time scale τ_{circ}. A dispersing disk will stop the migration eventually and basically freeze the final configuration. A knowledge of the present configuration will give a rough estimate of the distance that the planets have been travelling through the disk.

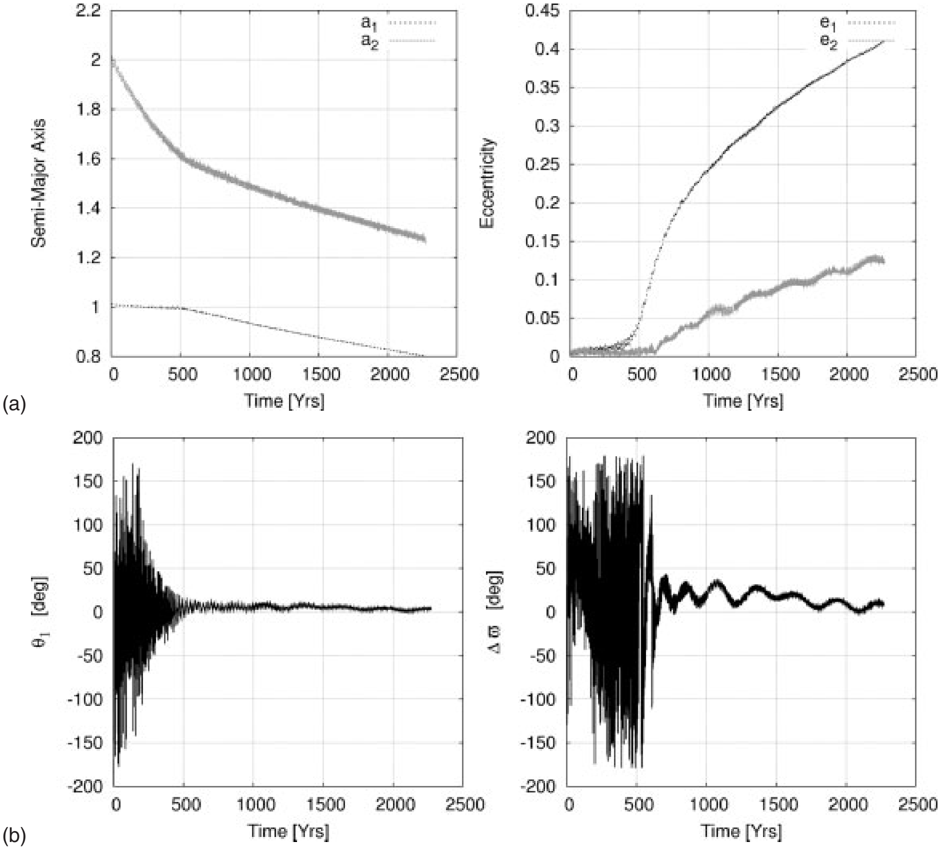


Fig. 5.3 (a) The semi-major axis (left) and eccentricity (right) evolution of two embedded planets in a protoplanetary disk. (b) The resonant angles Θ_1 (left) and $\Delta\varpi$ (right). At $t \approx 500$ yrs, capture in a 2:1 resonance occurs, the eccentricities increase, the orbits will align, and the planets migrate inwards simultaneously.

Upon capture, not only will the eccentricities increase, but the orbits will align to minimize dangerous close encounters. This is illustrated in the lower panel of Fig. 5.3 where Θ_1 and $\Delta\varpi$ are plotted as a function of time. Upon capture Θ_1 will librate first, followed later by $\Delta\varpi$ (or Θ_2). The most important results from hydrodynamical evolutions with two embedded planets is the fact that for converging inward migration of two massive planets which capture each other in a 2:1 mean-motion resonance, one always finds an alignment of the apsidal lines, i.e., apsidal co-rotation, where both resonant angles librate around zero [14, 43, 44]. As we will show below, this refers exactly to the observed state of GJ 873, and possibly to HD 82943.

5.4.2
Forced Migration

In the example above where full multi-dimensional hydrodynamical simulations have been performed, the heavy computational effort does not allow for detailed parameter studies. If required, those are typically performed applying a simplified approach where a standard three-body program (a star and two planets) is extended to include additional dissipative effects that mimic the migration and eccentricity damping induced by the disk. As has been shown above the end configuration will be a situation of two planets orbiting in a cavity surrounded by a disk. An additional forcing is applied in this case only to the outer planet. Such a forcing can be implemented directly as simple additional forces to the individual velocity components of the outer planets [14, 45], or by specifying the required damping rates τ_{mig} and τ_{ecc} of the semi-major axis and eccentricity for the outer planet and transform those into forces [24, 44]. This approach reduces the computational effort from weeks for one model to a few seconds. Typically the migration rate τ_{mig} is specified and then τ_{ecc} is given as a fixed fraction of the former as pioneered by Lee et al. [24].

$$\tau_{mig} = \frac{a}{\dot{a}} \qquad \text{and} \qquad \tau_{ecc} = \frac{e}{\dot{e}} = \frac{1}{K}\tau_{mig} \tag{5.4}$$

i.e., the eccentricity damping time-scale is a factor K shorter than the migration time. For small-mass planets where the interactions between the disk and the planets can be treated in the linear approximation, the value of K is roughly given by $(R/H)^2$ where R is the distance to the star and H the vertical thickness of the disk [46]. In our case of high-mass planets, the assumed parameter K can be calibrated by a comparison of such forced N-body models to detailed hydrodynamical simulations. This has been attempted in Kley et al. [44], where it has been found that K has to be chosen to lie between 1 and 10 for a good agreement with the hydrodynamics and forced three-body calculations, see also Fig. 5.4a. However, this low value for K is problematic when N-body results are compared to observations of GJ 876, the best analyzed system. Here, at least a value of $K \approx 100$ is required to keep the resonantly excited eccentricities sufficiently small [24].

In the hydrodynamical example, we started the planets directly beyond the 2:1 commensurability but still inside the 3:1 and 4:1 and found the standard result of adiabatic migration: capture in 2:1 and apsidal co-rotation. To investigate what happens if the initial orbits are further apart, we may utilize the simplified calibrated forced migration and perform detailed parameter studies, for example, for different initial separations. In Fig. 5.4b we present the outcome of resonant capture when two equal-mass planets ($m_1 = m_2$) are started at 4 and 12 AU from the star with initial eccentricities of $e = 0.02$, $K = 1$ and a damping time scale of $\tau_{mig} = 20\,000$ yrs. Below a certain mass threshold

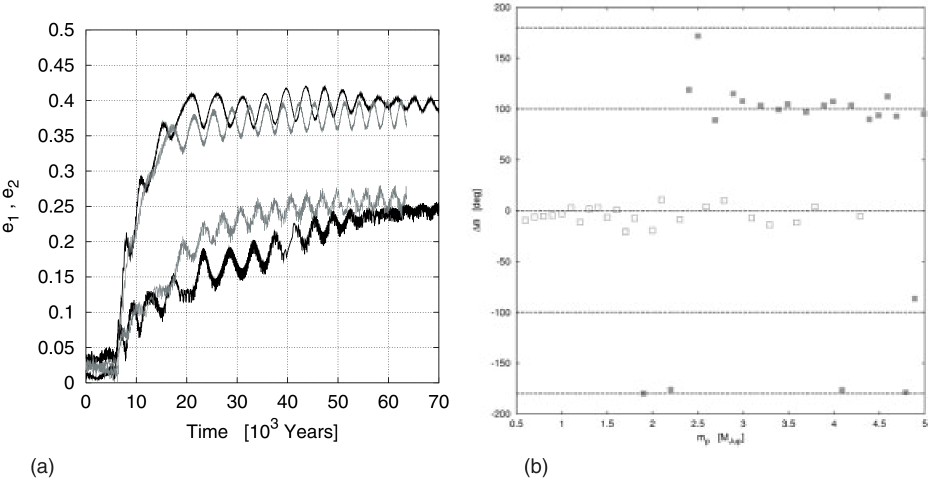


Fig. 5.4 (a) The eccentricity evolution of two massive planets that captured each other in 3:1 MMR for a full hydrodynamic simulations (black) and forced N-body (gray). A value of $K = 2.5$ has been used in the latter to obtain the fit. (b) The outcome of a parameter study to analyze the type of resonant capture for planets starting at $r = 4$ and 12 AU, respectively. Plotted is the resonant angle vs. the mass of the two planets (assumed to be of equal mass). Open squares denote capture in 2:1 and filled squares in 3:1 MMR. Note that all captures in 2:1 result in *aligned* apsidal co-rotation. Adapted from Kley et al. [44].

capture occurs exclusively in the 2:1 MMR. The other interesting feature seen in Fig. 5.4 relates to the second resonant angle $\Delta\varpi$. For capture in the 3:1 resonance we have an *asymmetric* configuration with $|\Delta\varpi| \approx 130^\circ$, whereas for the 2:1 capture we exclusively find a symmetric configuration with libration around $\Delta\varpi = 0$. This behavior in the 3:1 case depends also on the masses of the involved planets and has been analyzed using an averaged Hamiltonian approach by Beaugé et al. [47], who find exactly the same behavior.

5.4.3
Second Fundamental Model of Resonance

Capture into a kth-order resonance can be studied in the *second fundamental model of resonance*, which has been developed by Henrard et al. [48]. This model has proved to be very useful when studying the evolution of satellites under tidal interactions, or the resonant capture phenomenon.

In the following the Hamiltonian treatment of the resonant encounter will be summarized in the case of the 2:1 MMR. According to Murray and Dermott [13], the dominant part of the Hamiltonian in the case of two bodies, which are engulfed into a 2:1 resonance is

$$H = \alpha J + \beta J^2 + \varepsilon\sqrt{2J}\cos\theta \tag{5.5}$$

with the action-angle variables

$$J = \sqrt{\mu a}(1 - \sqrt{1 - e^2}) \qquad \theta = 2\lambda' - \lambda - \varpi \tag{5.6}$$

where a, e, λ, and ϖ are the semi-major axes, the eccentricity, the mean longitude and longitude of pericenter of the body under study, and $\mu = G^2(m_c + m)$ (G is the Gaussian gravitational constant). The primed quantities refer to the perturber. The coefficients in (5.5) are

$$\alpha = n^* - 2n^{*\prime} + \dot{\varpi} \qquad \beta = \frac{3}{2}\left[\frac{1}{ma^2} + \frac{4}{m'}a'^2\right] \qquad \varepsilon = |f_d| n^{\frac{3}{2}} \frac{m'a^2}{a'm_c} m^{\frac{1}{2}}, \tag{5.7}$$

where m, m', and m_c stand for the mass of the body under study, the perturber, and the central body, respectively, while f_d is related to the terms of the direct part of the perturbing function. Its numerical value exactly at the 2:1 resonance is $f_d = -1.19049\ldots$.

By introduction of new canonical variables

$$\Phi = \left(\frac{2\beta}{\varepsilon}\right)^{2/3} J \qquad \phi = \theta \tag{5.8}$$

the Hamiltonian (5.5) can be transformed to

$$H = -\frac{3}{2}(\delta + 1)(x^2 + y^2) + \frac{1}{4}(x^2 + y^2)^2 - 2x \tag{5.9}$$

where the Poincaré-like variables x and y are defined as

$$x = \sqrt{2\Phi}\cos\phi \qquad y = \sqrt{2\Phi}\sin\phi \tag{5.10}$$

The Hamiltonian (5.9) depends only on parameter δ, which is defined as

$$\delta = -\left(\frac{4\alpha^3}{27\beta\varepsilon^2}\right)^{1/3} - 1 \tag{5.11}$$

It is useful to recall that $\phi = 2\lambda' - \lambda - \varpi$ is the usual resonance variable θ and $\sqrt{2\Phi} \approx e$; i.e., in these variables the eccentricity is proportional to the distance between the origin and (x, y), see Eqs. (5.6) and (5.8). If ϕ librates around a mean value, the corresponding orbit is in a 2 : 1 resonance; when it circulates, the orbit is not in the resonance.

The resonant capture phenomenon can be understood by studying the phase portrait of the Hamiltonian (5.9), as shown in Fig. 5.5. When $\delta < 0$, there is one stable (elliptic-type) stationary point on the x-axis surrounded by phase trajectories, which in this case can be regarded as concentric circles (Fig. 5.5a). The radius of a circle corresponds to the eccentricity of the corresponding real orbit. Since the stationary point lies very close to the origin,

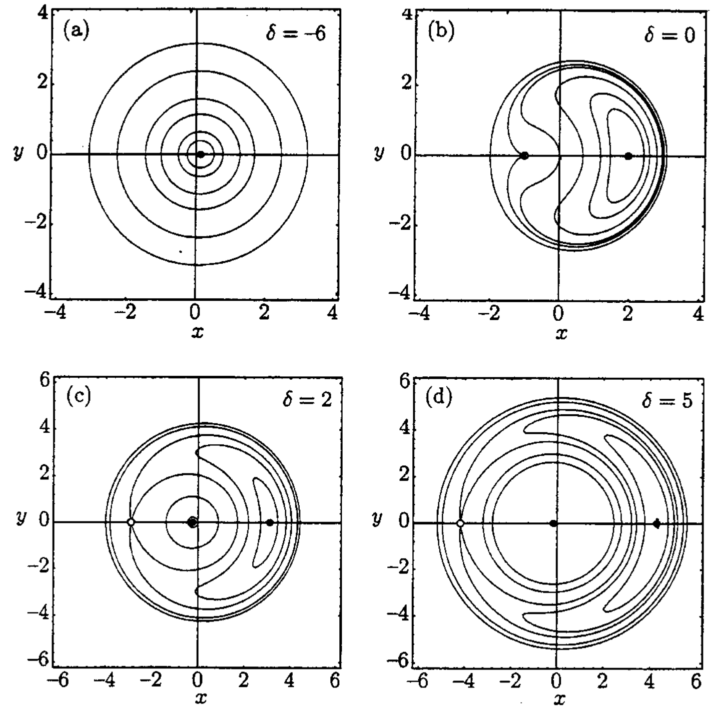


Fig. 5.5 The phase portrait of the Hamiltonian (5.9) for different values of the parameter δ. (Adapted from Murray and Dermott [13].)

the variable ϕ mostly circulates. However, for small values of e, libration is also possible. At $\delta = 0$, there appears a new stationary point on the x-axis at $x = -1$, which is the starting point of a separatrix separating libration from circulation (Fig. 5.5b). The already existing elliptic-type stationary point is shifted along the x-axis towards positive direction. Around this point the motion corresponds to libration. For $\delta > 0$ the stationary point that appears at $\delta = 0$ bifurcates into an elliptic and a hyperbolic point. Around the new elliptic point the motion is mainly circulation. However, for very small eccentricities, apocentric libration is also possible. The original elliptic point is shifted further into the positive direction. As δ increases, for smaller eccentricities the motion is circulation, and libration is possible for larger eccentricities. The two types of motion are separated by the separatrix starting from the hyperbolic stationary point, which is shifted towards the negative direction along the x-axis (Fig. 5.5c,d).

5.4.4
Outcome of Resonant Encounters

The main application of the second fundamental model of resonance is the study of *resonant encounters* that may occur in the course of the orbital evolution of two celestial bodies. As seen above, orbital evolution takes place in the case of migrating planets in the protoplanetary disk. The outcome of a resonant encounter can either be resonant capture or passing through a resonance without capture.

The Hamiltonian (5.9) depends only on the parameter δ. The action J is defined for a given value of δ as

$$J = \oint \Phi \, \mathrm{d}\phi = \oint x \, \mathrm{d}y \tag{5.12}$$

The variable J is the area enclosed by the phase trajectory on the $(\phi - \Phi)$ or $(x - y)$ planes. If, during one period of the phase trajectory, the variation of δ is negligible, J can be regarded as constant, and it is called the adiabatic invariant of the system. Then the area enclosed by the phase trajectory (J) is preserved, until it encounters the separatrix of the resonance. If the trajectory encounters the separatrix, the period of the trajectory tends to infinity, therefore the variation of δ is no longer negligible, meaning that J will no longer be an adiabatic invariant of the system. Thus when an orbit passes through a separatrix, the value of J changes suddenly. During its further evolution the new value of J is preserved.

In our case, the process of the resonant encounter is achieved through the migration of an outer planet towards the inner one when their mutual orbits are close to a mean-motion commensurability. Initially, the corresponding phase trajectory (of the inner planet) encloses a certain area. During the inward motion of the outer planet the shape of the trajectory is changing slowly (adiabatically) caused by the variation of δ, but the area enclosed by it is preserved. When the separatrix of the resonance gets close to the actual trajectory, a resonant encounter occurs, the outcome of which depends on the momentary conditions.

Since $\sqrt{2\Phi} \approx e$, there is a critical eccentricity e_c corresponding to the critical separatrix (at $\delta = 0$). Figure 5.6 shows the case when $e < e_c$. The shaded areas enclosed by the trajectories, corresponding to different δ, are the same. In Fig. 5.6c the critical separatrix ($\delta = 0$) appears. Since $e < e_c$, the area enclosed by the trajectory is smaller than the area enclosed by the critical separatrix, thus the orbit lies inside the separatrix, which means libration. Thus the orbit is captured into the 2:1 resonance. Then, by increasing δ, the orbit evolves according to Fig. 5.6 (d,e); its eccentricity increases, while the libration amplitude decreases, exactly as seen in the hydrodynamical simulations shown above.

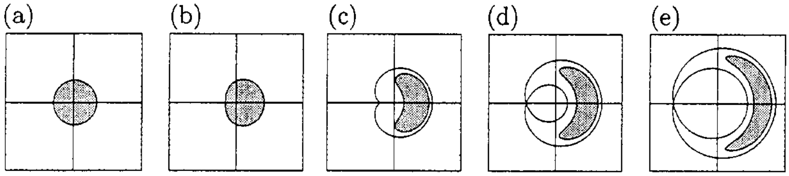


Fig. 5.6 Phase space of a resonant encounter, when $e < e_c$. The outcome of the encounter is resonant capture. (Adapted from Murray and Dermott [13]).

The critical eccentricity can be approximated as [13]

$$e_c = \sqrt{6}\left[\frac{3}{|f_d|}\frac{2^{\frac{2}{3}}m_c}{m'}\left(1+2^{\frac{2}{3}}\frac{m}{m'}\right)\right]^{-\frac{1}{3}} \tag{5.13}$$

which in the case of two Jupiter-mass planets and a solar-mass star is $e_c \approx 0.112$. Due to the eccentricity damping mechanism acting in protoplanetary disks, the eccentricities of the embedded giant planets are very small, therefore the condition $e_{inner} < e_c$ is typically satisfied when an inner giant is captured by an outer one migrating inwards.

If the initial eccentricity of the orbit is larger than the critical eccentricity, $e > e_c$, the result of the resonant encounter cannot be so easily forecasted. The area enclosed by the initial trajectory is larger than the area enclosed by the critical separatrix. By increasing δ, there exists a separatrix, which encloses an area equal to the area enclosed by the trajectory. On the other hand, if the trajectory encounters a separatrix, the action J changes suddenly. Either J changes to the area enclosed between the two branches of the separatrix, or changes to the area enclosed by the inner part of the separatrix. In the first case the orbit is captured into the 2:1 resonance and its eccentricity does not change, in the second case the trajectory passes through the resonance without being captured, while its eccentricity decreases suddenly. The more detailed treatment of the resonant capture, also including the cases of higher order resonances and capture probability estimates for $e > e_c$, can be found in [13].

5.5
Specific Systems

Now that we have laid out the theory of the capture process in the previous section, we will turn our attention to properties of real observed systems. We will demonstrate the evidence for a plain adiabatic migration and capture process for GJ 876, and will then show that, in explaining further systems such as

HD 128311 and HD 73526, one has to include short time perturbations to obtain a good match with the observations.

5.5.1
GJ 876: A Case of Adiabatic Migration

The first extrasolar system to be discovered which showed clear signs of a MMR was GJ 876 [16]. Here, two massive planets orbit a central star with orbits of $\approx$ 30 and 60 days. Later an additional close-in third planet (d) has been discovered [49] which does not, however, influence the dynamics of the resonant pair appreciably. The apsidal lines (of planets b, c) are aligned, and $\Delta\varpi$ and Θ_1 librate around zero with a amplitudes of about 34°, and only 7°, respectively [17]. In a detailed theoretical investigation Lee et al. [24] analyzed possible formation scenarios of the 2:1 resonant configuration in GJ 876. They performed a series of forced three-body simulations where, following the previously described hydrodynamical scenario, only the outer planet experiences a migration due to the ambient disk, and is slowly driven towards the inner planet. They used fixed constant values for the migration rate τ_{mig} and the eccentricity damping factor K as defined in Eq. (5.4), and found that the planets inevitably are captured into a 2:1 MMR in a symmetric configuration ($\Delta\varpi = 0$) and apsidal co-rotation. For sufficiently high values of the eccentricity damping they found that the eccentricities level off to constant values which are given solely by the magnitude of K and not by the migration time scale τ_{mig}. Interestingly, the rather small observed values of $e_1 = 0.23$ and $e_2 = 0.05$ for the inner and outer planet can only be attained in equilibrium for a quite large value of $K \approx 100$, i.e., through a strong eccentricity damping action of the ambient disk. In contrast, the aforementioned hydrodynamical simulations consistently yield much smaller values of K between 1 and 10 [44].

In a large parameter study using hydrodynamical as well as forced three-body simulations, Kley et al. [50] re-investigated formation scenarios of GJ 876. To possibly increase the eccentricity damping and avoid the problem of too small K values, a range of high viscosities, higher temperatures, and fully non-isothermal radiative disks have been investigated. However, none of these mechanisms are able to increase the eccentricity damping substantially. In the standard migration scenario the system is driven deep into the resonance where eccentricities continue to build up and the resonant angles converge to zero with only very small libration amplitude (very similar to Fig. 5.3). This leads to the conclusion that only a limited radial migration is allowed after the planets capture into their 2:1 resonance. Assuming that the resonant capture happened at the final stages in the formation process of GJ 876, the inclusion of a disk dispersal occurring on the viscous time scale resulted in values for the eccentricities which are consistent with the observa-

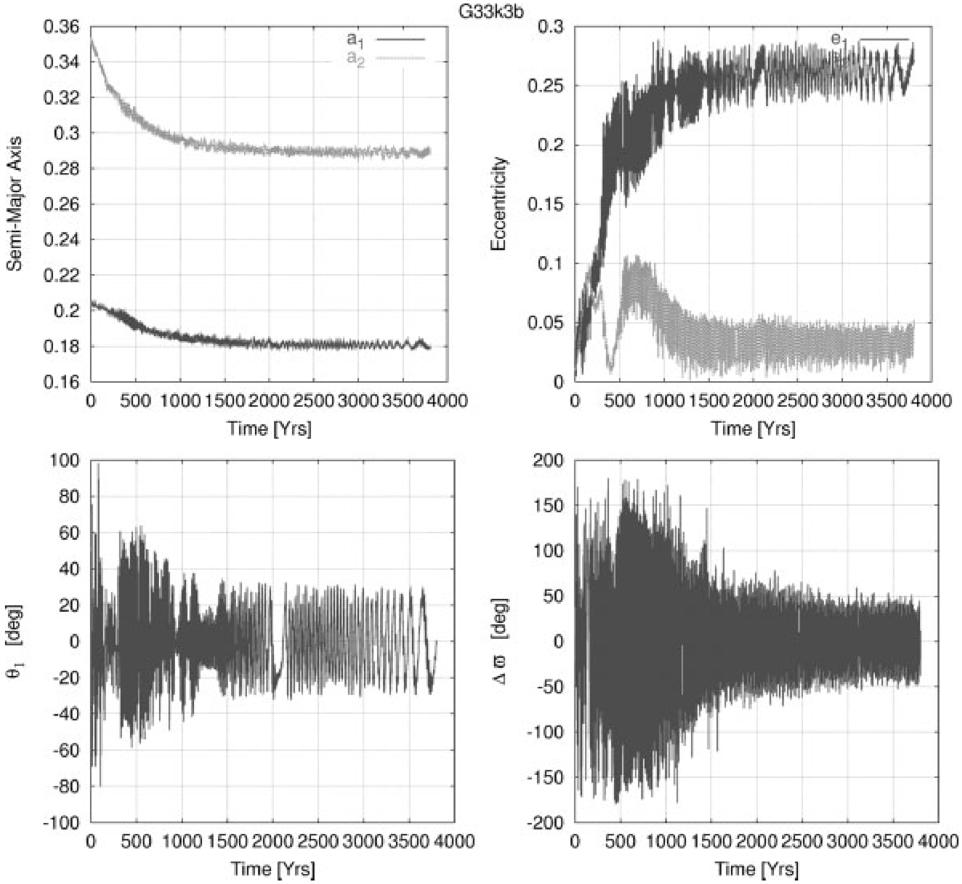


Fig. 5.7 The time evolution of the orbital elements $(a, e, \theta_1, \Delta\varpi)$ of a hydrodynamical evolution with two embedded planets for the system GJ 876. The disk has been chosen to be locally isothermal and a model of disk dispersal has been included. The mass ratios of the two planets are $q_1 = 1.75 \times 10^{-3}$ and $q_2 = 5.9 \times 10^{-3}$. (After Kley et al. [50]).

tions (see Fig. 5.7). Hence Kley et al. [50] concluded that the present resonant structure of GJ 876 clearly points towards resonant capture through an adiabatic migration process with resonant capture which has taken place during the final stages of the formation process.

5.5.2
Formation of Systems HD 128311 and HD 73526 through Mixed Scenarios

Recent radial velocity measurements suggest that the systems HD 128311 and HD 73526 contain pairs of giant planets engaged in a 2:1 protective MMR, [20, 22]. Numerical integrations, based on the published orbital data (see Table 5.2 for HD 128311 and Fit 0 of Table 5.3 for HD 73526) show that the giant planets in both systems are in a 2:1 resonance without apsidal co-rotation of the corresponding orbits. Moreover, the eccentricities of the giant planets show large-amplitude oscillations, being regular in the case of HD 128311, and irregular (chaotic) in the case of HD 73526 (see Fig. 5.8).

Table 5.2 Orbital data of HD 128311 as provided by Vogt et al. [20]. Here M denotes the mean anomaly.

Planet	Mass [M_J]	a [AU]	e	M [deg]	ϖ [deg]
b	1.56	1.109	0.38	257.6	80.1
c	3.08	1.735	0.21	166.0	21.6

Table 5.3 Dynamical orbit fits of the giant planets around HD 73526; Fit 0 corresponds to the published results of Tinney et al. [22], while Fit 1 belongs to one of new orbital solutions resulting in regular behavior of the giant planets as given by Sándor et al. [51].

Fit	Planet	Mass [M_J]	a [AU]	e	M [deg]	ϖ [deg]
0	b	2.9	0.66	0.19	86	203
	c	2.5	1.05	0.14	82	13
1	b	2.415	0.659	0.26	70.7	202.9
	c	2.55	1.045	0.107	170.7	253.7

In both cases the resonant angle Θ_1 librates around 0° with an amplitude $\sim 60^\circ$ (HD 128311), and $\sim 90^\circ$ (HD 73526), respectively. The lack of co-rotation in these system means that both Θ_2 and $\Delta\varpi$ circulate.

Additional stability investigations performed by the means of the Relative Lyapunov Indicator [52] show that the system around HD 73526 is located in a weakly chaotic region of the phase space [51]. On the other hand, radial velocity measurements to date also allow co-planar dynamic orbit fits for the system yielding regular behavior. One of the new fits, which we have been obtained by using the Systemic Console (http://www.oklo.org, generated by Greg Laughlin), is presented in Table 5.3 (Fit 1). The reason for looking for these orbit fits is that chaotic behavior is uncommon among the extrasolar planetary systems, and even a weak chaos may not guarantee an arbitrary long-term stability for the system. Additionally, as will be presented in the

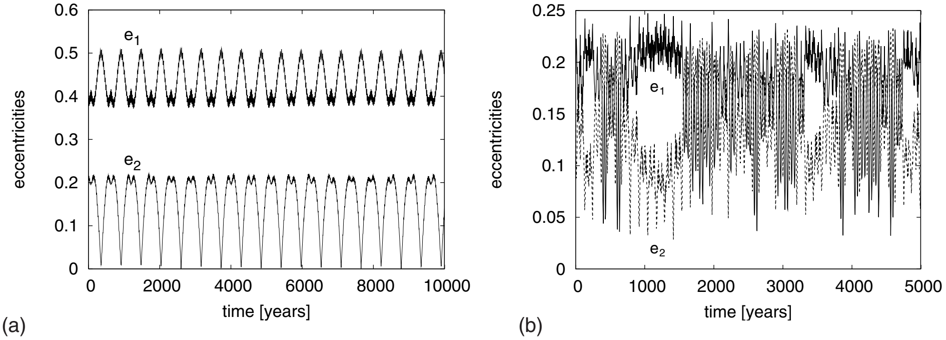


Fig. 5.8 Time evolution of the eccentricities of giant planets in the system HD 128311 (after Sándor et al. [27]) (a) and HD 73526 (b) obtained by numerical integration using initial conditions of Tables 5.2, and 5.3 (Fit 0), respectively.

following, the different formation scenarios clearly favor the regular orbital behavior of the giant planets [51]. Numerical integrations based on the newly derived orbital data exhibit very similar behavior of the eccentricities of the planets in HD 73526 (see Fig. 5.9) to the planets in HD 128311. The only difference between the dynamics of the systems is that, in the majority of the orbit fits, the giant planets of HD 73526 are in apsidal co-rotation, but with enlarged amplitudes in Θ_2 and $\Delta\varpi$.

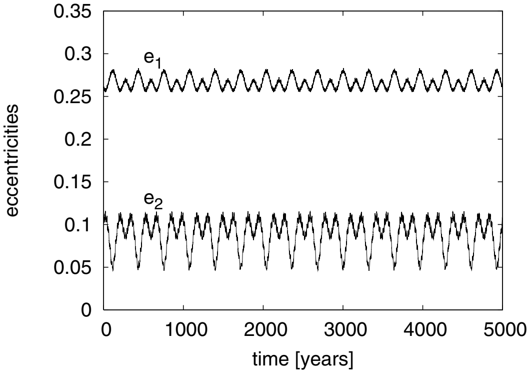


Fig. 5.9 Regular behavior of the eccentricities e_1 and e_2 of the giant planets in HD 73526 by using the initial orbital data given by Fit 1.

Inward Migration

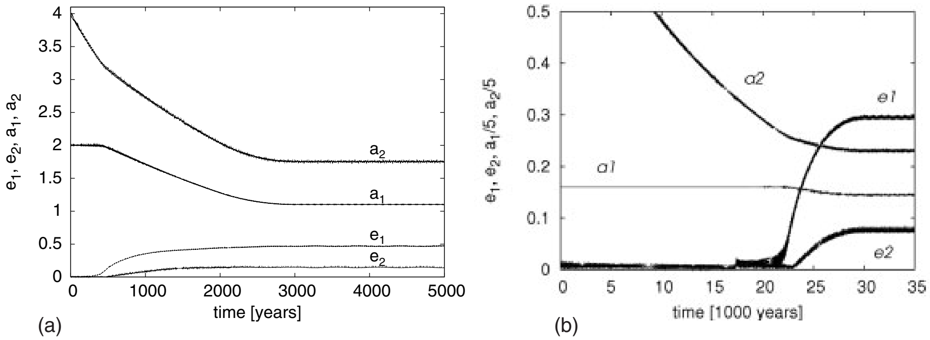
According to Lee et al. [24], in the case of a sufficiently slow migration, a planetary system engulfed into a 2:1 MMR should also be in apsidal co-rotation, and the eccentricities of the giant planets should yield only very small variations. Clearly, the orbital behavior of the giant planets around HD 128311 and HD 73526 violate this result, since (1) in the case of HD 128311 the orbits of the giant planets are not in apsidal co-rotation; (2) in both systems the eccentricities exhibit considerable variations. However, the presence of the resonance indicates that these systems may have, in the past, gone through an adiabatic migration process similar to GJ 876.

In order to resolve the above discrepancy, *mixed* evolutionary scenarios of migrating planetary systems have been suggested incorporating migration and other additional perturbative effects, such as sudden termination of migration, or a planet–planet scattering event [27].

As shown above, the migration of a single planet can be characterized by the migration rate $\dot{a}/a$ and the eccentricity damping rate $\dot{e}/e$, or by the corresponding e-folding times for the semi-major axes and eccentricities: τ_a and τ_e (τ_{mig} and τ_{ecc} above), respectively. It is important to recall that investigating the system GJ 876, Lee et al. [24] have found that, for a sufficiently slow migration, the final state of the system depends only on the ratio K of the e-folding times $K = \tau_a/\tau_e$.

The formation of HD 128311 has also been studied by using a migratory scenario [27]. According to the standard formation scenario, only the outer planet migrated inward, and after some time a resonant capture took place between the planets. Having reached the present values of the semi-major axes, the migration has been slowed down and stopped. This corresponds to the smooth dispersal of the protoplanetary disk. After this migration process, the system is locked into a deep 2:1 resonance, the resonant angles $\Theta_{1,2}$, and $\Delta\varpi$ librate around 0° with small amplitudes. The eccentricities of the giant planets e_1 and e_2 are almost constant (the index 1 refers to the inner and the index 2 to the outer planet). For $K = 5$, Sándor et al. [27] have found $e_1 = 0.46$ and $e_2 = 0.15$, see Fig. 5.10a. Smaller/larger values of K always yield systems deep in resonance and result in larger/smaller $e_{1,2}$, contradicting the observed behavior, shown in Fig. 5.8a.

Modeling the formation of HD 73526 by an inward migration of the outer giant planet, it has been found that, similar to the system GJ 876, the eccentricity of the outer planet needs a considerably damping $K \sim 15$–20 [51]. This ratio again exceeds $K \sim 1$ suggested by the hydrodynamical simulations [44]. Moreover, after the capture into the $2:1$ resonance during the migration process, the eccentricity of the inner planet is slightly increasing. Similar to HD 128311, the migration of the outer planet gradually stopped when the semi-major axes of the giant planets reached their actual values. The use of the relative high damping ratio K can be avoided in the case of HD 73526, if the resonant capture takes place just before the dispersal of the protoplanetary disk. This scenario has been proposed by Kley et al. [50] in the case of GJ 876,



(a)

(b)

Fig. 5.10 (a) The behavior of the semi-major axes and eccentricities during an inward migration of a model for HD 128311 with $\tau_a = 2 \times 10^3$ years and $K = 5$ (after Sándor et al. [27]). The migration is stopped between 2×10^3 and 3×10^3 years by applying a linear reduction. (b) The same for HD 73526 with $\tau_a = 2 \times 10^4$ years and $K = 5$. The disk's dispersion also obeys a linear reduction law between 2.5×10^4 and 3×10^4 years. In order to visualize the behavior of the eccentricities and semi-major axes together, the values of a_1 and a_2 are reduced by a factor 5.

and also applied to HD 73526 [51]. To delay the resonant capture, the inner giant planet moves initially at $a = 0.8$ AU, while the outer planet migrates inward. After the resonant capture, the eccentricity of the inner planet grows rapidly (see Fig. 5.10b), but its migration is gradually terminated, thus the inner planet's eccentricity does not exceed the limit provided by numerical integration based on the observation (which is $e_1 \leq 0.3$ in the case of HD 73526). During the above scenario, the damping ratio is only $K = 5$, which is a realistic value. On the other hand, the presence of a damping mechanism on the eccentricity of the inner planet cannot be excluded. However, this effect has not yet been studied in detail.

Clearly, the present behavior of the systems HD 128311 and HD 73526 is not the result of such an adiabatic inward migration process alone. Sándor et al. [27] have suggested two additional mechanisms, which may be responsible for the large oscillations of the eccentricities and breaking the apsidal co-rotation. These mechanisms are the sudden stopping of migration and scattering of one of the giant planets with an already existing small-mass planet which we shall briefly discuss now in turn.

Sudden Stopping of Migration

The first mechanism, which can induce large-amplitude variations of the eccentricities and the resonant angles is the sudden termination of the migration. Recent *Spitzer* observations of young stars show that the inner part of some protoplanetary disks may practically be emptied possibly due to photoevaporation induced by the central star [53, 54]. Thus, upon approaching the

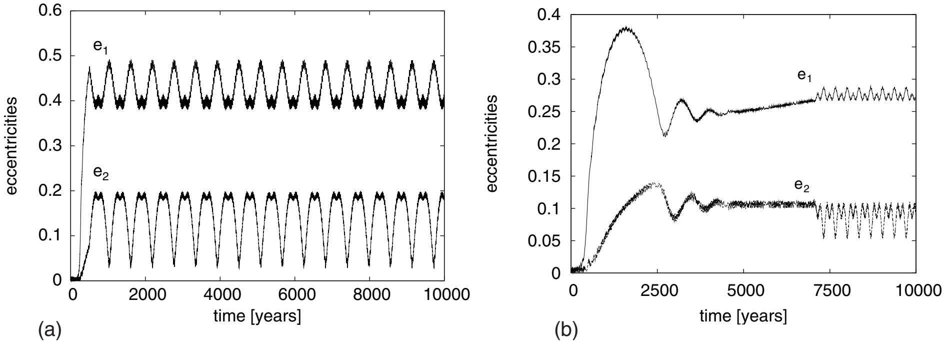


Fig. 5.11 The behavior of the eccentricities obtained by a sudden stopping of the migration of the outer planet. (a) (adapted from Sándor et al. [27]), corresponds to HD 128311, (b) to HD 73526.

inner rim of such a disk, the inward migration of a planet can be terminated rapidly.

In order to model this type of scenario in the case of HD 128311, Sándor et al. [27] have performed numerical simulations where the migration of the outer planet has been terminated abruptly upon reaching the actual value of its semi-major axis ($a_2 = 1.73$ AU). The inner planet is assumed to orbit in the empty region of the disk at $a_1 = 1.5$ AU, and the outer planet is started from $a_2 = 4$ AU forcing it to migrate inward very fast with a moderate damping of the eccentricity ($\tau_a = 500$ years, $K = 10$). It has been found that after the sudden stopping of the migration, the eccentricities of the giant planets behave in a very similar way to the observed case (compare Figs. 5.11a and 5.8a). The above described sudden stopping of the migration has not resulted in the breaking of the apsidal co-rotation. However, the libration amplitudes of the resonant angles have been increased considerably.

In a similar way to HD 128311, the behavior of the eccentricities of the giant planets in the system HD 73526 can also be modeled by a relatively fast termination of the migration. This again corresponds to a relatively fast dispersion of the protoplanetary disk. However, in the case of HD 73526 we have found that, contrary to HD 128311, where the migration has stopped abruptly, the disk dispersal time scale could be longer, $\Delta t_{\text{dis}} = 100$ yr. In order to model the present behavior of the system, a migration characterized by $K = 15$ and $\tau_a = 2 \times 10^3$ years has been assumed. After the termination of migration the eccentricities (displayed in Fig. 5.11b) and resonant angles of the giant planets, behave in a very similar way to the case of Fit 1 of Table 5.3, shown in Fig. 5.9.

Planet–Planet Scattering

The behavior of the eccentricities of the giant planets of the system HD 128311 and HD 73526 exhibit very similar behavior to that observed in the system around υ Andromedae. Ford et al. [55] suggested that such behavior in υ is most likely the result of a planet–planet scattering event. The present behavior of HD 128311 and HD 73526 have been also modeled by such an effect, i.e., one of the giant planets suffers a close encounter with an already existing smaller mass planet ($m_p = 10\ \mathrm{M}_\oplus$) [27, 51].

The most effective scattering occurs when a small-mass planet orbits close to the central star in a nearly circular orbit, and a pair of giant planets already locked into a 2:1 resonance migrate inward approaching their present positions. The giant planets perturb the motion of the inner planet, and most likely a resonant capture (into a 2 : 1 resonance) takes place between the inner giant planet and the small mass planet resulting in its eccentricity pumping during its forced migration. Thus it is very likely that the inner small-mass planet, is destabilized and practically ejected from the system by a close encounter with one of the giant planets. In the case of HD 128311 the perturbation induced by a close encounter with a small-mass planet could have been strong enough to break the apsidal resonance between the giant planets, as has been demonstrated by Sándor et al. [27] (Fig. 5.12). In the case of HD 73526, only the libration amplitude of the resonant angles have been increased, the giant planets remained in apsidal co-rotation. After the scattering event in both systems the variation of the eccentricities of the giant planets (see Fig. 5.13) exhibit very similar behavior to the observed ones.

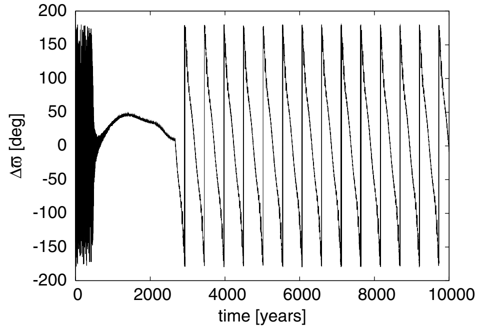


Fig. 5.12 The breaking of the apsidal co-rotation due to a scattering event induced by a small-mass inner planet in the system of HD 128311. (After Sándor et al. [27].)

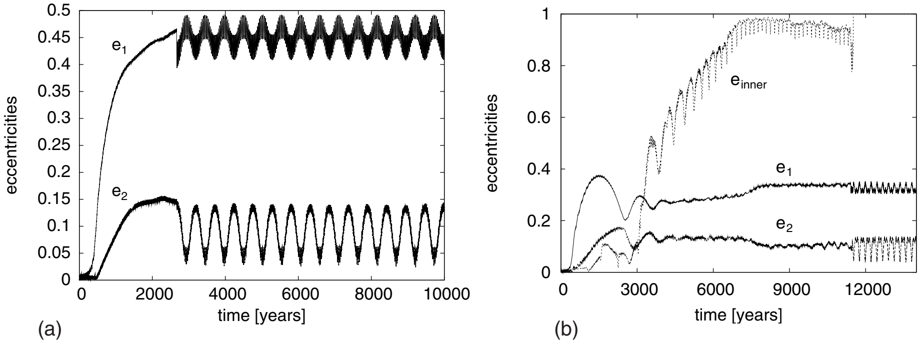


Fig. 5.13 Time evolution of the eccentricities of the giant planets before and after a scattering event with an inner low-mass planet. (a) (adapted from Sándor et al. [27]), corresponds to HD 128311, (b) to HD 73526. In the latter case the behavior of the eccentricity of the small-mass inner planet (which reaches values close to 1 before ejection) is also shown.

5.6
Summary

In this review we have studied possible formation scenarios of the observed mean-motion resonances in extrasolar planetary systems. As the majority of resonant systems are locked in a 2:1 resonance we concentrate in particular on these systems.

We have shown that planet–disk interaction between young embedded protoplanets and the protoplanetary disk may usually lead to a differential migration where the outer planet approaches the inner one. To analyze the outcome of the convergent migration we first perform full hydrodynamical simulations with two embedded planets and then compare these to forced three-body simulations where additional forces mimicking the planet–disk interaction have been added. We find that resonant capture into the 2:1 resonance indeed occurs frequently and that continued driving of the disk will move the system deep into the resonance with significant growth of the eccentricity of the two planets. Additionally, the final configuration of the orbits is symmetric where the periapses are aligned with only small libration. We refer to this state as apsidal co-rotation.

The well observed system GJ 876 is, in fact, in this symmetric configuration with only small libration of the two resonant angles. The relatively small planetary eccentricities in this system can only be explained if one assumes that after resonant capture occurred the length of migration was limited by disk dispersal. Alternatively, it may be possible that an additional damping mechanism has operated on the eccentricities during their migration process.

In the other two systems analyzed (HD 128311 and HD 73526) the eccentricities in the current observed state vary much more strongly and vanish periodically for one planet. We demonstrate that the dynamical structure of these two systems may be the results of the mixed evolutionary scenario, melting together inward migration and a sudden perturbation.

In summary, the complex dynamical behavior of the observed resonant planetary systems can be understood in the framework of their formation process. Interaction with the protoplanetary disk and possibly more sudden planet–planet scattering events are important ingredients in shaping planetary systems.

Acknowledgments

Part of this work has been supported by the Hungarian Scientific Research Fund (OTKA) under the grants D048424, T043739 and by the German Science Foundation (DFG) under grant 436 UNG 17/1/07.

References

1 Peale, S. J.: 1976, Orbital resonances in the solar system, *ARAA*, **14**, 215–246.

2 Goldreich, P.: 1965, An explanation of the frequent occurrence of commensurable mean motions in the solar system, *MNRAS*, **130**, 159.

3 Greenberg, R. J., Counselman, III, C. C., and Shapiro, I. I.: 1972, Orbit–orbit resonance capture in the solar system, *Science*, **178**, 747–749.

4 Greenberg, R.: 1973, Evolution of satellite resonances by tidal dissipation, *AJ*, **78**, 338.

5 Malhotra, R.: 1993, The origin of Pluto's peculiar orbit, *Nature*, **365**, 819.

6 Malhotra, R.: 1995, The origin of Pluto's orbit: implications for the solar system beyond Neptune, *AJ*, **110**, 420.

7 Roy, A. E. and Ovenden, M. W.: 1954, On the occurrence of commensurable mean motions in the solar system, *MNRAS*, **114**, 232.

8 von Weizsäcker, C. F.: 1948, Zur Kosmogonie, *Zeitschrift fur Astrophysik*, **24**, 181.

9 Patterson, C. W.: 1987, Resonance capture and the evolution of the planets, *Icarus*, **70**, 319–333.

10 Beaugé, C., Aarseth, S. J., and Ferraz-Mello, S.: 1994, Resonance capture and the formation of the outer planets, *MNRAS*, **270**, 21.

11 Cameron, A. G. W.: 1978, Physics of the primitive solar accretion disk, *Moons and Planets*, **18**, 5–40.

12 Mizuno, H.: 1980, Formation of the giant planets, *Progress of Theoretical Physics*, **64** 544–557.

13 Murray, C. D. and Dermott, S. F.: 1999, Solar system dynamics, in *Solar system dynamics by Murray, C. D.*

14 Nelson, R. P. and Papaloizou, J. C. B.: 2002, Possible commensurabilities among pairs of extrasolar planets, *MNRAS*, **333**, L26–L30.

15 Butler, R. P., Wright, J. T., Marcy, G. W., Fischer, D. A., Vogt, S. S., Tinney, C. G., Jones, H. R. A., Carter, B. D., Johnson, J. A., McCarthy, C., and Penny, A. J.: 2006, Catalog of nearby exoplanets, *ApJ*, **646**, 505–522.

16 Marcy, G. W., Butler, R. P., Fischer, D., Vogt, S. S., Lissauer, J. J., and Rivera, E. J.: 2001, A pair of resonant planets orbiting GJ 876, *ApJ*, **556**, 296–301.

17 Laughlin, G., Butler, R. P., Fischer, D. A., Marcy, G. W., Vogt, S. S., and Wolf, A. S.: 2005, The GJ 876 planetary system: a progress report, *ApJ*, **622**, 1182–1190.

18 Mayor, M., Udry, S., Naef, D., Pepe, F., Queloz, D., Santos, N. C., and Burnet, M.: 2004, The CORALIE survey for southern extrasolar planets. XII. Orbital solutions for 16 extrasolar planets discovered with CORALIE, *A&A*, **415**, 391–402.

19 McArthur, B. E., Endl, M., Cochran, W. D., Benedict, G. F., Fischer, D. A., Marcy, G. W., Butler, R. P., Naef, D., Mayor, M., Queloz, D., Udry, S., and Harrison, T. E.: 2004, Detection of a Neptune-mass planet in the ρ^1 Cancri system using the Hobby–Eberly telescope, *ApJ*, **614**, L81–L84.

20 Vogt, S. S., Butler, R. P., Marcy, G. W., Fischer, D. A., Henry, G. W., Laughlin, G., Wright, J. T., and Johnson, J. A.: 2005, Five new multicomponent planetary systems, *ApJ*, **632**, 638–658.

21 Correia, A. C. M., Udry, S., Mayor, M., Laskar, J., Naef, D., Pepe, F., Queloz, D., and Santos, N. C.: 2005, The CORALIE survey for southern extrasolar planets. XIII. A pair of planets around HD 202206 or a circumbinary planet?, *A&A*, **440**, 751–758.

22 Tinney, C. G., Butler, R. P., Marcy, G. W., Jones, H. R. A., Laughlin, G., Carter, B. D., Bailey, J. A., and O'Toole, S.: 2006, The 2:1 resonant exoplanetary system orbiting HD 73526, *ApJ*, **647**, 594–599.

23 Laughlin, G. and Chambers, J. E.: 2001, Short-term dynamical interactions among extrasolar planets, *ApJ*, **551**, L109–L113.

24 Lee, M. H. and Peale, S. J.: 2002, Dynamics and origin of the 2:1 orbital resonances of the GJ 876 planets, *ApJ*, **567**, 596–609.

25 Ferraz-Mello, S., Michtchenko, T. A., and Beaugé, C.: 2005, The orbits of the extrasolar planets HD 82943c and b, *ApJ*, **621**, 473–481.

26 Lee, M. H., Butler, R. P., Fischer, D. A., Marcy, G. W., and Vogt, S. S.: 2006, On the 2:1 orbital resonance in the HD 82943 planetary system, *ApJ*, **641**, 1178–1187.

27 Sándor, Z. and Kley, W.: 2006, On the evolution of the resonant planetary system HD 128311, *A&A*, **451**, L31–L34.

28 Naef, D., Mayor, M., Beuzit, J. L., Perrier, C., Queloz, D., Sivan, J. P., and Udry, S.: 2004, The ELODIE survey for northern extrasolar planets. III. Three planetary candidates detected with ELODIE, *A&A*, **414**, 351–359.

29 Psychoyos, D. and Hadjidemetriou, J. D.: 2005, Dynamics of extrasolar systems at the 5/2 resonance: application to 47 UMa, *Dynamics of Populations of Planetary Systems*, **IAU Colloq. 197**, 55–62.

30 Goldreich, P. and Tremaine, S.: 1980, Disk–satellite interactions, *ApJ*, **241**, 425–441.

31 Lin, D. N. C. and Papaloizou, J.: 1986, On the tidal interaction between protoplanets and the protoplanetary disk. III – Orbital migration of protoplanets, *ApJ*, **309**, 846–857.

32 Ward, W., R.: 1997, Protoplanet migration by nebula tides, *Icarus*, **126** 261–281.

33 Murray, N., Hansen, B., Holman, M., and Tremaine, S.: 1998, Migrating planets, *Science*, **279**, 69.

34 Ogilvie, G. I. and Lubow, S. H.: 2002, On the wake generated by a planet in a disc, *MNRAS*, **330**, 950–954.

35 Kley, W.: 1999, Mass flow and accretion through gaps in accretion discs, *MNRAS*, **303**, 696–710.

36 de Val-Borro, M., Edgar, R. G., Artymowicz, P., Ciecielag, P., Cresswell, P., D'Angelo, G., Delgado-Donate, E. J., Dirksen, G., Fromang, S., Gawryszczak, A., Klahr, H., Kley, W., Lyra, W., Masset, F., Mellema, G., Nelson, R. P., Paardekooper, S.-J., Peplinski, A., Pierens, A., Plewa, T., Rice, K., Schäfer, C., and Speith, R.: 2006, A comparative study of disc-planet interaction, *MNRAS*, **370**, 529–558.

37 Kley, W.: 1989, Radiation hydrodynamics of the boundary layer in accretion disks. I – Numerical methods, *A&A*, **208**, 98–110.

38 Kley, W.: 1998, On the treatment of the Coriolis force in computational astrophysics, *A&A*, **338**, L37–L41.

39 Crida, A., Morbidelli, A., and Masset, F.: 2006, On the width and shape of gaps in protoplanetary disks, *Icarus*, **181**, 587–604.

40 Nelson, R. P., Papaloizou, J. C. B., Masset, F., and Kley, W.: 2000, The migration and growth of protoplanets in protostellar discs, *MNRAS*, **318**, 18–36.

41 Bryden, G., Różyczka, M., Lin, D. N. C., and Bodenheimer, P.: 2000, On the interaction between protoplanets and protostellar disks, *ApJ*, **540**, 1091–1101.

42 Kley, W.: 2000, On the migration of a system of protoplanets, *MNRAS*, **313**, L47–L51.

43 Snellgrove, M. D., Papaloizou, J. C. B., and Nelson, R. P.: 2001, On disc driven inward migration of resonantly coupled planets with application to the system around GJ876, *A&A*, **374**, 1092–1099.

44 Kley, W., Peitz, J., and Bryden, G.: 2004, Evolution of planetary systems in resonance, *A&A*, **414**, 735–747.

45 Beaugé, C., Michtchenko, T. A. and Ferraz-Mello, S.: 2006, Planetary migration and extrasolar planets in the 2/1 mean-motion resonance, *MNRAS*, **365**, 1160–1170.

46 Tanaka, H. and Ward, W. R.: 2004, Three-dimensional interaction between a planet and an isothermal gaseous disk. II. Eccentricity waves and bending waves, *ApJ*, **602**, 388–395.

47 Beaugé, C., Ferraz-Mello, S., and Michtchenko, T. A.: 2003, Extrasolar planets in mean-motion resonance: apses alignment and asymmetric stationary solutions, *ApJ*, **593**, 1124–1133.

48 Henrard, J. and Lemaitre, A.: 1983, A second fundamental model for resonance, *Cel. Mech.*, **30**, 197–218.

49 Rivera, E. J., Lissauer, J. J., Butler, R. P., Marcy, G. W., Vogt, S. S., Fischer, D. A., Brown, T. M., Laughlin, G., and Henry, G. W.: 2005, A $\sim 7.5 M_\oplus$ planet orbiting the nearby star, GJ 876, *ApJ*, **634**, 625–640.

50 Kley, W., Lee, M. H., Murray, N., and Peale, S. J.: 2005, Modeling the resonant planetary system GJ 876, *A&A*, **437**, 727–742.

51 Sándor, Z., Kley, W., and Klagyivik, P.: 2007, On the formation of resonant planetary system HD 73526, *A&A*, `2007arXiv0706.2128S` **in press**.

52 Sándor, Z., Érdi, B., Széll, A., and Funk, B.: 2004, The relative Lyapunov indicator: an efficient method of chaos detection, *CMDA*, **90**, 127–138.

53 Calvet, N., D'Alessio, P., Watson, D. M., Franco-Hernández, R., Furlan, E., Green, J., Sutter, P. M., Forrest, W. J., Hartmann, L., Uchida, K. I., Keller, L. D., Sargent, B., Najita, J., Herter, T. L., Barry, D. J., and Hall, P.: 2005, Disks in transition in the Taurus population: Spitzer IRS spectra of GM Aurigae and DM Tauri, *ApJ*, **630**, L185–L188.

54 D'Alessio, P., Hartmann, L., Calvet, N., Franco-Hernández, R., Forrest, W. J., Sargent, B., Furlan, E., Uchida, K., Green, J. D., Watson, D. M., Chen, C. H., Kemper, F., Sloan, G. C., and Najita, J.: 2005, The truncated disk of CoKu Tau/4, *ApJ*, **621**, 461–472.

55 Ford, E. B., Lystad, V., and Rasio, F. A.: 2005, Planet–planet scattering in the upsilon Andromedae system, *Nature*, **434**, 873–876.

6
Impact of Stellar Activity on the Evolution of Planetary Atmospheres and Habitability

Helmut Lammer, Maxim L. Khodachenko, Herbert I. M. Lichtenegger, and Yuri N. Kulikov

Abstract

In order to understand the principles that generated earth's long-time habitable environment, compared with other terrestrial planets like Venus and Mars and possible terrestrial exoplanets, with orbits inside the habitable zones of late-type stars, one has to understand the evolutionary influence of the solar/stellar radiation and particle environment on planetary atmospheres. We show that the spectral type of a host star plays a major role in all atmospheric processes, where the photochemistry and evolution of planetary atmospheres, as well as their planetary water inventories must be studied within the context of the evolving stellar energy and particle fluxes. Only stable and dense enough atmospheres allow water to be liquid over geological time periods and protect the planetary surface from hostile radiation which may eventually result in the evolution of biospheres. It is shown that atmospheric expansion as well as thermal and nonthermal escape processes depend on the evolution of the stellar x-ray and EUV flux, and on the solar/stellar plasma flows. Model calculations indicate that earth-like exoplanets within close-in habitable zones of dwarf stars (< 0.05 AU) which are exposed to coronal mass ejections, may lose atmospheres equivalent to surface pressure values from tens to hundreds of bars, depending on the strength of their magnetic dynamo and absorbed XUV radiation.

6.1
Introduction

According to the present knowledge of the planetary population in our galaxy, Jupiter-like planets do exist outside the solar system. A concentrated research effort aimed at the detection of Neptune-class and larger and smaller terrestrial exoplanets with the CoRoT (CNES), Kepler (NASA), GAIA (ESA)and SIM (NASA) missions, is underway. Better understanding concerning the evolu-

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

tion of habitable planets is needed in the preparation phase for the planned Darwin (ESA) and TPF-C/I (NASA) missions, which will be designed for the detection of biomarkers in atmospheric spectra of known earth-like exoplanets. Terrestrial planets with biospheres are expected to orbit their mature stars along stable orbits inside their habitable zones. To allow for the development of biospheres, it must be determined whether a planet can keep its atmosphere and water inventory inside the habitable zone long enough for life to evolve. This includes interdisciplinary research activities related to stellar planetary relations, CO_2 surface weathering, effects of tidal-locking on the generation and maintenance of magnetic moments, plate tectonics, as well as climate models to understand under which conditions a habitable zone exists and how it evolves.

ESA and NASA recently began with the preparations of a road map for a precursor science program for the Darwin/TPF-C/I missions to investigate terrestrial-type exoplanets in terms of their origin, physical characteristics, evolution of their atmospheres and water inventories, and long time habitability. Although these missions will not be launched before the next decade, the basic decisions for their designs have to be made during the near future. One of the main question these missions must address is which star-types (M, K, G, F) may be good or at least preferred candidates in the search for habitable planets? Due to technical limitations, the target star selection of Darwin/TPF-C/I, is restricted to nearby stars in the distance range between 10–25 pc. A coronagraph design (TPF-C) would only allow observation of closer stars in this range, whereas the interferometer (Darwin/TPF-I) should be able to study stars up to the outer edge of this range. Since the large majority of stars, in the vicinity of our sun, fall in the lower mass domain, a detailed interdisciplinary study of the evolution of earth-like exoplanets within habitable zones of lower mass M and K-type stars is crucial for the characterization of the atmospheres of possible nearby exoplanets.

The present concept of the habitable zone for an earth-like planet is based on the orbital location where atmospheric CO_2 is able to sustain a sufficient amount of liquid H_2O on the planetary surface. If the distance to the host star is below a critical value, the planet experiences a runaway greenhouse effect like that on Venus. At the outer border of the habitable zone the surface temperature is too cold for liquid water to be stably present on the planet's surface. With this definition, the habitable zone depends mainly on the stellar luminosity and the planetary surface temperature, which is assumed to be stabilized slightly above zero degrees by the carbon–silicate cycle that controls partial CO_2 pressure. In such a scenario, vaporized H_2O starts to increase the surface warming and thus enhances the evaporation in a positive feedback. Although this general definition of the habitable zone developed for the sun-like G star planetary systems can be applied to all stellar spectral types, atmospheric evolution of terrestrial planets orbiting within the habitable zones of M, K and F-type stars will be different, so our present view of the habitable zone is incomplete. Low-mass M and K stars have their habitable zones at smaller distances, they have longer active x-ray and extreme ultraviolet (XUV) periods and different stellar winds than do solar-like stars.

Whether an H_2O-bearing terrestrial exoplanet in a dynamically stable orbit within its habitable zone can evolve into a habitable world like the earth, critically depends on its capability to survive the early period of heavy bombardment and high radiation of the young and active host star. The relevant wavelengths for the heating of upper atmospheres are those for which ionization takes place with $\lambda \leq 107$ nm, which contain only a small fraction of the host star's spectral power, but can lead to an efficient atmospheric escape triggered by hydrodynamic conditions. When a large amount of energy is deposited at the top of an atmosphere, preferably light species like H, H_2, and He can expand into the interplanetary space and may escape the planetary gravity field.

Astrophysical observations of solar proxies with various ages (representative for G stars) indicate an active phase of the young sun including continuous flare events and XUV radiation up to 100 times more intense than that of the present Sun, about 100 Myr after the star arrived at the Zero-Age Main-Sequence (ZAMS) [1]. After this very active stage, the solar XUV flux quickly decreases with time following a power-law relationship. Recent studies of K-type stars show that they remain at active emission levels somewhat longer than G stars and afterwards the activity also decreases as a power law in time. Early M-type stars appear to stay at high activity levels as long as about 1 Gyr, and then their luminosity decreases in an analogous way to G and K-type stars. Early K-type stars and early M stars may have XUV emissions of about 3–4 times and about 10–100 times higher, relative to their total luminosity, respectively, than solar-type G stars of the same age.

Apart from different stellar radiation and particle environments, terrestrial type planets inside the habitable zone of these stars can be partially or totally tidally-locked. The effect of tidal locking of terrestrial planets on climate variations, plate-tectonics, generation of internal magnetic moments, enhanced atmospheric erosion caused by stellar Coronal Mass Ejections (CME) and related magnetic clouds (MCs) may not be neglected. Recent model calculations show that tidal locking may result in weaker magnetospheres [2], which can be compressed by strong stellar winds during active periods to distances where the atmosphere builds an obstacle to the plasma flow similar to that on Venus and Mars [3]. Under such conditions, various nonthermal loss processes like ion pick up, sputtering, and ion clouds, triggered by plasma instabilities, can erode the atmosphere and significantly influence the planet's water inventory, even if the planet has an intrinsic magnetic moment. The atmospheres and oceans of tidally locked planets may freeze out to form a permanent icecap on the dark side of the planet. These uncertainties raise questions regarding atmospheric stability against thermal (hydrodynamic) and nonthermal escape and the evolution of planetary water inventories.

In this article we discuss the possible impact of these factors which are usually neglected in studies related to planetary habitability. In Section 6.2 we review the traditional concept of the circumstellar habitable zone of various star types, which is currently based on the constraint that a planet of the size and mass of the earth must contain large amounts of water and CO_2 reservoirs at a certain orbital distance from its host star; and we discuss various aspects which influence the evolution of planetary atmospheres not considered in the traditional habitable-zone concept.

In Section 6.3 we focus on the impact of stellar XUV radiation on the upper atmospheres of earth-like exoplanets and their initial water inventories within the habitable zones, while in Section 6.4 we study the stellar wind and CME plasma interaction with an upper atmosphere and related influences on the atmospheres of nonmagnetic and magnetized earth-like exoplanets within close-in habitable zones of low-mass stars. Finally, in Section 6.5 the implications of our preliminary results for planetary habitability and the habitable-zone concept in general, are discussed.

6.2
The Habitable Zone Around Main-sequence Stars

6.2.1
General Definition of the Habitable Zone

The habitable zone is defined as the space around a star, where the star's luminosity is sufficiently intense to maintain liquid water on the surface of a planet, without initiating a runaway greenhouse effect that dissociates water and sustains loss of hydrogen to space [4]. If climate conditions trigger a greenhouse effect, the surface temperature will be warmer compared to the effective temperature T_{eff} of the planet which can be written as

$$T_{\text{eff}} = \frac{(1-A)^{1/4}\,T_{\text{star}}}{\sqrt{2}}\left(\frac{R_{\text{star}}}{d}\right)^{1/2} \tag{6.1}$$

where R_{star} is the radius of the star, T_{eff} depends on the stellar radiative temperature T_{star} and thus on the luminosity of the star, the orbital distance d of the planet to the star and the planet albedo A. Although conditions necessary for the emergence, evolution and constancy of life are still unknown, long-time stability of a planetary atmosphere and liquid water are two main requirements which are widely accepted as an unavoidable necessity.

In the context of terrestrial planet-finding missions, it is important to distinguish between surface habitability, as opposed to possible habitats inside a planet. If life is or was present in the interior of Mars or of icy satellites like the Jovian moon Europa, biomarkers (gases related to biological activity) in their atmospheres would not be detected by remote sensing. Therefore, one should keep in mind that only an extended and productive biosphere, which is able to process the atmospheric and superficial materials can be detected.

An additional important point for the habitable zone is the cycling of CO_2 which is related to a planet climate regulation. On a terrestrial planet, the partial pressure of CO_2 is controlled by the carbonate–silicate cycle and kept at a value that maintains the average surface temperature slightly above the freezing point ($T_s = 273$ K). This self-regulation by the CO_2 cycle [5] relies on the fact that the formation of carbonates from atmospheric CO_2 requires liquid water. If the CO_2 level is too low to maintain substantial amounts of liquid water, the carbonate formation is stopped while the volcanic release of CO_2 keeps feeding the atmosphere. On the other hand, very high CO_2 levels stimulate the carbonate formation by enhancing the humidity of the lower atmosphere and the water cycle. These two feedback mechanisms work as long as CO_2 is released.

Planets orbiting at the inner edge of the habitable zone may experience two major problems. First, there is an upper limit of the infrared flux that can escape from a moist atmosphere. If energy above this limit is deposited into an atmosphere, a runaway greenhouse effect starts. In such a case the liquid water of a possible ocean can evaporate and the surface temperature rises to values beyond 1400 K [4].

The precise orbital distance where a runaway greenhouse effect occurs is uncertain, due to the effect of clouds on the albedo. For an atmosphere with a clear sky the process starts at about 0.85 AU for the present solar luminosity, but with highly reflecting clouds, this limit could be as small as about 0.5 AU. However, even when the runaway greenhouse conditions are not reached, the loss of hydrogen to space is strongly enhanced by the temperature profile of the atmosphere that becomes nearly isothermal, when the runaway threshold is approached. The resulting loss of water is substantial and affects the water content of the planet at orbital distances ≤ 0.95 AU. The long-term habitability of a planet located between this distance and the runaway threshold distance depends on the water reservoir and the time it takes to lose it completely.

At any orbital distance from the host star, one can estimate the amount of atmospheric CO_2 required to provide a mean surface temperature of about 273 K. This amount increases with the orbital distance of a terrestrial planet, until the outer limit of the habitable zone is reached. The existence of an outer edge is due to the increase of the infrared opacity and the albedo with increasing values of the CO_2 surface pressure, resulting respectively in the heating and cooling of the atmosphere. The determination of the outer boundary of the habitable zone, where the cooling effects dominate, is extremely difficult due to the complex role of CO_2-ice clouds [6]. However, with the present solar luminosity an outer limit of the habitable zone at about 2.4 AU can be estimated [7].

A habitable world, where the climate is controlled by abiotic CO_2 cycling can be very different from an inhabited world where the biosphere is involved in the fixation, cycling, and emission of atmospheric compounds. If the main greenhouse gas were CH_4 during a part of earth's history, before the rise of O_2 abundance and after the emergence of methanogenesis [8], the regulation of the climate would then be influenced by life itself and some complex coupling between them might arise. Therefore, a widespread and active biosphere may be able to make a planet suitable for life outside of the previously defined boundaries of the habitable zone. A CH_4-rich atmosphere sustained by biological activity and made up of CO_2, nitrogen oxides and, for instance, SO_2 compounds, may be opaque to the greater part of the thermal IR radiation and be habitable beyond the currently defined outer edge of the habitable zone. Hence, the general concept of the habitable zone may not be based solely on the radiative properties of CO_2 alone. Because CO_2 is the only efficient greenhouse gas that is compatible with an oxidizing atmosphere, the previously defined habitable zone may be restricted to biologically enriched O_2 atmospheres [9].

6.2.2
The Circumstellar Habitable Zone

For terrestrial planet-finding missions, the search for earth-like exoplanets will not be limited to solar-like G-type stars, but can be extended also to other lower-mass stars like M, K and slightly more massive stars like the F-type. By using the current definition of the habitable zone based on the sun, it is possible to estimate that it ranges from about 0.85 to 2.4 AU. As a reference, one can use evolution models for the luminosity of other main-sequence stars to predict the boundaries of the habitable zone for any star at any age as shown in Fig. 6.1.

One can also assume that the spectral shape of the stellar emission and the mass of earth-like planets do not change the limits of the habitable zone significantly. However, one should note that low-mass terrestrial planets cool more rapidly and may therefore not maintain a CO_2-rich atmosphere in order to prevent the planet from freezing. The boundaries of the habitable zone may change throughout a star's lifetime as the stellar luminosity evolves. This evolution is significant for very-low-mass stars during the first hundreds Myr, and for G and F stars, since their luminosity increases during the whole main sequence.

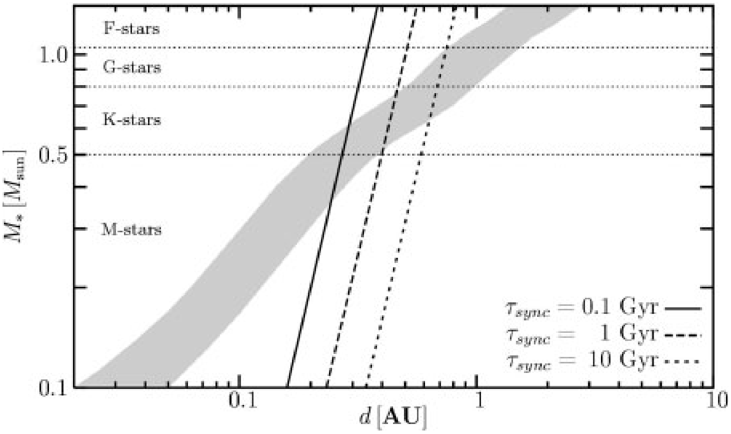


Fig. 6.1 Habitable zone for main-sequence stars (M, K, G, F) of different mass as a function of distance. Earth-like exoplanets become synchronized after 0.1 Gyr if they orbit within the habitable zone of M stars. (Courtesy of F. Selsis and J.-M. Grießmeier).

According to Fig. 6.1, earth-like planets in close orbits around their host stars are subjected to strong tidal interaction with the central body. This interaction can lead to many different effects, which directly influence the habitability of the planet. For planets in close-in habitable zones, strong tidal dissipation in the planet leads to gravitational locking on a very short time scale. The time scale τ_{sync} for synchronous rotation to establish

$$\tau_{\mathrm{sync}} \propto Q d^6 \tag{6.2}$$

depends on the orbital distance d and the planet's tidal dissipation factor Q [2, 10]. Because the rotation period equals the orbital period for gravitationally locked planets, fast rotation is not possible. As shown in Fig. 6.1, for stars with stellar masses below 0.6 M_{sun}, an Earth-mass planet orbiting in any part of the habitable zone becomes tidally-locked within the first Gyr after its origin. Therefore, additional questions regarding geophysical factors which can influence planetary habitability have to be considered.

6.2.3
Geophysical Factors which Influence Habitability

The consequences of tidal locking on planetary climate can be dramatic. Only a dense enough atmosphere (1.5–2 bars of CO_2) may provide homogeneous warming that prevents condensation and collapse of the atmosphere on the planet's night side [11]. Depending on the land surface and available amount of water, follow-up studies with a global circulation climate model indicate that the atmosphere of a terrestrial planet orbiting an M star may also be stable at pressure values < 1.5 bar [12].

Cycling of volatiles by plate tectonics contributes to regulating the composition of the earth's atmosphere, including the greenhouse gas CO_2 and, hence, the surface temperature and planet's habitability [13–17]. Plate tectonics is not only important for life on Earth due to the creation of land surfaces and enhancement of biodiversity through evolution on isolated continents, it is also an essential process with respect to the generation of a strong intrinsic planetary magnetic field, which protects the atmosphere from solar wind erosion and deflects high energy cosmic rays. The conditions which control plate tectonics are not known with certainty but the minimum requirements seem to be a sufficient planetary mass (and, hence, upward heat flow) to drive mantle convection and water to lubricate plate motion [18–20].

The effect of the lack of water can be seen on Venus: it has an earth-like mass but shows no plate tectonics [21]. Water affects the evolution of both a planetary mantle and a planetary tectonic engine. First, it makes the lithosphere sufficiently deformable for the subduction of the crust to occur. Second, it reduces the activation energy for creep and the solidus temperature of mantle rock, thereby enhancing the cooling of the interior and the efficiency of volcanic activity. Reservoirs of water and volatiles like CO_2 in the mantle also help to sustain the atmosphere through volcanic activity. On the other hand, water and CO_2 are recycled together with the subducting crustal rocks [22, 23].

Crust recycling through plate tectonics keeps the crust thin. If the crust is too thick, the lithospheric plate comprising the crust will be too buoyant to be subducted. A thin crust seems to be mandatory for plate tectonics to operate. Finally, plate tectonics helps to cool the interior efficiently which is required to maintain a strong thermally driven magnetic dynamo action for several Gyr.

By promoting very complex and global geological cycles, plate tectonics tends to generate conditions favorable for life. In the case of a one-plate planet (no subduction) the geologic interactions between the different subsets are reduced and more localized. Such a planet may not be able to sustain its internal magnetic field and to recycle volatiles into the mantle; finally it evolves toward a planet with an atmosphere which is either too thin (Mars) or too dense (Venus).

Moreover, tidal locking could also have a significant influence on large-scale convection in the planetary mantle and may inhibit plate tectonics partially or completely. Hence, there may be relatively few terrestrial planets within the tidal-lock radius showing plate tectonics. Tidally locked planets, which could not develop plate tectonics, may produce various life-frustrating scenarios like periodical outbreaks of Venus-type super volcanoes or different atmosphere–surface interaction processes, which may have a negative impact on the CO_2 cycle of a terrestrial planet.

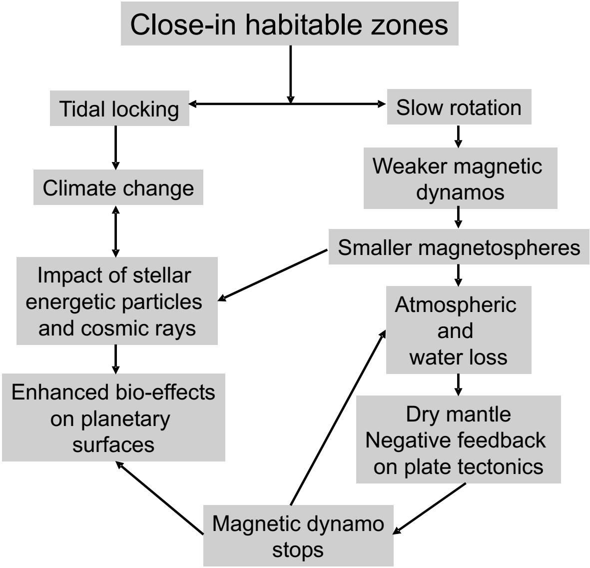


Fig. 6.2 Schematic illustration of geophysical factors which are important for the evolution of an earth-like habitable exoplanet.

In addition to the unfavorable geophysical conditions for the generation of a strong magnetic dynamo, the slow rotation of tidally locked exoplanets should further reduce or weaken the magnetic field (see Fig. 6.2). Commonly employed scaling laws for planetary magnetic moments yield rapidly decreasing moments with decreasing rotation rates. For these reasons weak magnetic moments are expected for close-in exoplanets [2, 24].

6.2.4
Stellar Radiation and Particle Forcing

One of the factors affecting the potential habitability of earth-like exoplanets around dwarf stars is their activity associated with the chromospheres and coronae. The relevant physical phenomena include intermittent and energetic flares, CMEs, stellar cosmic rays, enhanced coronal x-rays, and increased chromospheric UV emission. As discussed above, such events could severely inhibit the habitability of earth-like exoplanets within habitable zones close to the host star.

The activity of dwarf stars, like M-type stars, is thought to be the result of strong magnetic fields generated by their fully or partially convective interiors. Many of the relevant phenomena cannot be directly observed, except for flares. Even flaring rates and intensities require long-duration monitoring, so convenient proxies for this activity are used, such as optical Ca, H and K emission cores, H and Mg II emission, soft x-ray continuous emission, and a large number of UV- to soft x-ray emission lines [25]. A comprehensive discussion of flare stars can be found in the monograph by Gershberg [26].

Besides the long-lasting high-XUV emissions, another crucial parameter is the stellar plasma flow. Important factors in studies regarding the "stellar plasma – planetary atmosphere interaction" are the stellar wind density and velocity, which are highly variable with stellar age and depend also on the stellar spectral type and orbit location of the planet. The stellar mass loss and its resulting stellar wind was recently estimated by using the Hubble Space Telescope (HST) high-resolution spectroscopic observations [27] and by studying the hydrogen Lyman-α absorption features of several nearby main-sequence G and K-type stars. The observations revealed neutral hydrogen absorption features associated with the interaction between the stars fully ionized coronal winds and the partially ionized local interstellar medium. From these absorption features the mass flux of the stellar wind as a function of stellar activity was modelled. A small sample of observed solar-like stars suggests that the mass loss, and therefore the stellar wind mass flux, increases with stellar activity and stellar age.

From these observations one can conclude that younger solar-like stars have much denser and faster stellar wind compared with the present sun. A correlation between the mass-loss rates and the x-ray surface flux indicates an average solar wind density of up to 1000 times higher than today during the first hundred Myr after a solar-like star reaches the ZAMS [27]. However, it is important to note that these observations contained only a few K and G stars and more observations are needed in the future to get a detailed knowledge of the stellar wind mass flux and mass loss of young stars [27].

Furthermore, it is known from observations of our sun that every now and then strong eruptions (e.g., CMEs) occur and propagate as dense plasma structures through interplanetary space (see Fig. 6.3). These events should strongly affect the atmospheres and magnetospheres of terrestrial exoplanets in close-in habitable zones at orbit locations < 0.1 AU around low-mass M stars [3, 28]. Since M stars are very active in x-rays, they are expected to possess high flare rates and may thus generate permanent CME events [28, 29]. Because M stars are much more numerous compared to more massive K and G-type stars, it is of crucial importance that future studies investigate in detail the influence of CMEs on the formation and evolution of the atmospheres of terrestrial planets.

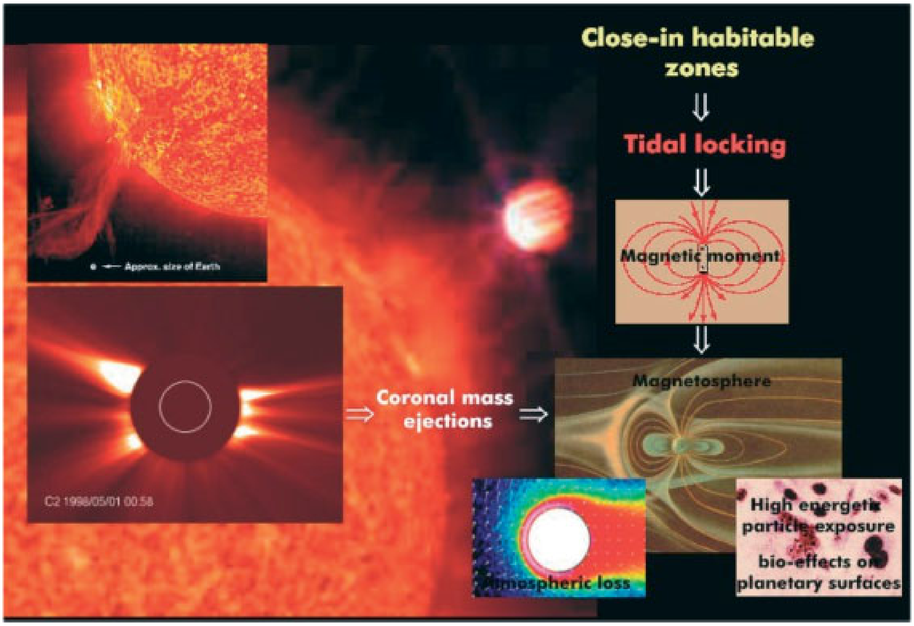


Fig. 6.3 Illustration of the expected impact of stellar winds and CME plasma on terrestrial exoplanets in orbits within close-in habitable zones of low-mass M and K-type stars. (Courtesy of N.K. Belisheva and N. Terada, SOHO/LASCO, ESA/NASA).

6.3
The Impact of the Stellar Radiation on the Upper Atmosphere

An XUV active star will be responsible for heating and expansion of the upper atmosphere. Under extreme conditions the temperature of the upper atmosphere can be so high that all light atoms overcome their gravitational binding to the planet and escape it in the form of planetary wind. However, a high CO_2 content in the atmosphere may prevent conditions for hydrodynamic outflow to develop due to intense IR-cooling. It is generally accepted that hydrodynamic escape of hydrogen atoms could be responsible for the heavy isotope enrichment which is observed in the atmospheres of Venus, Earth, and Mars with respect to their observed solar abundance [30–33]. From the isotope studies in the atmospheres of our solar system planets, we know that two stages of hydrodynamic escape are most likely required to fractionate; first, heavy Xe atoms (primordial); and later the lighter noble gases like Kr and Ar which are outgassed after the first stage of escape. The second stage of hydrodynamic escape should be of moderate magnitude, however, to avoid the fractionation of Kr, whose abundance is nearly solar-like. The high temperatures of the upper atmosphere and high atmospheric escape rates resulting from the intense XUV radiation which reaches saturation levels during the first several 100 Myr or even one Gyr of young sun-like stars or active dwarf stars, have a strong impact on atmospheric stability and the evolution of the planet's water inventory.

6.3.1
Heating, Cooling, and Expansion of the Upper Atmosphere

The most important heating and cooling processes in the upper atmosphere of the earth can be summarized as follows [34]:

- Heating due to N_2, O_2, and O photoionization by the solar XUV radiation (< 102.7 nm).
- Heating due to O_2 and O_3 photodissociation by the solar UV-radiation.
- Chemical heating in exothermic reactions.
- Neutral gas heat conduction.
- IR-cooling in the vibrational-rotational bands of CO_2, NO, O_3, OH, NO^+, $^{14}N^{15}N$, CO, etc.
- Heating and cooling due to the contraction and expansion of the thermosphere (to model the thermosphere diurnal variations).
- Turbulent energy dissipation and heat conduction.

The XUV heating in the earth's thermosphere yields an average dayside exospheric temperature of about 1000–1200 K [34, 35], while on Venus, which is closer to the sun but has a 96% CO_2 atmosphere, the exospheric temperature at the day side during average solar activity periods is only about 270 K [36–38]. The main explanation for the "cool" Venusian thermospheric–exospheric environment compared with the much "hotter" upper atmosphere of earth is very efficient cooling by IR emission in the 15 μm CO_2 band [39, 40].

By applying a diffusive-gravitational equilibrium and thermal balance model, one can investigate the heating of the thermosphere by photodissociation and ionization processes, by exothermic chemical reactions and cooling by CO_2 IR emission in the 15 μm band [38]. Figure 6.4 shows our modelled exospheric temperatures for a 96% CO_2 (4% N_2)atmosphere and for an atmosphere with an amount of only 10% CO_2 (90% N_2) mixing ratios as a function of the solar/stellar XUV flux for a Venus-mass and size planet. Depending on the spectral type of a star, as well as the calculated exospheric temperatures and exobase altitudes, atomic hydrogen could have been under diffusion-limited hydrodynamic "blow off" conditions even for a 96% CO_2

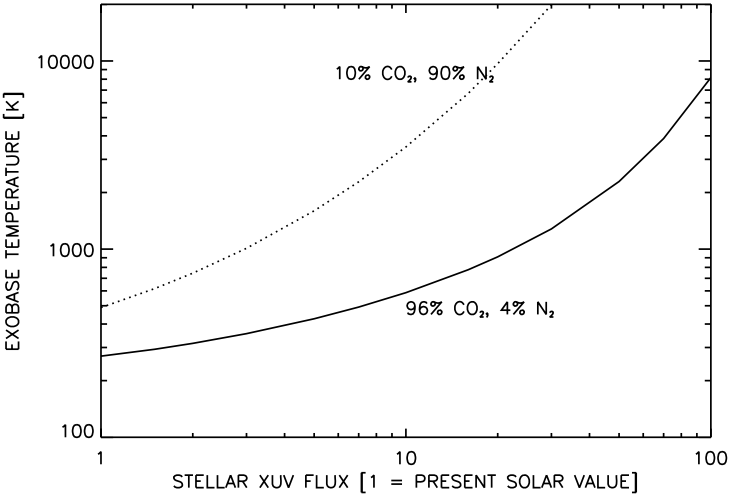


Fig. 6.4 Modelled exospheric temperatures for a 96% CO_2 Venus-like atmosphere (solid line) and for a 10% (dotted line) CO_2 and 90% N_2 atmosphere as a function of the solar/stellar XUV flux. Atomic hydrogen overcomes the gravitational potential of the planet if the temperature reaches a range of about 4000–4800 K.

Venus-like atmosphere for hundreds of million years after the star arrived at the ZAMS.

By assuming a 10% CO_2 and 90% N_2 atmosphere, the thermospheric temperatures on Venus could reach the critical value in the upper atmosphere for hydrogen atoms of about 4000 K about 3.8 Gyr ago and temperatures in excess of 20 000 K, which may result in non-hydrostatic equilibrium of the upper thermosphere, about 4 Gyr ago.

6.3.2
Thermal Escape of Atmosphere and Water

In the upper thermosphere the mean free path of the gas particles becomes large and collisions become negligible, so the light atmospheric constituents with a velocity which exceeds the gravitational escape velocity can escape from the planet. This region is called the exosphere, and the exobase begins at an altitude where the mean free path is about equal to the local scale height $H = kT_{\rm exo}/(mg)$ of the main atmospheric species, with k the Boltzmann constant, $T_{\rm exo}$ and g the temperature and gravitational acceleration at the exobase, and m the mass of the main atmospheric species.

When the exospheric temperature $T_{\rm exo}$ is large and the thermal escape parameter $X = GM_{\rm pl}m/kT_{\rm exo}r$ is smaller than 1.5, the exosphere becomes unstable and hydrostatic equilibrium no longer applies [41, 42]. For larger values of X the thermal escape flux can be determined by the Jeans equation

$$\phi_{\text{Jeans}} = \left(\frac{v_0}{2\sqrt{\pi}}\right) n_{\text{exo}}(1+X)e^{-X} \tag{6.3}$$

where $v_0 = (2kT_{\text{exo}}/m)^{1/2}$ is the most probable velocity of the particles at the exobase.

In an atmosphere where the exobase temperature may reach about 20 000 K during the active stellar period, as shown in Fig. 6.4, the Jeans escape flux for atomic oxygen according to (6.3) can be about 4.5×10^{10} cm^{-2} s^{-1}, resulting in loss rates of about 1.5×10^{29} s^{-1}. This loss rate is $\sim 10^4$ times larger than the present nonthermal oxygen loss rate expected at Venus [43]. These calculations emphasize the importance of a high CO_2 content for the stability of the terrestrial planetary atmospheres which are exposed to high XUV radiation fluxes.

One should also note that water-rich planets with hot lower atmospheres may result in humid upper atmospheres and very large loss rates of hydrogen. In a hydrogen-rich atmosphere the exobase moves to much larger distances compared with those calculated for a CO_2-rich atmosphere. In such a case atomic hydrogen formed from the H_2O vapor dissociation would dominate in the upper atmosphere and could escape, in a diffusion-limited regime, from the planet along with hydrodynamic wind which is similar to the observed case of the hydrogen-rich Jovian-type exoplanet HD209458b [44, 45]. In such conditions the whole amount of hydrogen contained in an earth-like ocean may escape from the planet during the active stellar period. However, according to Bounama [46] the earth currently contains an amount of water roughly corresponding to about four ocean masses, where about three ocean masses are in the mantle. Estimates for the total water contents of earth-like planets vary from about 1.5–20 ocean masses. Therefore, several ocean masses of water have to be lost to completely dry an earth-like planet.

6.4
Stellar Wind and CME Plasma Interaction with Planetary Atmospheres

If a planetary atmosphere is not protected by a strong earth-like magnetic field, the neutral gas in the upper atmosphere can be ionized due to charge exchange with the stellar plasma flow past the planet, by electron impact and by the XUV radiation, and picked up by the stellar wind. Therefore, the evolution of the stellar wind mass flux as a function of time has also important implications for the evolution of planetary atmospheres.

The current knowledge on CMEs comes from the study of the sun. A significant impact to this study has been made by the Large Angle and Spectrometric

Coronagraph (LASCO) on board of ESAs Solar and Heliospheric Observatory (SoHO), which observed more than 8000 CMEs since January 1996. The observational data on CMEs are related to two spatial domains: the near-sun region (up to 30 $R_{\rm Sun} \approx 0.14$ AU) remote-sensed by coronagraphs; and the outer region, including the geospace and beyond, where in-situ observations are made by spacecraft [47–49].

CMEs are associated with flares and prominence eruptions and their sources are usually located in active regions and prominence sites. A closed magnetic structure is the basic characteristic of the CME-producing regions. Recent studies of temporal correlation between CMEs and flares provide arguments in favor of the so-called common-cause scenario, according to which flares and CMEs are different manifestations of the same large-scale magnetic process on the star [50]. Despite the fact that the details of this process still remain unclear, it can already be definitely stated that intensive flaring activity of a star should be accompanied by an increased rate of CME production. The probability of CME-flare association increases with the duration of a flare [51]: $\sim 26\%$ for flare duration < 1 h; and 100% for flare duration > 6 h [28].

The latitude distribution of CMEs depends on the distribution of the closed-field magnetic regions (active regions) on the surface of the suns or stars. During the rising phase of the solar activity cycle (representative for active stars), CME latitudes spread gradually from those close to the equator (0°) up to all latitudes (±90°). At the same time observations of the sun indicate that the majority of plasma eruptions are located within the average latitude interval of $-60^\circ \leq \Theta \leq 60^\circ$ [49]. The CME occurrence rate shows good correlation with the sunspot number (SSN), with only small differences in detail and peaks with a delay of about two years after the peak in the SSN. This phenomenon is related to the fact that the sunspot activity is confined to the active region belt, whereas CME appear at all latitudes during the maximum [48, 49].

The maximum and minimum CME plasma densities $n_{\rm CME}$ at distances ≤ 0.1 AU are shown in Fig. 6.5 and can be estimated from the analysis of CME-associated brightness enhancements in the white-light coronagraph images. At larger distances (> 0.4 AU), $n_{\rm CME}$ is determined from in situ measurements by spacecraft of the density of the so-called magnetic clouds (MCs). The CME density dependence on the orbital distance d from the sun can be approximated by a power-law [28]

$$n_{\rm CME}(d) = n_0 \left(\frac{d}{d_0}\right)^{-3.6} \tag{6.4}$$

where $n_0 = 5 \times 10^5 \ldots 5 \times 10^6$ cm^{-3} and $d_0 = 3\,R_{\rm Sun}$ gives a good approximation for the values estimated from the SoHO/LASCO coronograph images. On the other hand, in situ observations of the plasma density in magnetic

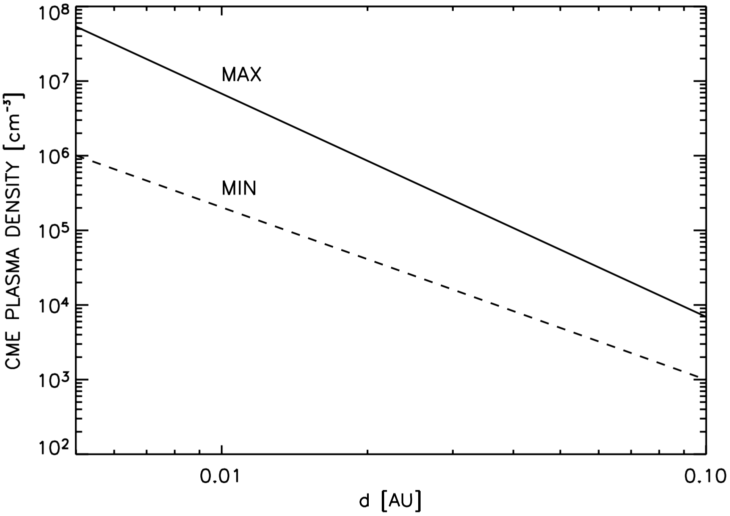


Fig. 6.5 Maximum and minimum CME plasma density between 0.005 and 0.1 AU.

clouds at distances > 0.4 AU give the power law

$$n_{\mathrm{MC}}(d) = n^0_{\mathrm{MC}} \left(\frac{d}{d_0}\right)^{(-2.4\pm0.3)} \tag{6.5}$$

with $n^0_{MC} = 6.47 \pm 0.85\,\mathrm{cm}^{-3}$ and $d_0 = 1$ AU. Based on CME and MC data one can provide density approximations according to Khodachenko [28]

$$n^{\min}_{\mathrm{ejecta}}(d) = 4.88d^{-2.31} \qquad n^{\max}_{\mathrm{ejecta}}(d) = 7.10d^{-2.99} \tag{6.6}$$

where d is taken in AU. The average mass of CMEs, M_{CME}, is $\approx 10^{15}$ g and the average duration of CMEs, τ_{CME}, at distances $(6\ldots10)R_{\mathrm{Sun}}$ is ≈ 8 hours.

6.4.1
Collision Probability Between Planets and CMEs

CMEs collide with a planet in a discrete way. The frequency of CME-planet collisions can be estimated from

$$f_{\mathrm{impact}} = \frac{\left(\Delta_{\mathrm{CME}} + \delta_{\mathrm{pl}}\right)\sin\left[(\Delta_{\mathrm{CME}} + \delta_{\mathrm{pl}})/2\right]}{2\pi\sin\Theta} f_{\mathrm{CME}} \tag{6.7}$$

where f_{CME} is the CME occurrence rate, Δ_{CME} and δ_{pl} are CME and planetary angular sizes, Θ is related to the CME ejection latitude interval $\pm\Theta$. When

$f_{\text{impact}} \cdot \tau_{\text{CME}} \geq 1$, i.e., when

$$f_{\text{CME}} \geq f^{c}_{\text{CME}} = \frac{2\pi \sin\Theta}{(\Delta_{\text{CME}} + \delta_{\text{pl}}) \sin\left[\frac{(\Delta_{\text{CME}}+\delta_{pl})}{2}\right] \tau_{\text{CME}}} \tag{6.8}$$

where f^{c}_{CME} is the critical CME production rate, a planet can be considered to be under continuous exposure of CME plasma [52]. In this case the speed and density of the stellar wind around the planet will be, in fact, the CME speed and density. Assuming $\Delta_{\text{CME}} = \pi/3$, τ_{CME} 8h, $\delta_{\text{pl}} \to 0$ and $\Theta = \pi/2$, Eq. (6.8) yields the critical frequency $f^{c}_{\text{CME}} \approx 36$ CMEs per day.

Figure 6.6 illustrates the f^{c}_{CME} dependence on the angular size of CMEs, Δ_{CME}. The two curves in Fig. 6.6, which correspond to $\delta_{\text{pl}} = 0$ and $\delta_{\text{pl}} = 0.1$, respectively, show the small influence of the planetary angular size, δ_{pl}, on the value of f^{c}_{CME}. The values obtained for the f^{c}_{CME} are not much higher than the CME production rate of the present day active sun $f_{\text{CME}} \sim 6 \ldots 8$ CMEs per day.

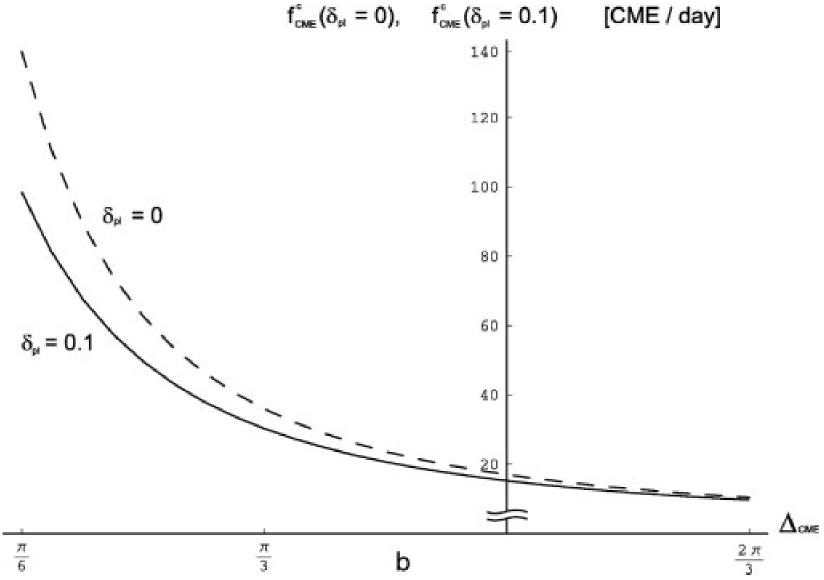


Fig. 6.6 Dependence of the critical value f^{c}_{CME} of stellar CME production rate, for which a planet appears under a continuous action of CME plasma flow, on CME angular size Δ_{CME}, for different planetary angular sizes: $\delta_{\text{pl}} = 0$ and $\delta_{\text{pl}} = 0.1$.

It is known from astrophysical observations that low-mass M-stars undergo frequent flare events [53]. The appearance of flares with a certain energy obeys a power law [28] $dN/dE \approx 0.06 L_x^{0.95} E^{-1.8}$, where dN is the number of flares per unit time in the energy interval $[E, E + dE]$, and L_x is the x-ray luminosity of a star characterizing its flaring activity. For an M star with a saturated x-ray activity level ($L_x \approx 7 \times 10^{28}$ erg/s) a few strong ($E > 10^{32}$erg) flares per day are expected, which is three orders of magnitude higher than that of the present sun.

6.4.2

Atmospheric Mass-loss Caused by CME Exposure

As discussed above, earth-like exoplanets within close-in habitable zones of low-mass M stars at orbital distances < 0.2 AU become tidally locked at time periods < 100 Myr, due to gravitational interaction with the star. The slow rotation rates of tidally locked planets lead to smaller planetary magnetic moments $M \sim (0.022 \ldots 0.15)M_E$, where M_E is the earth's magnetic moment [2, 24, 28].

This lower magentic moment decreases the magnetopause standoff distance which can be compressed down to very small distances above the planetary surface ($\geq 1/2$ earth radii), so that the atmosphere will directly interact with the stellar wind and CME plasma flow [3]. Figure 6.7 shows our modelled atomic oxygen number density, based on the temperature profiles shown in

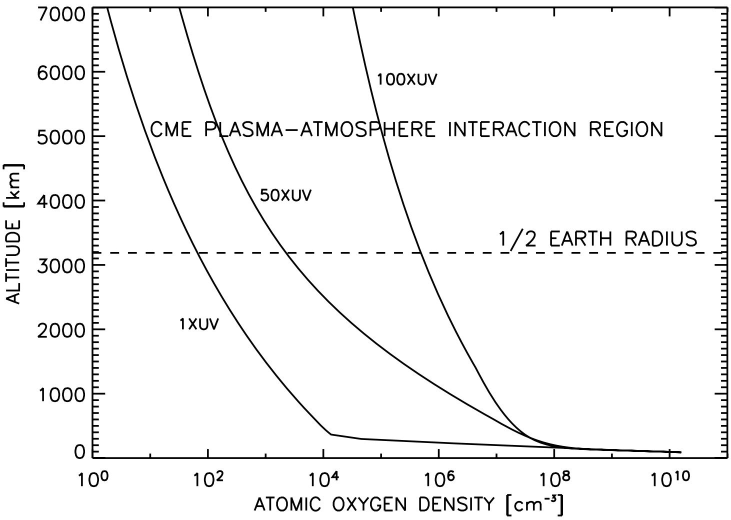


Fig. 6.7 Modeled atomic oxygen density profiles as a function of altitude and stellar XUV flux. The dashed line indicates the planetary obstacle (magnetopause) of 1/2 earth radii above the planetary surface. The oxygen atoms above this obstacle can be eroded by the CME plasma flux.

Fig. 6.4 as a function of altitude and XUV flux. One can see that the neutral atmosphere expands due to the heating by the XUV radiation. Therefore, planetary atmospheres which are exposed to high XUV fluxes are strongly affected by the CME plasma flux and related to its atmospheric erosion.

The application of a numerical test particle model to the modeled oxygen density profiles shown in Fig. 6.8 indicates that a tidally-locked weakly magnetized earth-like exoplanet at an orbital distance of about 0.05 AU inside the habitable zone of an XUV and CME active ($f_{\mathrm{CME}} > 14$) M star may lose 10 to 100 bars of atmosphere after 1 Gyr of continuous CME plasma exposure (see Fig. 6.8) if the M star XUV activity is more than 50 times higher compared with that of the present sun [3].

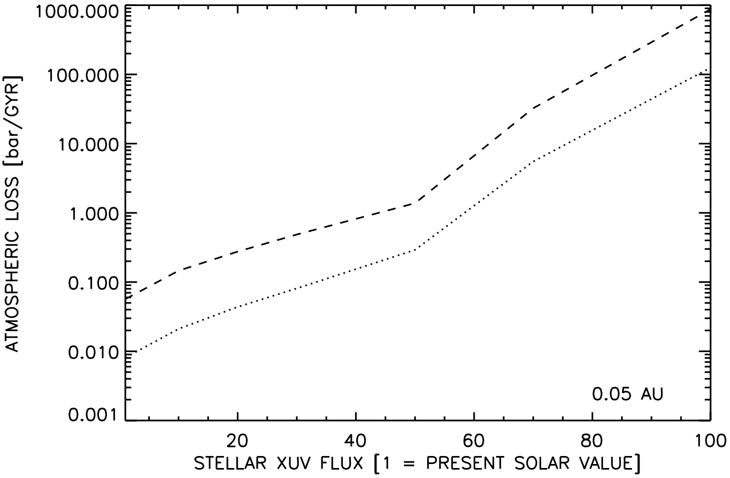


Fig. 6.8 Time dependent erosion from an earth mass and size exoplanet with a Venus-like CO_2 atmosphere as a function of minimum (lower dotted line) and maximum (upper dashed line) CME plasma flux at an orbital distance of 0.05 AU, for various XUV flux values normalized to that of the present sun at 1 AU and for a magnetopause obstacle at 0.5 earth radii above the planetary surface.

As discussed before, M stars could reach very high x-ray fluxes – up to 100 times or even more compared with G-type stars. Hence, heated and extended upper atmospheres of earth-like exoplanets at an orbital distance of about 0.05 AU can experience very high loss rates. These effects will have an important impact on the planetary habitability. It should also be noted that the loss rates shown in Fig. 6.8 correspond to an atmosphere with a high Venus-like CO_2 mixing ratio. Real earth-like N_2 atmospheres with lower CO_2 but higher N_2 content will be much more strongly affected by the stellar radiation and charged particle impact as discussed in this section.

Considering all these aspects, we conclude that it is very unlikely that habitable “earth-analogues” (planets with the same mass and size and atmospheric composition as the present earth) may be discovered within the habitable zones of active dwarf stars.

However, larger terrestrial planets, so-called “super-earths ” or “ocean planets” [54, 55] with sizes of about two Earth-radii will have masses of almost 10 times that of the earth. These exoplanets will have larger cores, which may result in stronger magnetic moments compared to earth-sized exoplanets and their higher mass will produce much less extended thermosphere–exosphere regions. Therefore, we expect that stronger magnetic moments of larger earth-like exoplanets with less extended upper atmospheres would reduce the damaging effect of the dense stellar plasma flow.

6.5
Discussion

We have shown that, although the current simplified definition of the circumstellar habitable zone [4] is useful, it is far from being sufficient and several additional factors which can influence the habitability of a planet have to be considered. While the prerequisites necessary for the emergence of life are still unknown, some conditions that may inhibit the evolution of a habitable planet can be, nevertheless, identified. Also, it seems clear that earth-like exoplanets within close-in habitable zones of low-mass stars, may evolve in a different way from similar planets in the habitable zones of solar-like G stars.

Besides open questions regarding the impact delivery of water and the role of giant planets in the formation and evolution of terrestrial exoplanets in dwarf-star systems, we would like to draw attention to the three main factors which are certainly different when compared with the earth-like exoplanets in orbits within G-star habitable zones. These factors according to geophysical, plasma physical, and aeronomical conditions can be summarized as follows:

- Planets inside the habitable zones of M and some K stars are tidally locked, therefore plate tectonics may not develop and big hot-spot volcanoes may frustrate life periodically or affect the long-time habitability of such a planet.

- Tidal locking will result in climate effects and weaker intrinsic magnetic moments so that dense stellar winds and CME plasma impacts can compress their magnetospheres and erode upper atmospheres.

- The habitable zones of low-mass stars are exposed to strong XUV irradiation during longer time periods and to denser stellar plasma fluxes as well. This radiation and particle exposure can erode the atmosphere of an earth-sized planet and may dramatically limit a range of the purely climatologically defined habitable zone.

Observation of tidally locked exoplanets without and with atmospheres around low-mass stars by a Darwin-like terrestrial planet-finding mission, would give us an idea of at which orbital distances a terrestrial planet inside the habitable zone of a dwarf star can retain an atmosphere.

Because close-in terrestrial exoplanets are unlikely to be protected by an extended earth-like magnetosphere, energetic particles which are related to CMEs and active regions of their host stars, together with cosmic rays can reach almost the whole surface area of the upper atmospheres [2]. Under these conditions energetic particles can strongly interact with the atmosphere where they change the chemistry, possibly even the O_3 content, and climate. If the pressure conditions are similar to those on earth, secondary energetic particles can penetrate to the planet's surface where they will have an impact on biological systems [2].

Only multidisciplinary efforts including astronomy, planetary and biological sciences will allow a detailed definition of earth-like planets beyond the solar system. In order to find constraining parameters for the evolution of life in the universe and the existence of biospheres, atmospheric, climate, photochemical, radiative transfer, magnetohydrodynamic, test particle and hybrid models for studying solar/stellar wind/atmosphere interaction processes, as well as magnetospheric, geological, and more types of models will have to be developed and applied within the framework of an elaborate precursor science program dedicated to the characterization of extraterrestrial planets, which are expected to be discovered by the future terrestrial planet-finding missions like Darwin or TPF-C/I during the next decade.

Acknowledgments

M. L. Khodachenko acknowledges support from the Austrian "Fonds zur Förderung der wissenschaftlichen Forschung" (project P16919-N08), ÖAD-RFBR Scientific and Technical Collaboration Program (project No. 18/2006), and the ÖAD-Acciones Integradas Program (project No. 11/2005). H. Lammer, H. I. M. Lichtenegger and Yu. N. Kulikov are also indebted to the Austrian Academy of Sciences, "Verwaltungsstelle für Auslandsbeziehungen" and the Russian Academy of Sciences for supporting working visits to the PGI/RAS in Murmansk, Russian Federation. The authors acknowledge also a support from the "Büro für Akademische Kooperation und Mobilität" of the Austrian Academic Exchange Service under the ÖAD-Acciones Integradas project No. 12/2005. Further, the authors thank the Austrian Ministry for Science, Education and Culture (bm:bwk) for supporting the CoRoT project. This study was also supported by the International Space Science Institute (ISSI, Bern) and was carried out under the framework of the ISSI Team "Evolution of Habitable Planets".

References

1 Ribas, I., Guinan, E. F., Güdel, M., and Audard, M.: 2005, *ApJ*, **622**, 680.

2 Grießmeier, J.-M., Stadelmann, A., Motschmann, U., Belisheva, N. K., Lammer, H., and Biernat H. K.: 2005, *Astrobiology*, **5**, 587.

3 Lammer, H., Lichtenegger, H. I. M., Kulikov, Yu. N., Grießmeier, J.-M., Terada, N., Erkaev, N. V., Biernat, H. K., Khodachenko, M. L., Ribas, I., Penz, T., and Selsis, F.: 2007, *Astrobiology*, **in press**.

4 Kasting J. F.: 1988, *Icarus*, **74**, 47.

5 Walker, J. C. G., Klein, C., Schidlowski, M., Schopf, J. W., Stevenson, D. J., and Walter M. R.: 1983, *Earth's Earliest Biosphere*, (ed. J. W. Schopf), Princeton University Press, 160.

6 Forget F. and Pierrehumbert R. T.: 1997, *Science*, **278**, 1273.

7 Mischna, M. A., Kasting, J. F., Pavlov, A., and Freedman, R.: 2000, *Icarus*, **145**, 546.

8 Pavlov A. A., Kasting J. F., Brown L. L., Rages K. A., and Freedman R.: 2000, *J. Geophys. Res.*, **105**, 11981.

9 Selsis F.: 2000, The evolving Sun and its influence on planetary environments, (eds. B. Montesinos, A. GimÃl'nez, E. Guinan), *ASP Conf. Series*, **269**, 273.

10 Goldreich, P. and Soter, S.: 1966, *Icarus*, **5**, 375.

11 Joshi, M. M., Haberle, R. M., and Reynolds, R. T.: 1997, *Icarus*, *129*, 450.

12 Joshi, M. M.: 2003, *Astrobiology*, **3**, 415.

13 Sundquist, E. T.: 1993, *Science*, **259**, 934.

14 Kasting, J. F.: 1993, *Science*, **259**, 920.

15 Franck, S., Block, A., von Bloh, W., Bounama, C., Schellnhuber, H.-J., and Svirezhev, Y.: 2000, *Planet. Space Sci.*, **48**, 1099.

16 Franck, S., von Bloh, W., Bounama, C., Steffen, M., Schönberner, D., and Schellnhuber, H.-J.: 2000, *J. Geophys. Res.*, **105**, 1651.

17 Wostencroft, R. D. and Raven, J. A.: 2002, *Icarus*, **157**, 535.

18 Regenauer-Lieb, K., Yuen, D. A., and Branlund, J.: 2001, *Science*, **294**, 578.

19 Solomatov, V. S.: 2003, *34th Annual Lunar and Planetary Science Conference Proceedings*, **no.1406**.

20 Solomatov, V. S.: 2004, *J. Geophys. Res.*, **109**, 10.1029/2003JB002628.

21 Jakosky, B. M. and Phillips, R. J.: 2002, *33rd Annual Lunar and Planetary Science Conference*, **no.1195**.

22 Breuer, D., Zhou, H., Yuen, D. A., and Spohn, T.: 1996, *J. Geophys. Res.*, **101**, 7531.

23 Breuer, D., Yuen, A., and Spohn, T.: 1997, Phase transitions in the Martian mantle: Implications for partially layered convection. *Earth Planet. Sci. Lett.*, **148**, 457.

24 Grießmeier, J.-M., Motschmann, U., Stadelmann, A., Penz, T., Lammer, H., Selsis, F., Ribas, I., Guinan, E. F., Biernat, H. K., and Weiss, W. W.: 2004 *A&A*, **425**, 753.

25 Ayres, T. R.: 1997, *J. Geophys. Res.*, **102**, 1641.

26 Gershberg, R. E.: 2005, Solar-Type Activity in Main-Sequence Stars, Springer, Berlin.

27 Wood, B. E., MÃijller, H.-R., Zank, G., and Linsky, J. L.: 2002, *ApJ*, **574**, 412.

28 Khodachenko, M. L., Ribas, I., Lammer, H., Grießmeier, J.-M., Leitner, M., Selsis, F., Eiroa, C., Hanslmeier, A., Biernat, H. K., Farrugia, C. J., and Rucker, H. O.: 2007, *Astrobiology*, in press.

29 Scalo J., Kaltenegger, L., Segura, A., Fridlund, M., Ribas, I., Kulikov, Yu. N., Grenfell, J. L., Rauer, H., Odert, P., Leitzinger, M., Selsis, F., Khodachenko, M. L., Eiora, C., Kasting, J., and Lammer, H.: 2007 *Astrobiology*, in press.

30 Pepin, R. O.: 1991 *Icarus*, **92**, 2.

31 Pepin, R. O.: 1994, *Icarus*, **111**, 289.

32 Pepin, O. R.: 1997, *Icarus*, **126**, 148.

33 Pepin, R. O.: 2000, *Meteoritics & Planetary Science*, **35**, 126.

34 Gordiets, B. F., Kulikov, Yu. N., Markov, M. N., and Marov, M. Ya.: 1982, *J. Geophys. Res.*, **87**, 4504.

35 Jacchia, L. G.: 1977, Thermospheric Temperature, Density and Composition: New Models, *Spec Rep. Smithson. Inst. Astrophys. Obs., Cambridge Mass., USA*, 375.

36 Nieman, H. B., Hartle, R. E., Hedin, A. E, Kasprzak, W. T., Spencer, N. W., Hunten, D. M., and Carignan, G. R.: 1979, *Science*, **205**, 54.

37 Hedin, A. E., Nieman, H. B., Kasprzak, W. T., and Seiff A., *J. Geophys. Res.*, **88**, 73.

38 Kulikov, Yu. N., H. Lammer, Lichtenegger, H. I. M., Terada N., Ribas, I., Kolb, C., Langmayr, D., Lundin, R., Guinan, E.

F., Barabash, S., and Biernat, H. K.: 2006, *Planet. Space Sci.*, **54**, 1425.

39 Gordiets, B. F. and Kulikov, Yu. N.: 1985, *Adv. Space Res.*, **5**, 113.

40 Bougher, S. W., Engel, S., Roble, R. G., and Foster, B.: 1999, *J. Geophys. Res.*, **104**, 16,591.

41 Chamberlain, J. W.: 1963, *Planet. Space. Sci.*, **11**, 901.

42 Öpik, E. J.:1963, *Geophys. J. Roy. Astron. Soc.*, **7**, 490.

43 Lammer, H., Lichtenegger, H. I. M., Biernat, H. K., Erkaev, N. V., Arshukova, I. L., Kolb, C., Gunell, H., Lukyanov, A., Holmstrom, M., Barabash, S., Zhang, T. L., and Baumjohann, W.: 2006, *Planet. Space Sci.*, **54**, 1445.

44 Lammer, H., Selsis, F., Ribas, I., Guinan, E. F., Bauer, S. J., and Weiss, W. W.: 2003, *ApJ L*, L1221.

45 Vidal-Madjar, A., Lecavalier des Etangs, A., Désert, J.-M., Ballester, G. E., Ferlet, R., Hébrand, G., and Mayor, M.: 2003, *Nature*, **422**, 143.

46 Bounama, C., Franck, S. and von Bloh, W.: 2001, *Hydro. Earth Sys. Sci.*, **5**, 569.

47 Gopalswamy, N., Kaiser, M. L., Thompson, B. J., Burlaga, L. F., Szabo, A., Vourlidas, A., Lara, A., Yashiro, S., and Bougeret, J. L.: 2000, *Geophys. Res. Lett.*, 27, 1427.

48 St. Cyr, O. C., Howard, R. A., Sheeley, N. R. Jr., Plunkett, S. P., Michels, D. J., Paswaters, S. E., Koomen, M. J., Simnett, G. M., Thompson, B. J., Gurman, J. B., Schwenn, R., Webb, D. F., Hildner, E., and Lamy, P. L.: 2000, *J. Geophys. Res.*, **105**, 18169.

49 Gopalswamy, N.: 2003, *Adv. Space Res.*, **31**, 869.

50 Zhang, J., Dere, K. P., Howard, R. A., Kundu, M. R., and White, S. M.: 2001, *ApJ*, **559**, 452.

51 Sheeley, N., Howard, R.A., Koomen, M. J., and Michels, D. J.: 1983, *ApJ*, **272**, 349.

52 Khodachenko, M. L., Lammer, H., Lichtenegger, H. I. M., Langmayr, D., Erkaev, N. V., Grießmeier, J.-M., Leitner, M., Penz, T., Biernat, H. K., Motschmann, U., and Rucker, H. O. 2007, *Planet. Space Sci.*, in press.

53 Haisch, B., Strong, K. T., and Rodono, M.: 1991, *ARA&A*, **29**, 275.

54 Selsis, F., Chazelas, B., Borde, P., Ollivier, M., Brachet, F., Decaudin, M., Bouchy, F., Ehrenreich, D., Griessmeier, J. -M.; Lammer, H.; Sotin, C., Grasset, O., Moutou, C., Barge, P., Deleuil, M., Mawet, D., Despois, D., Kasting, J. F., and Leger, A.: 2007, *arXiv:astro.ph/0701608*.

55 Léger, A., Selsis, F., Sotin, C., Guillot, T., Despois, D., Mawet, D., Ollivier, M., Labéque, F. A., Valette, C., Brachet, F., Chazelas, B., and Lammer, H.: 2004, *Icarus*, **(**169), 499.

7
Dynamics of the Extrasolar Planetary Systems

Tatiana A. Michtchenko, Sylvio Ferraz-Mello, and Christian Beaugé

Abstract

This chapter is a review of the long-period dynamics and stability of planetary systems. The relevance in the understanding of the planetary system behavior is described in Section 7.1. In Section 7.2, the 21 multi-planetary extrasolar systems are grouped with respect to their dynamical behavior into three classes: the secular, the hierarchical and the resonant classes, the latter including a subclass of nearly resonant systems with low-eccentricity planets. In Section 7.3, we introduce the main techniques commonly used for exploring the dynamical features of the systems, such as numerical integrations of exact equations of motion, dynamical map constructions and analytical modeling, and the advantages of each technique are discussed briefly. In Section 7.4 we describe the nonresonant behavior of the systems with two planets in co-planar orbits and extend the planar problem to the three-dimensional case. We conclude this section by describing the effects of the proximity of a secular system to mean-motion resonances. The phenomena related to mean-motion resonances are presented in Section 7.5, through a study of the stability of the GJ 876 and HD 82943 planetary pairs. Finally, a special subclass of nearly resonant systems, which includes the solar system, is described in Section 7.6.

7.1
Introduction

With the discovery of the first planets orbiting another stars, PSR 1257+12 Wolszczan and Frail [1] and 51 Peg [2], we can now continuously improve our basic concepts of planetary dynamics. The newly discovered worlds challenge our imagination by their unusual orbital configurations and raise new questions in our understanding of their dynamical evolution. After a centuries-old paradigm, with clock-working revolution of the low-eccentricity and low-inclination planets, our own solar system now emerges as an atypical structure in the current sample of the 21 known multi-planetary extrasolar systems.

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

In the light of new discoveries, the understanding of planetary dynamics is receiving particular attention for several reasons. These include:

1. Detection of the extrasolar planets and determination of their orbits. Almost all the extrasolar planets have been detected indirectly by radial-velocity measurements. This technique explores the dynamical effects that the planets produce in their parent star's motion. The determination of the planetary orbits from the observational data involves the basic concepts of celestial mechanics. Two approaches are often used: the fitting of the observational data by non-interacting Keplerian orbits and the N-body model accounting for perturbations to Keplerian orbits. In both cases, to obtain good fits, the understanding of planetary dynamics is indispensable.

2. Long-term evolution and stability of the extrasolar systems. The discovery of extrasolar planets has brought into focus a long standing, but yet unsolved, problem of the long-term stability of planetary systems. The extrasolar planets exhibit an astonishing variety of physical and orbital patterns, such as large masses, very small distances to the central star and high eccentricities. Also, planets have been detected in binary and triple-stars systems and around pulsars. Several multi-planet systems exhibit resonant dynamical states which are not seen among the planets of our solar system.

3. The theories of formation/migration and dynamical evolution of the planet systems. Most of the theories were designed to explain the planets orbiting our sun. Nowadays, the striking contrasts between extrasolar and the solar systems activate the substantial revision of previous models and the appearance of alternative theories. An important tool in planetary dynamics studies has been the classical perturbation theory developed to explain the planetary orbits in the solar system, however, restricted to the systems with small eccentricities and inclinations. The extension of these theories to the larger domains of orbital parameters is required to explain the orbital characteristics of the extrasolar systems.

4. Planetary habitability studies and the search for life. Which conditions are necessary to make a planet habitable and for life to appear and survive? The answer to this question includes an extended list of requirements and, no doubt, some of them will be related to the dynamical stability of planets inside the habitable zones.

7.2
Classification of the Systems with Respect to their Dynamical Behavior

Currently (December 2006), there are 21 multiple-planet systems catalogued in The Extrasolar Planets Encyclopaedia[1]; seven of these systems are composed of more than two planets. Keeping in mind that some of the planets were quite recently detected and their data can be substantially modified by further observation, we will use this current sample to construct a broad picture of planetary configurations.

From the dynamical point of view, the mutual distance between planets is the most important characteristic of a system. According to Kepler's third law, it is related to the ratio of the orbital periods of planets. It is worth noting that orbital periods are primary parameters of the observation fitting and therefore free from mass uncertainties. Eccentricities are also important, because, they possibly carry information on the dynamical evolution of the system. Using these two parameters, we select the pairs of planets in consecutive orbits and plot them in Fig. 7.1. In this figure, we also plot the nominal location of main mean-motion resonances, as given by Kepler's third law.

The position of a pair of planets with respect to mean-motion resonances defines its qualitative comportment. To distinguish kinds of planet behavior, we use three classes proposed by Ferraz-Mello et al. [3]. For the purpose of our study, we slightly modify that classification. Class I includes the close pairs ($1 > T_1/T_2 > 0.2$) involved in one of the main mean-motion resonances. The resonant dynamics described in Section 7.5 is specific for systems from this class. There is also a peculiar subclass of systems with low-eccentricity planets, which are not trapped inside a mean-motion resonance, but lie close to one of them. The behavior of these quasi-resonant systems is discussed in Section 7.6.

Class II includes planets which are far from mean-motion resonances, but still closely approach each other ($1 > T_1/T_2 > 0.14$), and therefore have strong gravitational interactions. We refer to these systems as secular ones and describe their dynamics in Section 7.4. Finally, Class III includes the pairs of planets which are largely separated ($T_1/T_2 < 0.14$) and, consequently, are weakly interacting; these systems are often referred to as hierarchical systems [4, 5]). It is worth a warning here about the possible difficulties in distinguishing between the populations of Class I and Class II. These difficulties arise mainly due to uncertainties in determination of the physical and orbital parameters of the planets; formal uncertainties in the eccentricities are on the order of 20% [6]. However, the studies of the dynamical stability of the systems can still impose constraints on their motion.

---

1) homepage `http : //vo.obspm.fr/exoplanetes/encyclo/index.php` (Schneider).

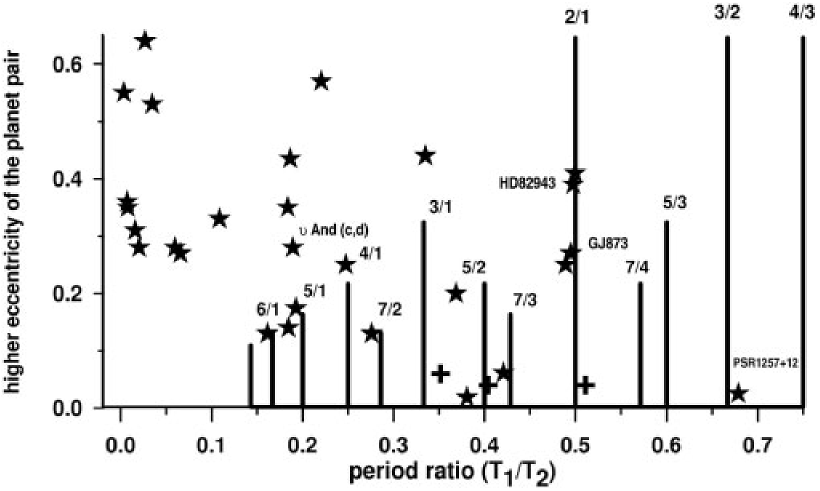


Fig. 7.1 Distribution of the pairs of planets in consecutive orbits on the plane (T_1/T_2, e), where T_1 and T_2 are orbital periods of the inner and outer planets, respectively, and e is the highest eccentricity of the two orbits. The nominal location of main mean-motion resonances are shown by peaks whose heights are inversely proportional to the order of the corresponding resonance. The extrasolar systems are shown by stars, while the pairs Jupiter–Saturn, Saturn–Uranus and Uranus–Neptune are shown by crosses. Some systems that we consider in this article are labeled.

The population of Class I is generally characterized by large masses and high eccentricities. Two very close planets ($T_1/T_2 > 0.45$) in very eccentric orbits are unable to remain stable if not tied by a mean-motion resonance. The distribution of the exosystems in Fig. 7.1 shows that the 2/1 resonance is the most populated; there are four systems inside this resonance, GJ 876(c,b), HD 82943, HD 73526 and HD 128311. Two of the planets of 55 Cnc seem to be involved in a 3/1 mean-motion resonance [7, 8], although additional observations are necessary to confirm this configuration. Recently, one more pair, the HD 202206 planets, was suggested to evolve inside the 5/1 resonance [9]. The subclass of nearly resonant planets, on the other hand, is populated by low-eccentricity planets: the giant planets of our solar system and the terrestrial planets of the pulsar PSR 1257+12. The nearly circular 47 Uma planets are also considered to belong to this subclass.

The members of Class II are located in the regions between the main mean-motion resonances in Fig. 7.1. The first discovered multi-planet system υ Andromedae belongs to this category [10]. Some of the systems have very-high eccentricities ($e > 0.4$) and can be considered as putative Class I candidates, since updated orbital parameters could place them in low-order mean-motion resonances (as an example see the case of the μ Ara system described in Goździewski et al. [11]).

Finally, the population of Class III is easily identified at small values of T_1/T_2 in Fig. 7.1. Due to the weak mutual perturbations, these systems reside in extended zones of stability in which they can easily survive for a very long time. It is worth noting that some of the planets have large masses corresponding to those of brown dwarfs and all of them have high eccentricities.

7.3
Numerical Integrations, Dynamical Maps and Modeling of the General Three-body Problem

Direct numerical integrations of the full equations of motion are widely used in the studies on the long-term behavior of the extrasolar systems. Being a powerful tool for exploring the dynamical features of the systems they do, however, have several limitations. The most serious of them concerns the correct interpretation of the results obtained. Indeed, in the multi-planet systems, the mutual gravitational interactions may produce an intricate set of complex dynamical effects, hard to explain without the basic concepts of analytical studies. Also, the inadequacy of the space chosen to present the results of integrations may be the source of spurious appearances, leading to erroneous interpretation (one example will be shown in Section 7.4.1).

Another limitation is connected to the problem of observational uncertainties in orbital parameters of the extrasolar systems. To investigate the long-term stability of these systems, we have to perform numerical integrations, which require initial conditions of the planets. As a result, numerical integrations give information on the stability, which is limited to the initial conditions used. The orbital elements of the extrasolar systems and their physical parameters are continuously changing due to new observational data. Even in the case of slight modification, the numerical integrations must be done again, notwithstanding the high cost of them. An alternative method of overcoming this problem is the construction of dynamical maps. Dynamical maps are also founded on purely numerical integrations, but over a large set of initial conditions, which cover a neighborhood of the system under study. The time intervals of the simulations can be sufficiently short, but still appropriate for the detection of the occurrence of mean-motion and secular resonances. Analyzing the topological structure of dynamical maps, we can detect the mechanisms that can induce instabilities in the planetary motion, and determine domains of their action. We will often use the dynamical maps here to supplement the results of analytical modeling.

Analytical investigations, providing basic concepts, are fundamental in stability studies. Only these approaches are able to explore large sections of the parameter space of the problem and provide an answer to the question of the global stability of the system under study. In addition, the modeling introduces the possibility of studying distinct dynamical states of the systems, separately. According to their type of motion, the systems are generally separated between secular and resonant ones. To have a better understanding of the planetary behavior, we apply different models for each of these dynamical states. It is common practice in dynamical studies to analyze what were previously general aspects of systems, in order to make a decision about the approach to be used.

In this text, we only briefly introduce the modeling of the general three-body problem (for more details see [12–19]). The basic model consists of two planets with masses m_1 and m_2, for the inner and outer planets respectively, revolving in the same sense around the central star of mass m_0. We develop a semi-analytical approach, employing a numerical averaging over the short-period perturbations (of the order of orbital periods) in the mutual interaction of the two planets, to obtain the corresponding Hamiltonian function. The analytical approach allows us to investigate a large domain of the phase space of the three-body problem without time-expensive numerical integrations of the equations of motion, and without any restriction on the magnitude of the planetary eccentricities and inclinations.

7.4
Secular Dynamics: υ Andromedae Planets c and d

In the absence of commensurabilities between the orbital periods of two planets, their Keplerian orbits undergo long-term secular variations due to the mutual gravitational interactions. The classical Laplace–Lagrange planetary theory [20, 21], which well describes the secular behavior of the planets with small eccentricities and inclinations, has shown that the secular perturbations mostly affect the shape and the orientation of the planetary orbits, while the semi-major axes remain invariable. In particular, the linear and first order in masses theory has demonstrated the long-term stability of the solar system and provided the main features of the planetary secular motion.

In contrast with our solar system, the known exoplanets have generally larger planet masses, smaller semi-major axes, and higher orbital eccentricities; these factors, particularly the high eccentricities, limit the application of classical theory for these systems. For this reason, we use a semi-analytical approach which is, in essence, an extension of the linear approximation to the domain of the moderate-to-high eccentricities and inclinations. It employs a numerical averaging of the short-period gravitational interactions of the planets (of the order of orbital periods), to determine the Hamiltonian describing the secular perturbations in the system. The secular total energy of the system is then defined as

$$\overline{\mathcal{H}}_{\text{sec}} = -\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} \frac{G\, m_1\, m_2}{a_2}\, R(a_i, e_i, I_i, M_i, \omega_i, \Omega_i)\, dM_1 dM_2 \qquad (7.1)$$

where the integrand is the disturbing function of the canonical astrocentric semi-major axes, eccentricities and inclinations of the planets (a_i, e_i and I_i, respectively) and the angular elements: mean anomalies M_i, arguments of pericenter ω_i and longitudes of node Ω_i ($i = 1, 2$, stands for the inner and outer planets, respectively). The relationship between canonical and usual osculating heliocentric orbital elements is described in detail in Ferraz-Mello et al. [22]. Note that the contribution of the Keplerian part is constant (since the semi-major axes are constant) and therefore need not be considered.

Together with the total energy $\overline{\mathcal{H}}_{\text{sec}}$, we introduce such global quantities as the total Angular Momentum (AM) and the Angular Momentum Deficit (AMD) [23, 24], which, up to second order in masses, can be written:

$$\begin{aligned} \text{AM} &= m_1\, n_1\, a_1^2 \sqrt{1 - e_1^2}\cos I_1 + m_2\, n_2\, a_2^2 \sqrt{1 - e_2^2}\cos I_2 \\ \text{AMD} &= \sum_{i=1}^{2} m_i\, n_i\, a_i^2 - \text{AM} \end{aligned} \qquad (7.2)$$

The use of AMD is easier because it has a clear interpretation. Indeed, the AMD has a minimum value (zero) for circular and co-planar orbits and increases with increasing eccentricities and inclinations. It should be emphasized that, in the secular dynamics, AMD and AM are equivalent and the behavior of AM is reverse of that of AMD.

The conservation of the total energy and of the total angular momentum along one secular path, defines the main features of the secular motion of planetary systems. Also, it should be noted that the secular phase space structure of the three-body (two-planet) system depends only on two parameters, the ratio of the planet masses and the ratio of their orbital semi-major axes.

The pair of the υ Andromedae planets **c** and **d** is a classical example of secular behavior. Actually, the best-fitting results of the radial velocity measurements show that this system is composed of three planets, but the third planet, **b**, is far inward and has a smaller mass; thus, in first approximation, we can neglect the perturbations due to this planet on the motion of the pair **c–d**. Assuming an edge-on orbital configuration of the planets and neglecting the mutual inclinations between their orbits, we numerically integrate the secular paths, which are shown in Fig. 7.2. The first feature notable in Fig. 7.2a is the anti-phase evolution of the eccentricities of the planets **c** and **d**: when one eccentricity reaches its maximal value, the other is exactly at its minimal value, and vice versa. A simple examination of the expression (7.2), under

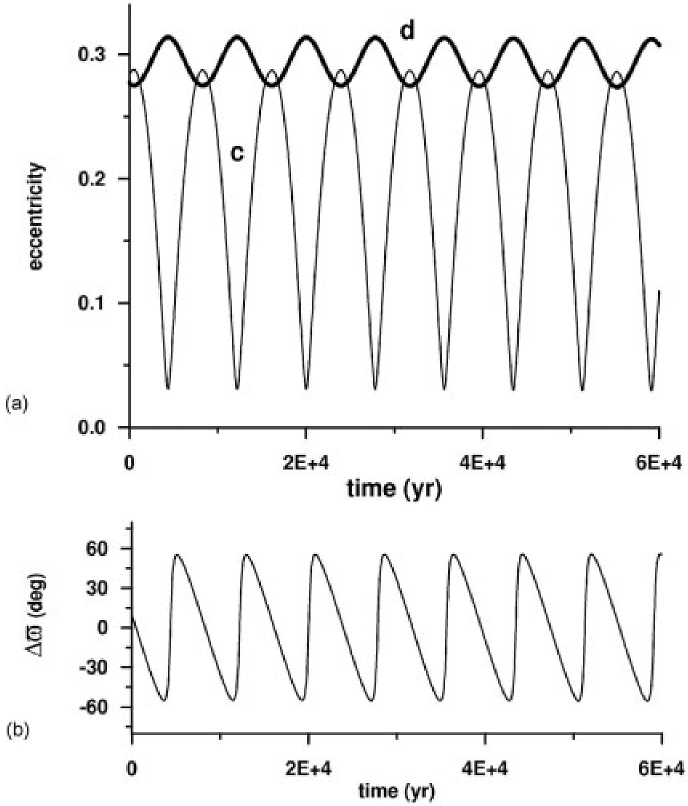


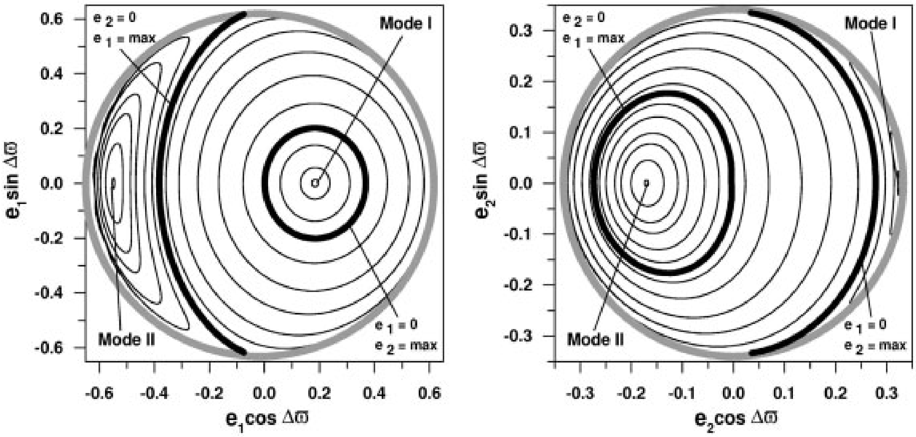
Fig. 7.2 (a) Time evolution of the eccentricities of the orbits of the υ Andromedae planets **c** and **d**, for the present best-fit initial conditions. (b) The secular variation of the angle $\Delta\varpi = \varpi_2 - \varpi_1$.

conditions $I_1 = I_2 = 0$ and constant a_1 and a_2, reveals that the coupled variation of the eccentricities is a direct consequence of the conservation of the total AM (or AMD).

7.4.1
Planar Dynamics

One important feature of secular dynamics can be observed in Fig. 7.2b. It is the almost exact alignment of the apsidal lines of the outer planets. The difference of the longitudes of periastra of the two planets oscillates with small amplitude around 0. From the known extrasolar systems, several pairs are found in the aligned modes of motion. In our solar system, the angle between Jupiter's and Uranus's perihelions oscillates around 180° (anti-alignment), with an amplitude of 70° and a period of 1.1 Myr [25]. This phenomenon is sometimes referred to as “secular resonance” or “apsidal resonance” in the literature (see, for example [26]). However, as we will see later, the apsidal alignment/anti-alignment in this case can be interpreted as a simple circulation in the phase space around a center displaced from the origin and is not related to any resonance.

Let us consider the problem of the secular motion of two planets in detail. Applying the Hamilton–Jacobi approach (7.1), we calculate the secular planar orbits of the edge-on outer υ Andromedae planets, with the fixed value of AMD (corresponding to the current υ Andromedae system). Figure 7.3 displays the solutions of the equations of secular motion, in the two eccentricity phase planes, $(e_i \cos \Delta\varpi, e_i \sin \Delta\varpi)$: the left panel shows the secular variations of the inner planet and the right panel shows the secular variations of the outer planet. In each plane we can see two fixed points (centers), known as Mode I and Mode II. Mode I corresponds to the aligned configuration of the planets with the minimum possible energy of the system, for a fixed value of AMD and $m_2/m_1 > 1$. At the same time, Mode II describes anti-aligned configuration and corresponds to the maximum values of the energy. For $m_2/m_1 < 1$, on the other hand, the energies of the Mode I and Mode II solutions are maximum and minimum, respectively. The initial conditions chosen exactly in one of these centers correspond to stationary secular solutions, with eccentricities constant in time and $\Delta\varpi = 0$ or 180°.



(a) (b)

Fig. 7.3 Secular solutions of the υ Andromedae planets **c** and **d**, at AMD $= 0.025$. Locations of the Mode I and Mode II are given by the fixed points in (a) and (b). The thick black curves represent orbits passing through the $e_1 = 0$ and $e_2 = 0$. The broad gray curves bound the energy manifold.

The smooth curves surrounding each of the two fixed points are quasi-periodic solutions. Even though the motion of the angle $\Delta\varpi$ may be either oscillation (about 0 or 180°) or circulation, there is no separatrix between them. To better understand this feature, we plot in thick black curves, two particular solutions in Fig. 7.3. These solutions are associated with singularities in the equations of motion, which occur at $e_1 = 0$ and $e_2 = 0$. One solution was obtained with initial condition $e_1 \approx 0$ and is seen as a smooth curve passing through the origin in the e_1 phase plane (a). In comparison, the solution, obtained with the initial condition $e_2 \approx 0$, appears as a "false" separatrix in the e_1 phase plane, separating the domains of the two different kinds of motion. In the e_2 plane (b), this solution is a smooth curve passing through the origin and the other one separates the domains of two types of motion.

The separatrix-like curves in Fig. 7.3 are due only to the topological inadequacy of the plane to represent the solutions of the secular system. As shown in Pauwels [27], these solutions are better seen on a sphere (Pauwels' sphere), which two opposite sides are shown in Fig. 7.4. On the sphere, the longitude of one point is $\Delta\varpi$ and its latitude ϕ is given as $\sin\phi = 1 - 2\,I_1/\mathrm{AMD}$, where $I_1 = m_1\sqrt{G\,m_0\,a_1}(1 - \sqrt{1 - e_1^2})$. The north pole corresponds to $e_1 = 0$ and the south pole to $e_2 = 0$. The axis through the centers Mode I and Mode II is inclined with respect to the vertical axis and the planetary orbits appear as a family of parallels with respect to this axis. The motions between the thick black curve and the Mode I center, the paths of which do not enclose the ver-

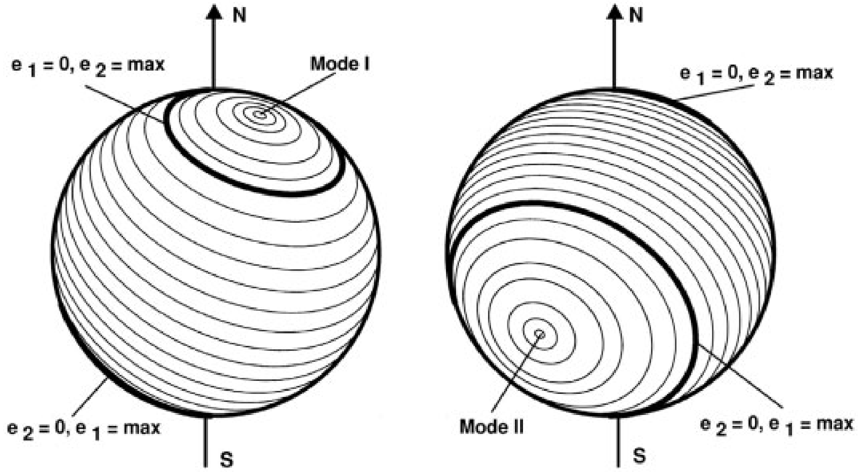


Fig. 7.4 The same phase space as Fig. 7.3 presented on the Pauwels' sphere, of which two sides are shown.

tical axis, are oscillations of the angle $\Delta\varpi$. The same is true for the motions between the other thick curve and the Mode II center. The only motions enclosing the vertical axis, between the two thick curves, show circulations of the angle $\Delta\varpi$. Figure 7.4 clearly shows that oscillations and circulations may be continuously transformed from one into another, because no topological separatrix exists in the phase space. To make this fact clear in the following, we will use the composite word "circulation/oscillation" to indicate the regime of motion where the two kinds of motion co-exist, but are not topologically separated.

7.4.2
Dynamical Maps

Figure 7.5a shows the locations of Mode I (at $\Delta\varpi = 0$) and Mode II (at $\Delta\varpi = 180°$) centers in the (e_1, e_2) plane of initial conditions, constructed with the dynamical parameters of the edge-on outer υ Andromedae system. Negative values of e_1 correspond to $\Delta\varpi = 180°$. We also show the domains of oscillation and circulation of $\Delta\varpi$ and the location of the actual system (star symbol). In Fig. 7.5b we show the dynamical map of the space with the same initial conditions. The map was obtained by direct numerical integrations of the unaveraged system with initial eccentricities distributed on a grid of 100×61 values and with the aligned (e_1 positive) and anti-aligned (e_1 negative) periastra. The initial inclination was fixed at 1°. A spectral indicator was used to distinguish between regular and chaotic regions [28]. The white regions correspond to very regular oscillatory motion around the Mode I and Mode II (the loci of the periodic orbits corresponding to the two modes are shown by the continuous lines). The gray regions correspond to regular circulations (the gray tone just indicates a larger number of harmonics in the power spectrum of the solution). The black boundary corresponds to chaotic motion beyond which lies the hatched region where planetary collisions occur within 0.5 Myr.

The secular behavior of the υ Andromedae system is stable over a large domain of initial eccentricities. Figure 7.5a shows the domain of stability bounded by the thick curve. This curve was calculated with the Hill criterion, which provides a critical (minimal) separation between two planets, given, to lowest order in masses, by Gladman [29], Marchall and Bozis [30], Roy et al. [31]

$$\Delta_c = 2.4 \left(\frac{m_1}{m_0} + \frac{m_2}{m_0} \right)^{1/3}$$

The high eccentricity domain above the thick curve is the region where the planets approach closely each other. The comparison of the two panels shows that for large eccentricities this criterion is very optimistic.

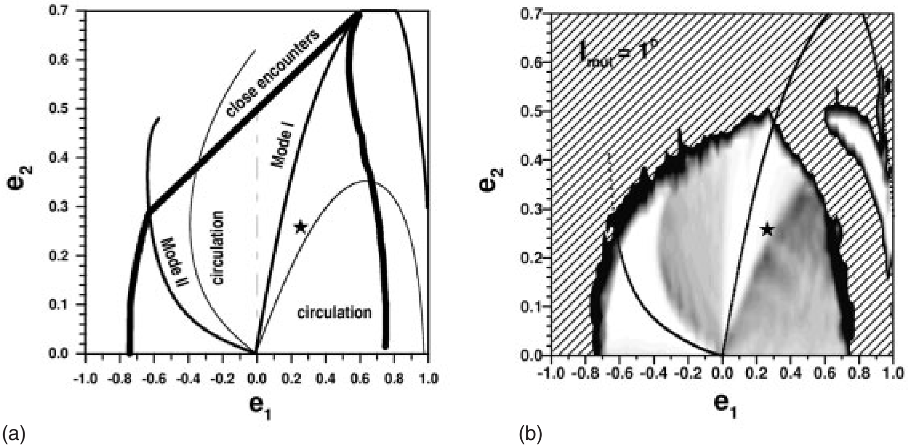


Fig. 7.5 (a) Periodic orbits and domains of oscillation and circulation in the (e_1,e_2) plane of initial conditions, obtained for the parameters of the υ Andromedae system. (b) Dynamical map calculated with $I_{\text{mut}} = 1°$. The curves are the location of analytically calculated periodic orbits.

The dynamical map in Fig. 7.5b also shows a robust region of regular motion near the right border, where the orbits of the inner planet can reach eccentricities as high as 0.95, without the appearance of any instabilities for 530 000 years. This feature is associated with the advent of a true nonlinear secular resonance [18]. The nonlinear secular resonance may be a common feature of the planetary system, particularly, when the masses and axes ratios are not close to 1. When the planetary masses are equal and their mutual distances are small, the nonlinear secular resonance does not occur; However, under these conditions, the probability of finding the system in oscillation (around 0 or 180°) is very close to 1. When the semi-major axis ratio is far from 1 and $m_2/m_1 > 1$, the domains of oscillation of $\Delta\varpi$ decrease, and the nonlinear secular resonance zone appears. Its location on the (e_1, e_2) plane of initial conditions is in the regions of the very high eccentricities of the inner planet. The orbit of the outer planet in this case can be nearly circular. Interestingly, the nonlinear secular resonance does not occur for all considered values of the axes ratio, if the mass ratio is less than 1. Up to now, no real system is known, which can be associated with this dynamical state. However, the regime of planetary parameters where this feature is significant is not particularly extreme, and it would not be surprising if such systems are identified in the future.

7.4.3
3D Dynamics

In this section we present results on the non-coplanar secular motion of the planets. It is the case that, at present, the inclinations of almost all the extrasolar systems are unknown. In fact, the spectroscopic radial velocity technique measures only the line-of-sight component of the velocity of the star and the fitting of Keplerian ellipses is unable to detect the inclinations and nodes of the planetary orbits. However, dynamical stability considerations can provide constraints on the individual planetary masses and inclinations and, despite the detection difficulties, it is important to understand the spatial evolution of the systems.

In a 3D system, the conservation of the total angular momentum allows exchanges between planet eccentricities and inclinations. If the invariable plane (a plane orthogonal to the total angular momentum vector) is chosen as a reference plane, both inclinations are not independent, but are constrained, in the case of two planets, by the condition [32]

$$G_1 \sin I_1 = G_2 \sin I_2 \tag{7.3}$$

where, up to second order in masses, $G_i = m_i n_i a_i^2 \sqrt{1 - e_i^2}$ $(i = 1, 2)$. Thus, instead of two inclinations, it is enough to consider the mutual inclination between the planetary orbits, which is just $I_{\text{mut}} = I_1 + I_2$, when the invariable reference plane is used. Moreover, for a fixed value of AM, the mutual inclination is determined by the eccentricities through

$$\cos I_{\text{mut}} = \frac{\text{AM}^2 - G_1^2 - G_2^2}{2 G_1 G_2} \tag{7.4}$$

To show all possible regimes of the 3D secular motion, we present in Fig. 7.6 the dynamical map of the system, for two values of AMD. One is the value corresponding to the actual edge-on outer υ Andromedae system, AMD $= 3 \times 10^{-3}$ (in units of the solar mass, astronomical unit and year). The other value is larger, AMD $= 8 \times 10^{-3}$, and corresponds to the actual planetary orbits, but with a mutual inclination of 60°. In the construction of the dynamical maps of Fig. 7.6, a grid of initial conditions was defined on the (e_1, e_2) planes, with steps $\Delta e_1 = 0.02$ and $\Delta e_2 = 0.01$. On the (e_1, I_1) planes, a grid of initial eccentricity and inclination of the inner planet was defined with steps $\Delta e_1 = 0.02$ and $\Delta I_1 = 1°$. The signs $+$ or $-$ preceding the variable e_1, indicate aligned and anti-aligned periastron lines, respectively. The signs $+$ or $-$ preceding the variables e_2 and I_1 indicate that the initial values of $2\omega_1$ are zero or 180°, respectively. The location of the edge-on outer υ Andromedae system is indicated by a star.

The shading scale used in the dynamical maps in Fig. 7.6 is related to the degree of stochasticity of the solutions: the lighter regions in the dynamical

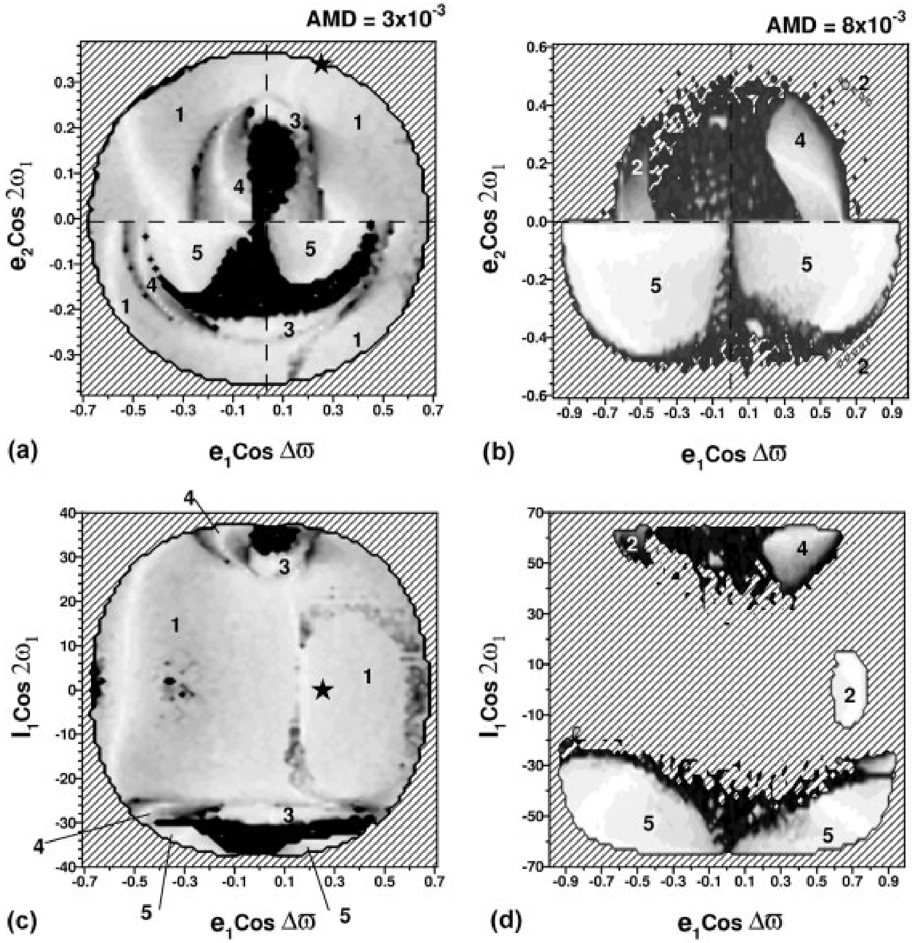


Fig. 7.6 Dynamical maps for AMD $= 3 \times 10^{-3}$ on the (e_1,e_2) plane (a) and (e_1,I_1) plane (c). The domains of the different regimes of motion marked by labels from **1** to **5** are described in the text, (b) and (d) Dynamical maps for AMD $= 8 \times 10^{-3}$.

maps correspond to initial conditions of regular motion, darker tones indicate increasingly chaotic motion. The hatched regions are regions of large-scale instability followed by disruption of the system within the time-interval of 530 000 years. The analysis of the structure of the dynamical maps reveals interesting topological properties and reflects important features of the 3D secular dynamics which we summarize in the following (for more detail see [19]).

1. The low-to-moderate eccentricity and mutual inclination regime of motion (domain **1** with $e_1 < 0.6$ and $I_{\mathrm{mut}} < 30°$). This is a general circulation/oscillation regime, similar to that found in the planar case. The sys-

tems always exhibit two main kinds of secular motion, characterized by circulation of $\Delta\varpi$ or its oscillation around 0 or 180°. The arguments of the pericenter are in regular direct circulation. The edge-on outer υ Andromedae system is in this regime of motion and its secular behavior is stable; weak instabilities seem to occur only in the vicinity of the high-order 16:3 mean-motion resonance (see Fig. 7.7).

2. The high eccentricity and low-to-moderate inclination regime (domain **2** with $e_1 > 0.6$ and $I_{\mathrm{mut}} < 10°$) is characterized by large-scale instabilities, due to close approaches of the planets, followed by disruption of the system within a few thousands of years. The only surviving solutions in this region are those inside the nonlinear secular resonance and are bounded by a zero-frequency separatrix. The secular angle $\Delta\varpi$ librates around 0 and the variations of e_1 and e_2 are strongly coupled. This feature of the secular dynamics of two-planet systems is the same as the one that appeared in the studies of the planar problem.
3. The high-inclination regimes of motion (domains **3–5** with $I_{\mathrm{mut}} > 30°$): complex dynamical behavior with the presence of several regimes of resonant motion. The dominating behavior is the e–I coupling, or the Lidov–Kozai resonance [33], characterized by the coupled variation of the eccentricity and inclination of the inner planet and the libration of the angle ω_1 around ±90°. At variance with the analogous phenomenon in restricted problems, the variation of the planet inclinations and eccentricities is constrained by the total angular momentum conservation. The large eccentricity/inclination excursions induced by the Lidov–Kozai resonance in restricted problems cannot occur in the planetary problem. A regime of motion with $\Delta\varpi$ in the secular resonance also exists in the high-inclination region. In this case, the secular angle $\Delta\varpi$ librates either around 0 or 180°.

7.4.4
Inclination and Planetary Mass Indetermination

The main problem in applying the 3D secular model to the known planetary systems, is that the observational data derived from radial velocity measurements do not constrain the individual orbital inclinations. Hence, at the first instance, we always assume that both planetary orbits are observed edge-on, so the spatial configuration of this system corresponds to two co-planar planetary orbits placed in the invariable plane, whose inclination i to the sky plane is 90°.

Up to second order in masses, the secular behavior of the systems with small mutual inclinations is independent of the inclination of the invariable plane to the plane of the sky. This is a consequence of the fact that the main features of secular motion of planetary systems are independent of the individual values of the planetary masses and semi-major axes, but only on their ratios. In the case of co-planar orbits, the mass ratio may be determined from the observations, being unaffected by the indetermination of the individual masses. As a consequence, the results of the edge-on system obtained in the study of the co-planar dynamics for a given AM, are valid for non edge-on systems with the angular momentum equal to AM/ $\sin i$. This conclusion allows us to partially overcome the problem of the mass indetermination, as far as the limit $\sin i \sim 0$ (the planetary masses tend to be very large) is not approached. If $I_{\text{mut}} \neq 0$, the determination of the mass ratio is affected by the unknown factor $\sin i_1 / \sin i_2$, where i_1 and i_2 are the individual inclinations to the sky plane of the inner and outer orbits, respectively.

7.4.5

Proximity to Mean-motion Resonances

When we study the stability of the real systems, special care must be taken with the proximity of the systems to the main mean-motion resonances. In the case of the outer υ Andromedae system, this system is not far from the 5/1 mean-motion resonance (Fig. 7.1). At variance with the secular case, there is no evidence that the nearly resonant behavior depends only on the mass ratio. Therefore the dynamical structure of the phase space may be strongly sensitive to the inclination of the orbits to the plane of the sky.

Figure 7.7 shows the dynamical maps of the neighborhood of the current position of the outer planet **d** in two numerical experiments. (a) shows the edge-on outer υ Andromedae system, with the masses $m_1 = 1.83$ M_{Jup} and $m_2 = 3.97$ M_{Jup}, and (b) shows the same system, but with the masses $m_1 = 2 \times 1.83$ M_{Jup} and $m_2 = 2 \times 3.97$ M_{Jup}, which corresponds to inclinations of 30° to the plane of the sky. (Note that the mass ratio is the same in both sets). The shading scale was used to distinguish between the regions of regular and chaotic motion. In both panels, the hatched regions are characterized by large-scale instabilities followed by disruption of the system within the time interval of 530 000 years. The domains of chaotic motion (dark tones) are associated with mean-motion resonances. We note the dominating presence of the 5:1 resonance in the region. Weaker resonances of higher order appear as black strips of variable width: they are labeled by the corresponding symbols on the top of the panels. The location of the υ Andromedae system is marked by a star symbol in both panels.

The edge-on system shown in Fig. 7.7a is far away from the region of the strong 5:1 resonance; on the other hand, it is close to the weak 16:3 mean-motion resonance. The situation does not seem to be dramatic in this case: due to its high order, the 16: 3 resonance is very narrow. Moreover, due to

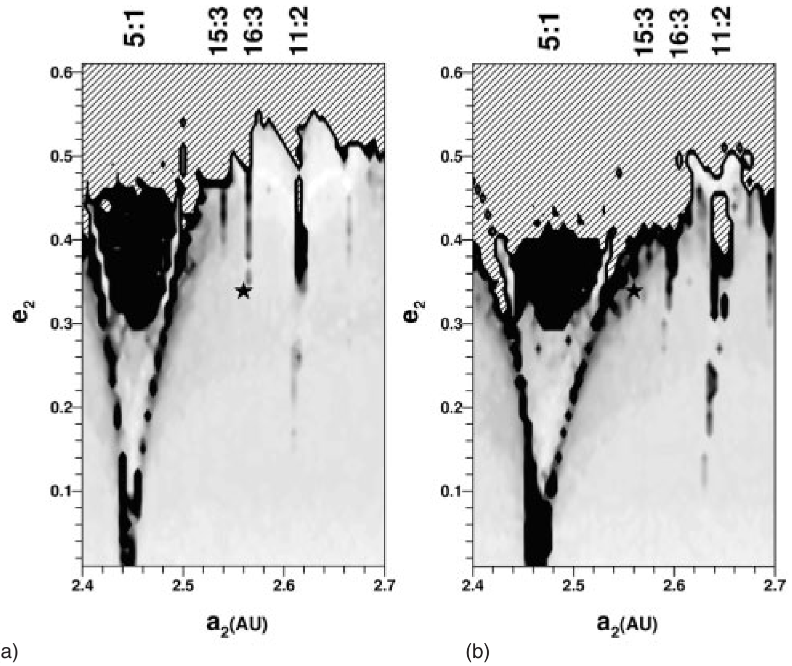


Fig. 7.7 Dynamical maps of the region around the υ Andromedae system on the (a_2, e_2) plane of initial semi-major axis and eccentricity of the outer planet **d**. The initial mutual inclination is equal to 1°. (a) Map calculated with planetary masses of the current edge-on system, $m_1 = 1.83\,\mathrm{M_{Jup}}$ and $m_2 = 3.97\,\mathrm{M_{Jup}}$. (b) The same, but with masses $m_1 = 2 \times 1.83\,\mathrm{M_{Jup}}$ and $m_2 = 2 \times 3.97\,\mathrm{M_{Jup}}$. The high-order mean-motion resonances are labeled on the top of the graph. The domains, where planetary collisions occur within 0.5 Myr, are hatched. The position of the edge-on outer υ Andromedae system is indicated by a star.

constraints imposed by the conservation of the total angular momentum, this system may survive for a time comparable with the age of the star [34]. The situation may be considered dramatic in the case of the second set of planetary masses, corresponding to $\sin i = 0.5$ (Fig. 7.7b). In that case, the region of large-scale instabilities increases and all mean-motion resonances are shifted towards higher values of the semi-major axes. The system under study is now located very close to the strong 5:1 mean-motion resonance and undergoes large-scale instabilities.

7.5
Resonant Dynamics: Planetary Systems in Mean-Motion Resonances

When the ratio of the orbital periods of two planets is close to a ratio of two simple integers, we say that the planets are close to a mean-motion resonance. A mean-motion resonance (hereafter just a resonance) is often written in generic form as $(p+q)/q$, where p and q are integers, the latter representing the order of the resonance. The critical (or resonant) angles of a resonance are defined as $\sigma_i = (p+q)\lambda_2 - q\lambda_1 - q\varpi_i$, where λ_i are mean longitudes of the planets. The dynamical behavior of the critical angles defines the location of the system with respect to the resonance: when one of these angles is oscillating, the system is inside the resonance.

The averaging procedure shown in Eq. (7.1), together with short-period terms, removes all effects of mean-motion commensurabilities. In the resonance case, it is necessary to adopt a different rule: the averaging of the system with respect to the synodic angle $Q = \lambda_2 - \lambda_1$. Therefore, the secular and resonant perturbations, which are of long period, remain in the Hamiltonian. The averaged Hamiltonian (total energy) is defined, to second order in masses, by

$$\overline{\mathcal{H}}_{\text{res}} = \sum_{i=1}^{2} \frac{G m_0 m_i}{2\, a_i} - \frac{1}{2\pi} \int_0^{2\pi} \mathcal{R}(a_i, e_i, I_i, Q, \sigma_i, \omega_i)\, dQ \tag{7.5}$$

where $\mathcal{R}$ is the disturbing function. The Hamiltonian function is very complicated, even in the planar case, and generally possesses several extrema, which define the stationary solutions of the averaged problem. These solutions are often referred to as *apsidal co-rotation resonances* (ACR).

In contrast with the secular behavior, the semi-major axes are no longer invariant, but their variation is constrained by

$$(p+q)\, m_1\, n_1\, a_1^2 + p\, m_2\, n_2\, a_2^2 = \text{const.} \tag{7.6}$$

This means that the averaged semi-major axes of the planets always oscillate with opposite phases and with amplitudes that are inversely proportional to the planet masses. Finally, it is worth emphasizing that the total angular momentum preserves the general form given in Eq. (7.2) and is conserved. However, because of Eq. (7.6), it no longer implies a correlated variation of the eccentricities.

7.5.1
GJ 876 Planets c and b

In the planar case, after averaging over Q, the Hamiltonian is reduced to two degrees of freedom, to which correspond two angular variables. In what follows, we adopt the secular angle between the apsidal lines, $\Delta\varpi = \varpi_1 - \varpi_2$, and the resonant angle σ_1 (the second resonant angle σ_2 is equal to $\sigma_1 + q\Delta\varpi$). The structure of the phase space can be studied through the construction of surfaces of sections. As an example, in Fig. 7.8 we present the surfaces of

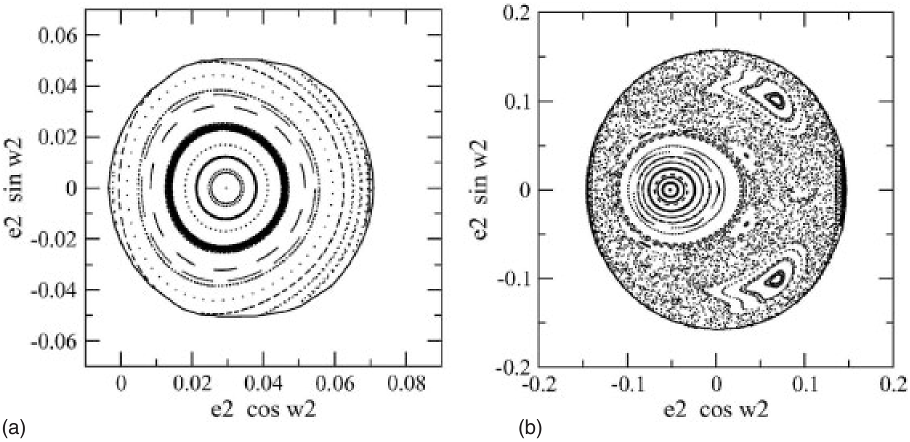


Fig. 7.8 Surfaces of section of an outer planet, by the plane $\sin \sigma_1 = 0$ when $\dot{\sigma}_1 < 0$, for two different energy levels. The sections in (a) were constructed along the energy level corresponding to the current best-fit configuration of the GJ 876 system. The continuous curves show the boundaries of the energy manifold. The sections corresponding to the actual GJ 876 system are shown by thick curves. Sections in (b) correspond to a lower energy system with the same masses.

section of the GJ 876 system (planets **c** and **b**), which evolve inside the 2/1 mean-motion resonance (see Fig. 7.1), for two different energy levels.

Figure 7.8a corresponds to the best-fit configuration of the GJ 876 system and (b) corresponds to a lower energy system with the same masses. The first set is close to the global stable stationary solution corresponding to the energy maximum and no domains of chaotic motion are seen in the section. Both critical angles and the secular angle are in libration around 0; in this case, we say that the system is near to (0, 0) ACR. The lower energy system is distant from the stationary solution and presents a complex dynamical structure (b). The zones of stable motion are reduced to small islands surrounded by a sea of chaotic motion. In the main libration regime (the island near the right-hand boundary in the section of the inner planet and near the origin in the section of the outer planet), the critical angle σ_1 remains librating around 0, but the secular angle $\Delta\varpi$ is now librating around 180°. The other two islands are characterized by the libration of σ_1 and the circulation of $\Delta\varpi$. When the energy decreases, the size of the regions of stable motion also decreases and they finally shrink into unstable equilibrium points. Thus, the surfaces-of-section study shows a robust domain of stable motion close to an apsidal corotation resonance.

ACRs are classified in different types, depending on the motion of the resonant and secular angles. The possible combinations are (0, 0), (π, 0), (0, π) and (π, π), where the numbers inside the brackets indicate the stationary values of σ_1 and $\Delta\varpi$. For example, a (0, π) ACR is a solution for which $\sigma_1 = 0$

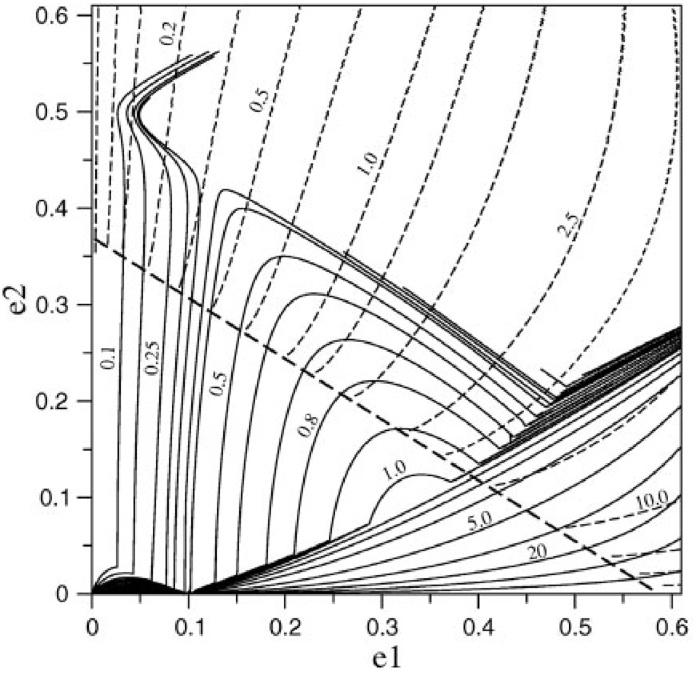


Fig. 7.9 Stable apsidal co-rotation solutions, for given mass ratios m_2/m_1, in the 2/1 resonance, for eccentricities up to 0.6. Continuous curves: $(0,\pi)$, (0,0) and asymmetric solutions. Broken curves: (π,π) co-rotations.

and $\Delta\varpi = \pi$. These solutions are known as symmetric ACRs. There are also asymmetric solutions which are characterized by libration of the critical angles about values which are different from zero or π. The global view of ACRs in the 2/1 mean-motion resonance is shown in Fig. 7.9 [35] (for the other resonances see [36]).

In Fig. 7.9 we present the values of the eccentricities corresponding to apsidal co-rotation resonances. Each curve corresponds to a mass ratio m_2/m_1. The continuous lines show the symmetric $(0, \pi)$, (0, 0) and asymmetric solutions, while the (π, π) solutions are presented in broken lines (the family of $(\pi, 0)$ ACR does not exist in this resonance). The dashed line is the locus of configurations leading to possible collisions between the planets. This curve can be expressed by $a_1(1 - e_1) = a_2(1 - e_2)$. Note that (π, π) co-rotations are beyond the collision curve. For very small masses, these solutions can even be detected in the close vicinity of the collision curve. But it is worth observing that, for finite planetary masses, the stability of (π, π) co-rotations cannot be guaranteed below the Hill stability criterion (see Section 7.4.1).

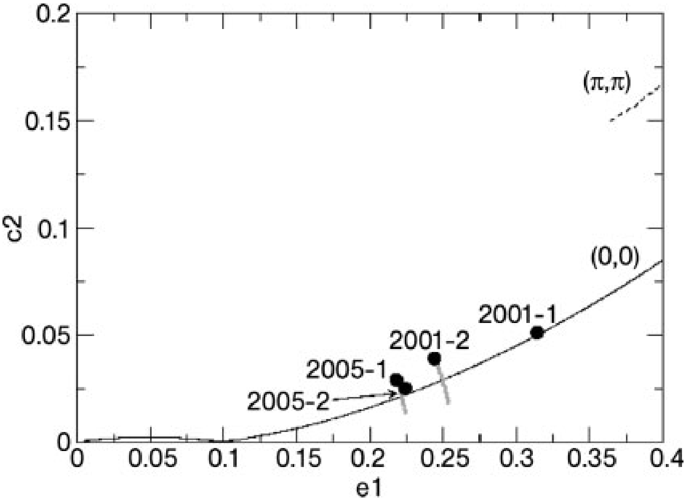


Fig. 7.10 Orbital evolution of GJ 876 **c**,**b** system on the (e_1,e_2) plane. Filled circles show four different orbital fits, while gray curves show the corresponding orbital paths. All fits are characterized by small amplitudes of libration around ACR solutions. For mass ratio $m_2/m_1 = 3.2$, (0,0) and (0,π) ACR are drawn by continuous black lines, while (π,π) solutions are shown by broken curve. (From Beaugé et al. [7]).

Figure 7.10 shows the eccentricities in different orbital fits of the system GJ 876 (**c**,**b**) [37–39] and their temporal variation during a 10^5 yr time span. It is noteworthy that all observational fits lie very close to the zero-amplitude resonant solutions, for the mass ratio 3.2. This fact suggests that resonant exoplanets acquired their present orbits as a result of migration from initially quasi-circular orbits and capture into commensurabilities [7]. Indeed, hydrodynamical simulations of the evolution of two planets immersed in a gaseous disc, have always shown co-rotational final orbits (e.g. [40–44], etc).

7.5.2
HD 82943 Planets c and b

The first available orbital fit for this system was posted in the Geneva extrasolar planet search programme homepage in 2002. According to the posted data, the system should be trapped in a (π, π) co-rotation, which lies above the collision curve [46]. Nevertheless, this configuration is not consistent with a smooth planetary migration. Indeed, starting with initially circular orbits, the system could reach the (π, π) ACR domain only through collisional events [47]. In 2005, updated observational data permitted to obtain a new orbital fit by Mayor et al. [45], significantly different from the previously published values. This fit placed the system in oscillation about the (0, 0) co-rotation solu-

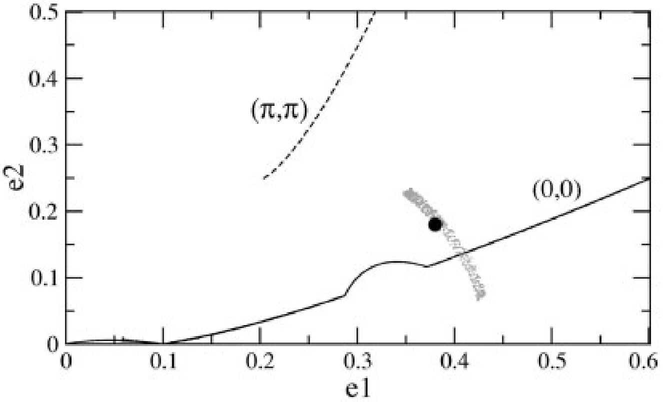


Fig. 7.11 The same as the previous figure, but for the HD 82943 planets, with $m_2/m_1 = 1$ and the orbital fit from [45]. (From Beaugé et al. [7]).

tion. Figure 7.11 shows the best-fit solution for orbital parameters [45] and the temporal variation of planet eccentricities calculated over 2×10^3 yr time interval. We note that, although a co-rotation around the (0,0) symmetric family is confirmed, it shows a very large amplitude.

A longer numerical simulation shows that the published orbits of the HD 82943 planets undergo a major catastrophe in about 50 000 yr, and one of the planets is expelled from the system (or collides with the star) in less than twice that time [48]. In the first 10 000 yr of the simulation, the solution lies in the 2:1 resonance zone, and one of the two critical angles related to this resonance appears in libration. However, the permanence of the system in a resonant state is only temporary. Since the age of the star is of the order of 10^9 yr, this orbital fit does not correspond to an acceptable system.

A new effort to determine the orbits of the HD 82943 planets was made by Ferraz-Mello et al. [48]. It was shown that a relatively large number of orbital fits are possible which satisfy the observational data of HD 82943 almost as well as the least-squares solutions. A dynamical analysis of these fits showed that many of them still yield unstable orbital configurations. However, some of the "good fits" allow the system to remain in the neighborhood of a (0, 0) apsidal co-rotation for long time intervals and, consequently, are compatible with a co-rotation/migration scenario. A new attempt, adding new observations to the set considered in the previous determination, leaded to best-fit solutions showing the same instabilities as the one discussed above [49]. Further observations may help towards a better understanding of the system. So far, the orbital uncertainties are such that they allow us even to obtain a

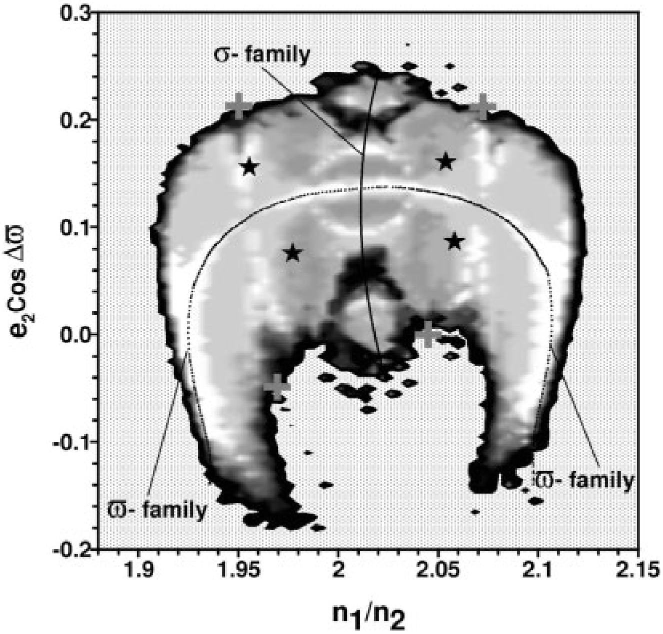


Fig. 7.12 Dynamical map of the 2/1 resonance on the plane (n_1/n_2,e_2), obtained for $e_1 = 0.41, \sigma_1 = 0, \omega_1 = 90°$. The initial mutual inclination is equal to $1°$. The conditions with initially aligned periapses lie in the positive half-plane and anti-aligned in the negative half-plane. Periodic solutions of the resonant Hamiltonian, corresponding to librating σ_1 and $\Delta\varpi$ form the σ_1 and $\Delta\varpi$ families, respectively. Intersections of the best-fit solution are shown by crosses and the solution B from [48] by stars.

completely different configuration which fits the observed RV variations of the parent stars as being induced by a pair of Trojan planets (that is, in 1:1 MMR) [50].

To understand the dynamics of this system, we consider a dynamical map covering the region around the best-fit solution and present it on the (n_1/n_2, e_2) plane (Fig. 7.12). This dynamical map was constructed adopting the initial mutual inclination of $1°$. We can see the domains of stable motion of the resonance in gray tones, while the hatched regions are regions of large-scale instability followed by disruption of the system within the simulation time interval (530 000 years). The intersections of the best-fit solution given by Mayor et al. [45] with this plane are displayed by crosses, and the intersections of the good fit labeled as B in Ferraz-Mello et al. [48] are displayed by stars. It is clear from Fig. 7.12 that the best-fit solution places the system on the border of the stable domain of the 2/1 resonance; this fact fully explains the short-living chaotic behavior of the system. At variance, all intersections of the given good fit lie in the domain free from instabilities. It is worth noting that the r.m.s. of the residuals in the given good fit and in the best fit are almost the same.

The interpretation of the topological properties of the 2/1 resonance is facilitated by plotting two families of periodic solutions in Fig. 7.12. These solutions were obtained with the analytical model. The almost-vertical curve presents the location of stable periodic solutions where $\sigma_1 = 0$. The almost-horizontal curve shows the periodic solutions in which $\Delta\varpi = 0$. By definition, the intersection of two families gives the location of the (0, 0) ACR, which is characterized by constant eccentricities and semi-major axes. The narrow white strips coincide with the location of the families of periodic solutions. The stable domain around (0, 0) ACR is dominated by a gray background of regular resonant motion of the system. We can see, in Fig. 7.12, that it is possible to find initial conditions of stable motion (light regions) with large libration amplitudes; the only condition is the proximity of these solutions to the family $\Delta\varpi = 0$.

For small mutual inclinations, the resonant motion is composed by two normal components: one is an oscillation around periodic solutions of the family $\sigma = 0$, and the other one is an oscillation around periodic solutions belonging to the family $\Delta\varpi = 0$. For high mutual inclination, the analysis of the third dimension requires a more detailed study.

7.6
Nearly Resonant Systems with Small Eccentricities

The most curious feature which can be observed in Fig. 7.1 is the existence of no more than a few pairs of planets with small eccentricties (< 0.1). Beside the three pairs of the outer solar system (dots), the only known systems with this characteristic are the three planets in the PSR 1257+12 system and the 47 Uma planets. It should be noted that there is no evidence that low-eccentricity orbits are able to produce apparent bias in observational data.

We place these pairs of planets in the subclass with near-resonant motion. Such systems are not trapped in a mean-motion resonance, but reside in close proximity to one, in such a way that resonant effects may be observable on their dynamics. The existence of near resonances in the outer solar system has been known for many years. All Jovian planets are in the immediate vicinity of significant commensurabilities between their mean motions. The most famous of these commensurabilities involves Jupiter and Saturn, whose orbital periods are such that $T_{\mathrm{Sat}}/T_{\mathrm{Jup}} \approx 5/2$, which is responsible for the large-longitude perturbations known as *Great Inequality*. The associated mode of motion has a period of $\sim 880\,\mathrm{yr}$ and plays an important role in the depletion processes among the minor planets' population [51–53].

The other near-resonances in the solar system involve the orbits of Saturn and Uranus which lie close to the 3/1 commensurability, and the orbits of

Uranus and Neptune, close to the 2/1 commensurability. In the extrasolar systems, the pair of the planets **c** and **d** around the pulsar PSR1257+12 is close to the 3/2 mean-motion resonance [54], and the pair **b** and **c**, to the 5/2 resonance. The closeness of the outer pulsar planets to commensurability produces perturbations large enough to be observed from Earth thus confirming the existence of the planets [55, 56] and allowing us to determine their masses, thus solving the inclination indetermination. Finally, the orbits of two planets of the 47 Uma system seem to be close to the 7/3 resonance.

To understand the dynamics of the near-resonant configurations, we present a dynamical map in the neighborhood of the present position of Saturn in Fig. 7.13. The dynamical map was obtained by simulation of the actual elements of the outer planets, except Saturn's osculating semi-major axis and eccentricity, which were varied on a 251×21 grid [34]. The actual position of Saturn is marked by a plus sign. The values of the spectral indicator, obtained in the range from 1 to 100 over 1.5 Myr, are coded by gray levels that vary logarithmically from white to black and are plotted on the (a_S, e_S) plane of initial osculating orbital elements. The lighter regions indicate regular motion, whereas the darker regions indicate chaotic motion. The domains where planetary collisions occur within 50 Myr are hatched.

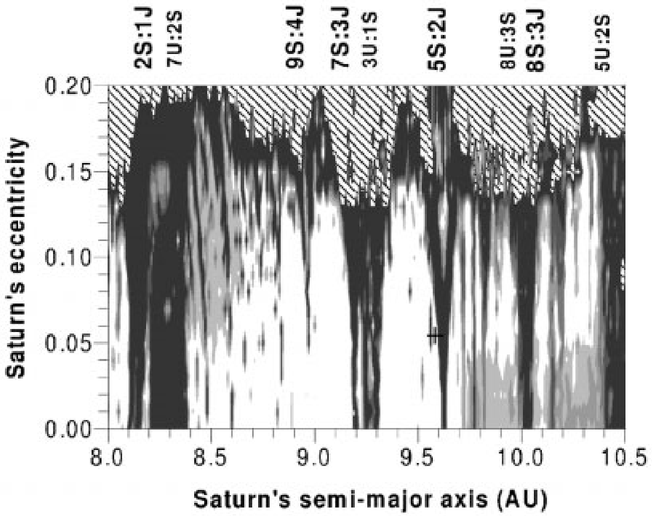


Fig. 7.13 Dynamical map of the region around Saturn. The actual position of Saturn is indicated by a plus sign.

The main mean-motion resonances between the planets exhibit chaotic motion and appear in Fig. 7.13 as black "stalactites" of different widths, depending on the order of the resonance. Two-planet mean-motion resonances are labeled on the top of the graph in the form $n\,P_i : m\,P_j$, where P denotes the planet (J, S and U for Jupiter, Saturn and Uranus, respectively) and n and m are the integers that appear in the critical argument of the corresponding resonance. There are also several narrow vertical bands of chaotic motion associated with three-planet resonances (see [34]).

The regions around the planets are densely filled by mean-motion resonances. Why are the planets near, but not quite in, the exact resonances? We can expect that the simultaneous resonant coupling between the planetary motions would allow a largely stable dynamical behavior. The most well known example is the system of three Galilean satellites of Jupiter, which are involved in a complex set of mean-motion resonances. We can see in Fig. 7.13 that important two-planet resonances are overlapping each other, generating large instabilities in planetary motion. This is due to the fact that the planets are only close to these resonances, but not inside them. As a consequence, the overlap of the resonances destroys stable domains of an individual resonance and the only possible robust domains of stable motion are in the low-eccentricity region.

We note that Saturn is presently located in an apparently regular region, very close to the 5S:2J resonance with Jupiter. We can conjecture that the actual configuration of the outer solar system is a natural arrangement of complex systems composed by several planets and that only nearly resonant configurations and low eccentricities permit us to put together relatively close massive planets. A similar phenomenon is not observed in the inner solar system. We might think that the mass factor could be responsible, because the width of a resonance is related to the individual planetary masses. Many questions related to the planet-system configurations have arisen from the discovery of extrasolar planets. We believe that this discovery will also provide many opportunities for improving our understanding of the evolution of planetary systems.

References

1 Wolszczan, A., and D. A. Frail: 1992, A planetary system around the millisecond pulsar PSR1257 + 12. *Nature* **355**, 145–147.

2 Mayor M. and D. Queloz: 1995, A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355–359.

3 Ferraz-Mello, S., T.A. Michtchenko, C. Beaugé, and N. Callegari: 2005. Extrasolar planetary systems. In *Chaos and Stability in Planetary Systems* (R. Dvorak et al., eds.), Springer-Verlag Berlin Heidelberg.

4 Ford, E.B., B. Kozinsky, and F. Rasio: 2000, Secular evolution of hierarchical triple star systems. *ApJ* **535**, 385–401.

5 Lee, M.H. and S.J. Peale: 2003, Secular evolution of hierarchical planetary systems. *ApJ* **592**, 1201–1216.

6 Ford, E.B.: 2005, Quantifying the uncertainty in the orbits of extrasolar planets. *AJ* **129**, 1706–1717.

7 Beaugé, C., T. A. Michtchenko, and S. Ferraz-Mello: 2006, Planetary migration and extrasolar planets in the 2/1 mean-motion resonance. *MNRAS* **365**, 1160–1170.

8 Zhou, L.-Y., H. J. Lehto, Y.-S. Sun, and J.-Q. Zheng: 2004, Apsidal corotation in mean-motion resonance: the 55 Cancri system as an example. *MNRAS* **350**, 1495–1502.

9 Correia, A. C. M., Udry, S., Mayor, M., Laskar, J., Naef, D., Pepe, F., Queloz, D., Santos, N. C.: 2005, The CORALIE survey for southern extrasolar planets. XIII. A pair of planets around HD202206 or a circumbinary planet? *A& A* **440**, 751–758.

10 Butler, R.P., G.W. Marcy, E. Williams, H. Hauser, and P. Shirts: 1997, Three new "51 Pegasi–Type" planets. *ApJ* **474**, L115–L118.

11 Goździewski, K., M. Konacki, and A. J. Maciejewski: 2005, Orbital Solutions to the HD 160691 (μ Arae) Doppler signal. *ApJ* **622**, 1136–1148.

12 Beaugé, C., and T. A. Michtchenko: 2003, Modelling the high-eccentricity planetary three-body problem. Application to the GJ876 planetary system. *MNRAS* **341**, 760–770.

13 Callegari, N., Jr., T. A. Michtchenko, and S. Ferraz-Mello: 2004, Dynamics of two planets in the 2/1 mean-motion resonance. *CMDA* **89**, 201–234.

14 Callegari, N., Jr., S. Ferraz-Mello, and T. A. Michtchenko: 2006, Dynamics of two planets in the 3/2 mean-motion resonance: application to the planetary system of the pulsar PSR B1257+12. *CMDA* **94**, 381–397.

15 Laskar, J., and P. Robutel: 1995, Stability of the planetary three-body problem: I. Expansion of the planetary Hamiltonian. *CMDA* **62**, 193–217.

16 Malige, F., P. Robutel, and J. Laskar.: 2002, Partial reduction in the N-body planetary problem using the angular momentum integral. *CMDA* **84**, 283–316.

17 Michtchenko, T. and S. Ferraz-Mello: 2001, Modeling the 5:2 mean-motion resonance in the Jupiter-Saturn planetary system. *Icarus* **149**, 357–374.

18 Michtchenko, T.A., and R. Malhotra: 2004, Secular dynamics of the three-body problem: application to the v Andromedae planetary system. *Icarus* **168**, 237–248.

19 Michtchenko, T.A., S. Ferraz-Mello, and C. Beaugé: 2006, Modeling the 3-D secular planetary three-body problem. Discussion on the outer v Andromedae planetary system. *Icarus* **181**, 555–571.

20 Brouwer, D., and G.M. Clemence: 1961, *Methods of Celestial Mechanics*. Academic Press Inc., New York.

21 Laplace, P.S.: 1799, *Mécanique Céleste*. English translation by N. Bowditch, Chelsea Pub. Comp. Edition, N.Y., 1966.

22 Ferraz-Mello, S., T.A. Michtchenko, and C. Beaugé: 2006, Regular motions in extrasolar planetary systems. In *Chaotic Worlds: From Order to Disorder in Gravitational N-Body Systems*, (B.A. Steves, ed.), Kluwer Acad. Publ. (in press). 22.

23 Laskar, J.: 2000, On the spacing of planetary systems. *Phys. Rev. Letters* **84**, 3240–3243.

24 Poincaré, H.: 1897, Sur une forme nouvelle des équations du problème des trois corps. *Bull. Astron.* **14**, 53–67.

25 Milani A., and A. M. Nobili: 1984, Resonance locking between Jupiter and Uranus. *Nature* **310**, 753–755.

26 Barnes, R., and Greenberg, R.: 2006. Extrasolar planetary systems near a secular separatrix. *ApJ* **638**, 478–487.

27 Pauwels, T.: 1983, Secular orbit–orbit resonance between two satellites with nonzero masses. *Cel. Mech.* **30**, 229–247.

28 Michtchenko, T.A., D. Lazzaro, S. Ferraz-Mello, and F. Roig: 2002, Origin of the basaltic asteroid 1459 Magnya: A dynamical and mineralogical study of the outer main belt. *Icarus* **158**, 343–359.

29 Gladman, B.: 1993, Dynamics of systems of two close planets. *Icarus* **106**, 247–263.

30 Marchal, C., and G. Bozis: 1982, Hill stability and distance curves for the general three-body problem. *Cel. Mech.* **26**, 311–333.

31 Roy, A. E., I. W. Walker, A. Carusi, and G. B. Valsecchi: 1984, The use of the energy and angular momentum integrals to obtain a stability criterion in the general hierarchical three-body problem. *A& A* **141**, 25–29.

32 Jacobi, C.G.J.: 1842, “Sur l’elimination des noeuds dans le problème des trois corps”. *Astronomische Nachrichten*, Bd XX, 81–102.

33 Lidov, R.: 1963, Analiz evolucii orbit iskustvennich sputnikov. *Problemi dvigenia iskustvennich nebesnich tel*, Izd. Akademii Nauk SSSR, Moscow, 119–134.

34 Michtchenko, T.A., and S. Ferraz-Mello: 2001, Resonant structure of the outer Solar System in the neighborhood of the planets. *AJ* **122**, 474–481.

35 Beaugé, C., S. Ferraz-Mello, and T. A. Michtchenko: 2003, Extrasolar planets in mean-motion resonance: apses alignment and asymmetric stationary solutions. *ApJ* **593**, 1124–1133.

36 Michtchenko, T.A., C. Beaugé, and S. Ferraz-Mello: 2006, Stationary orbits in resonant extrasolar planetary systems. *CMDA* **94**, 411–432.

37 Laughlin, G., and J. E. Chambers: 2001, Short-term dynamical interactions among extrasolar planets. *ApJ* **551**, L109–L113.

38 Laughlin, G., Butler, R. P., Fischer, D. A., Marcy, G. W., Vogt, S. S., Wolf, A. S.: 2005, The GJ 876 planetary system: a progress report. *ApJ* **622**, 1182–1190.

39 Rivera, E. J., Lissauer, J. J., Butler, R. P., Marcy, G. W., Vogt, S. S., Fischer, D. A., Brown, T. M., Laughlin, G., and Henry, G. W.: 2005, A 7.5 M planet orbiting the nearby star, GJ 876. *ApJ* **634**, 625–640.

40 Kley, W., DÁngelo, G., and Henning, T.: 2001, Three-dimensional simulations of a planet embedded in a protoplanetary disk. *ApJ* **547**, 457–464.

41 Kley, W.:2003, Dynamical evolution of planets in disks. *CMDA*, **87**, 85–97.

42 Nelson, R. P., and Papaloizou, J. C. B.: 2002, Possible commensurabilities among pairs of extrasolar planets. *MNRAS* **333**, L26–L30.

43 Papaloizou, J. C. B.: 2003, Disc-planet interactions: migration and resonances in extrasolar planetary systems. *CMDA* **87**, 53–83.

44 Snellgrove, M. D., Papaloizou, J. C. B., and Nelson, R. P.: 2001, On disc driven inward migration of resonantly coupled planets with application to the system around GJ876. *A& A* **374**, 1092–1099.

45 Mayor, M., Udry, S., Naef, D., Pepe, F., Queloz, D., Santos, N. C., and Burnet, M.: 2004, *A& A*, **415**, 391–402.

46 Ji, J., Kinoshita, H., Liu, L., Li, G., and Nakai, H.: 2003, The apsidal antialignment of the HD 82943 system. *CMDA* **87**, 113–120.

47 Lee, M.H.: 2004, Diversity and origin of 2:1 orbital resonances in extrasolar planetary systems. *ApJ* **611**, 517–527.

48 Ferraz-Mello, S., T.A. Michtchenko, and C. Beaugé: 2005, The orbits of the extrasolar planets HD 82943 c and b. *ApJ* **621**, 473–481.

49 Lee, M.H., Butler, R.P., Fischer, D.A., Marcy, G.W., and Vogt, S.S.: 2006, On the 2:1 orbital resonance in the HD 82943 planetary system. *ApJ*, **641**, 1178–1187.

50 Goździewski, K., and M. Konacki: 2006, Trojan Pairs in the HD 128311 and HD 82943 planetary Systems? *ApJ* **647**, 573–586.

51 de La Barre, C. M., Kaula, W. M., and Varadi, F.: 1996, A study of orbits near Saturn's triangular Lagrangian points. *Icarus* **121**, 88–113.

52 Liou, J.-C., and R. Malhotra: 1997, Depletion of the outer asteroid belt. *Science* **275**, 375–377.

53 Michtchenko, T.A., and S. Ferraz-Mello: 1997, Escape of asteroids from the Hecuba gap. *Planet. Space Sci.* **45**, 1587–1593.

54 Ferraz-Mello, S. and T.A. Michtchenko: 2002, Extra-solar Planetary Systems. *Revista Mexicana de Astronomía y Astrofísica (Serie de Conferencias)* **14**, 7–10.

55 Malhotra, R., Black, D., Eck, A., and Jackson, A.: 1992, Resonant orbital evolution in the putative planetary system of PSR1257 + 12. *Nature* **356**, 583–585.

56 Rasio, F. A., Nicholson, P. D., Shapiro, S. L., and Teukolsky, S. A. 1992. A observational test for the existence of a planetary system orbiting PSR1257 + 12. *Nature* **355**, 325–326.

8
Planets in Double Stars

Elke Pilat-Lohinger and Rudolf Dvorak

Abstract

Stability studies of planetary motion in binary systems are very important, since we expect that more and more planets in such stellar systems will be discovered in the future – due to the fact that most of the stars in the solar neighborhood form double or multiple star systems. Currently (January 2007) we know only few planetary systems, which have a “close” stellar companion, out of a – meanwhile long – list of extrasolar planetary systems. The total number of binaries with planets is about 30. In most of these double-star systems the distance of the two stellar components is between 100 and more than 12000 AU. Therefore, it is obvious that the detected planets were found to move around one stellar component.

This chapter gives an overview about planetary orbits in binary systems, where the different types of motion are discussed. First, we present some general stability studies that define the stable region for planetary motion in such systems, where different mass ratios of the stellar components were considered. In this context we discuss the influence of both eccentricities – that of the binary and that of the planet. We show that circular planetary motion cannot always give the necessary information, whether a planet is in a stable zone or not, especially when high eccentric planetary orbits are close to the border of stable motion. Some applications to real binary systems will underline the importance of such general investigations. Moreover, we show some results about the influence of the secondary companion on the region between the host star and the giant planet.

8.1
Introduction

The famous discovery of the first planet orbiting the sun-like star 51 Peg more than 10 years ago [1] was certainly a milestone in astronomy. Since that time, we have known that the planets of our solar system are not the only ones.

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

From the more than 200 extrasolar planets near sun-like stars (cf. web-site of J. Schneider)[1] that have been discovered during the last decade, we can distinguish the following systems:

1. Single-star single-planet systems.
2. Single-star multi-planet systems.
3. Planets in double-star systems.

The last two are of special interest from a dynamical point of view, since the stable planetary motion is restricted to certain regions of the phase space in those systems due to the gravitational interactions between three or more celestial bodies. Up to now we have knowledge of about 25 multi-planetary systems and about 30 binary systems [2] that host one or more planets.

In this chapter we will concentrate on planetary motion in double-star systems. Stability studies for such systems are very important, as it is well known that more than 60% of the stars in the solar neighborhood build double- or multiple-star systems (see [3]). Due to this fact, we expect that, in future, more and more planets will be found in binary systems.

However, we have to point out that more than one decade before the first planet in a binary system had been discovered, astronomers, working in Dynamical Astronomy, carried out stability studies for the different types of motion (see, e.g., [4–9]) using the elliptic restricted three-body problem (see Section 8.2) for numerical simulations. Between 1988 and 1998 Benest studied several binaries in a series of papers (see [10–15]). The discovery of planets in such systems encouraged many research groups either to examine special double star systems (see, e.g., [16–20]) or to study the stability in general again (see, e.g., [21–23]). Additionally, there are investigations, that used the general three-body problem (see e.g. [24–26] and more recently by Innanen et al. [27] or Musielak et al. [28]).

This chapter gives a summary of planetary motion in binary systems. After a short introduction to the models and methods that were used for the different studies, we first concentrate on general stability studies that define the stable region for such systems with different mass ratios of the stellar components. Moreover, we show the importance of taking the planet's eccentricity into account, which normally reduces the stable zone. The significance of such stability studies can be seen from the application to real systems, that produces immediate information for the observer, whether or not a detected planet is in the stable region. And finally, we visualize the influence of the secondary companion.

---

[1] http://vo.obspm.fr/exoplanetes/encyclo/catalog.php

8.2
Computations and Methods

For computer simulations of dynamical systems one has to choose an appropriate dynamical model, that will give a good approximation for the system under consideration. In a first step we will apply the simplest model which is most probably the fastest one for the computations. This is, in our case, the *elliptic restricted three-body problem (ER3BP)*: where the motion of a mass-less body (m_3) is studied in the gravitational field of two massive bodies, i.e., the so-called primaries (m_1 and m_2). Since m_3 does not influence the motion of the primaries (i.e., in our case the two stars), they move on Keplerian orbits around their center of mass. This model is commonly used in studies of celestial mechanics and gives quite reasonable results if the mass of the third body is small compared to the other two.

If the third body is quite massive it is advisable to use the *three-body problem (TBP)*, where we are faced with gravitational interactions between all three bodies.

For binaries, where a giant planet has already been detected, one has to use the following models instead of the ER3BP and the TBP, namely:

- the *restricted four-body problem (R4BP)*: where the motion of a mass-less body is studied in the gravitational field of three massive bodies; and
- the *four-body problem*: where all bodies are massive, so that they interact gravitationally. Its application is also used to verify the results of the R4BP.

For the computations we used two integration methods which are known to work correctly in the case of close encounters, namely the *Bulirsch–Stoer method* [29] – which applies the Richardson extrapolation, where a large interval H is divided step by step into n finer subintervals $h = H/n$ – and the *Lie integration method* [30, 31] – where a special linear differential operator (called the Lie operator) produces a Lie series which is used to solve the equations of motion; the adaptive step-size control of the program allows for the computation of close encounters. With both methods one can obtain high-accuracy solutions for ordinary differential equations.

To determine the state of motion of the computed orbits, one has to use a chaos indicator, on the one hand, or long-term orbital computations and analysis on the other hand.

As a chaos indicator the *fast Lyapunov indicator (FLI)* (see [32]) was used, which is quite a fast tool to distinguish between regular and chaotic motion. According to the definition – which is the length of the largest tangent vector:

$$\psi(t) = \sup_i \|v_i(t)\| \qquad i = 1, \dots n$$

(n denotes the dimension of the phase space) – it is obvious that chaotic orbits can be found very quickly because of the exponential growth of this vector in the chaotic region. For most chaotic orbits only a small number of primary revolutions is needed to determine the orbital behavior. In order to distinguish between stable and chaotic motion we define a critical value for the FLIs, which is usually set between 10^7 and 10^9 depending on the computation time. This method has often been applied to studies of Extrasolar Planetary Systems (see, e.g., [18–20, 22, 23, 33–36] and others).

A simple method used to analyze the orbital parameters of an extensive computation is, e.g., the *sup method*, where the variation of the sup of an action variable is examined. This method was introduced by Laskar [37] and later by Froeschlé and Lega [38]. For the planetary motion it is useful to apply this method to the eccentricity, since the variation of sup-e determines the orbital behavior quite well – as was shown by Laskar in his long-term studies of our solar system (see [39]).

Another good characterization of a region in a planetary system is achieved by the *max-e*, which is the maximum eccentricity of an orbit calculated over the whole integration time. The max-e stability maps are very useful – especially for studies in the so-called habitable zone of a sun-like star (see Dvorak, this volume) – in a sense that one can easily determine the stable regions. This method has been successfully applied to studies of extrasolar planetary systems (see, e.g., [18–20, 36, 40–42] and others.)

8.3
General Stability Studies of Planetary Motion in Double-star Systems

From a dynamical point of view, we can distinguish three types of motion in double-star systems (cf. [4]):

1. *The satellite-type (or S-type) motion*, where the planet moves around one stellar component.
2. *The planet–type (or P-type) motion*, where the planet surrounds both stars in a very distant orbit.
3. *The libration-type (or L-type) motion*, where the planet moves in the same orbit as the secondary companion but 60° ahead or behind, and furthermore, they are locked in 1:1 mean-motion resonance.

In the following sections, we discuss the characteristics of these motions.

8.3.1
S-type Motion

Most of the general stability studies of S-type motion (also called circumstellar motion) used the planar ER3BP and determined the stable region as a function of the binary's eccentricity, where the motion of the planet was initially circular (see, e.g., [6] or [21]). Only the numerical investigation by Pilat-Lohinger and Dvorak [22] also analyzed the influence of the planet's eccentricity. In principle, these three cited investigations determined the stable regions of planetary motion in a similar way. The host star, which the planet (i.e., the massless body) moves around, is always denoted m_1, and the following conditions apply:

1. The initial conditions of the binaries are:
   (a) a fixed semi-major axis of 1 AU,
   (b) variation of the eccentricity between 0 and 0.9 with a step-size of 0.1, and
   (c) two starting positions for the second star m_2 at the pericenter and the apocenter.
2. The initial conditions of the planets are:
   (a) a semi-major axis between 0.1 and 0.9 AU with a various step δa, and
   (b) four starting positions were used for each orbit (i.e., mean anomaly $= 0°$ or $90°$ or $180°$ or $270°$),
   (c) the initial eccentricity was 0 in [6] and [21] and was varied between 0 and 0.5 with a step of 0.1 for all mass ratios and in some cases up to 0.9 (in [22]),
   (d) all three investigations studied planar planetary motion.

While in [6] and [21] they determined the stable zone through escape orbits within a certain computation time, in [22] a chaos indicator was used to characterize the long-term behavior of motion. Even if the computation time of the latter seems to be quite short, one will have to take into account that the results are valid for a longer timespan due to the application of the FLI. Moreover, test computations using a rougher grid for the FLI maps and computations times over $10^4, 10^5$ and 10^6 periods of the primaries did not change the result significantly.

As an example we present the results for a binary, where the two stellar components have equal masses, i.e., the mass ratio $\mu = 0.5$. The 3D plot (Fig. 8.1) shows the initial eccentricities of the binary and of the planet on the x- and y-axis, respectively. The z-axis gives the extension of the stable zone which is defined by the semi-major axis (in AU) of the last stable orbit, i.e., the largest distance between the planet and m_1, where the orbits of all four starting positions remained stable for the integration time. The results can be summarized as follows:

- One can clearly see a stronger reduction of the stable zone along the x-axis, i.e., if the eccentricity of the binary is increased from 0 to 0.8.

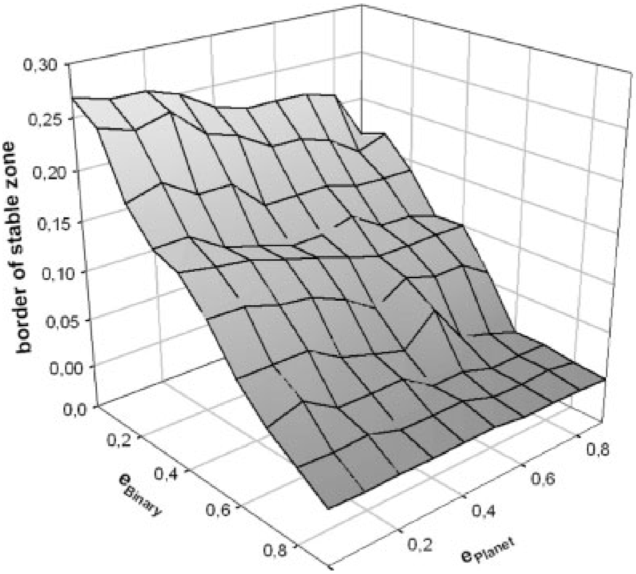


Fig. 8.1 The size of the stable zone (z-axis) of S-type motion in a binary with mass ratio $\mu = 0.5$ depending on the eccentricity of the binary (x-axis) and of the planet (y-axis). It is clearly seen that the variation in e_{binary} influences the extension of the stable zone more strongly than the variation in e_{planet}.

- The appearance of minima and maxima, as at $(e_{binary}, e_{planet}) = (0.1, 0.2)$ indicates, that an increase in the eccentricity does not reduce the stable zone immediately.
- The influence of the binary's eccentricity on the size of the stable zone can be seen when we follow a line along the x-axis (i.e., e_{Binary}) from 0 to 0.8 for a certain eccentricity of the planet. The reduction of the stable zone is between 0.20 and 0.24 [units of length (UL)] for the different eccentricities of the planet.
- When we follow a line along the y-axis (i.e., e_{Planet}) from 0 to 0.8 then the stable zone is reduced between 0.025 and 0.09 [UL].
- Consequently, the influence of e_{binary} is between 2.5 and nearly 8 times stronger than that of e_{planet}.

To show the necessity of taking the planet's eccentricity into account, to determine whether a detected planet in a binary is in the stable zone or not, a comparison of the results of [21] and [22] for $\mu = 0.5$ was carried out. In this connection, the relation published in [21], that defines a zone for the stability borders of a certain mass ratio and a certain eccentricity of the binary is

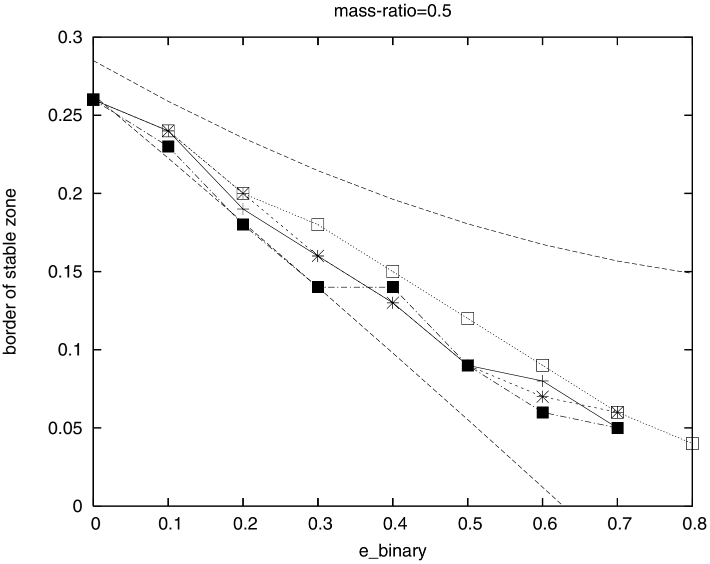


Fig. 8.2 Stability borders for planetary motion in a binary with equal masses. The region between the two dashed lines (which were calculated from Eq. 3 of [21]) shows where one can expect the border of stable motion for planetary orbits in such a double-star system. The border lines for $e_{planet} = 0$: full line with crosses; $e_{planet} = 0.1$: dashed line with stars; $e_{planet} = 0.2$: dashed-dotted line with black squares are taken from [22]. The dotted line with white squares shows the result of the computations in [21].

applied. Since in [21] they did the computation only for circular planetary motion, we verify how the results for eccentric planetary motion obtained in [22] fit the region defined by this relation, which is marked by the two dashed lines in Figs. 8.2 and 8.3, that show the comparison.

In the first figure, only results up to $e_{planet} = 0.2$ are shown: for circular orbits see the full line with crosses, for $e_{planet} = 0.1$ the dashed line with stars, and for $e_{planet} = 0.2$ the dashed-dotted line with black squares. Furthermore, the dotted line with white squares shows the result obtained in [21]. One can see clearly, that all calculated stability borders lie inside the limiting curves. For $e_{planet} = 0.2$, there are two cases ($e_{binary} = 0.2$ and 0.3) where the calculated border is at the lower limit. Moreover, the results in [22] for the circular case and for $e_{binary} = 0.1$ are similar, except for $0.4 \leq e_{binary} \leq 0.6$. Figure 8.3 summarizes the results for all eccentricities of the planets, where the majority of the stability borders lie outside the given zone defined by the relation of Holman and Wiegert. This means that, for eccentric planetary orbits ($e_{planet} \geq 0.2$) we have to shift the border of the stable zone towards the host star (m_1) and in case of $e_{binary} = 0$, from 0.26 to 0.22 [UL].

A summary of the results is given in Tables 8.1 and 8.2. In Table 8.1 one can see the border of the stable region (i.e., the semi-major axis of the last stable

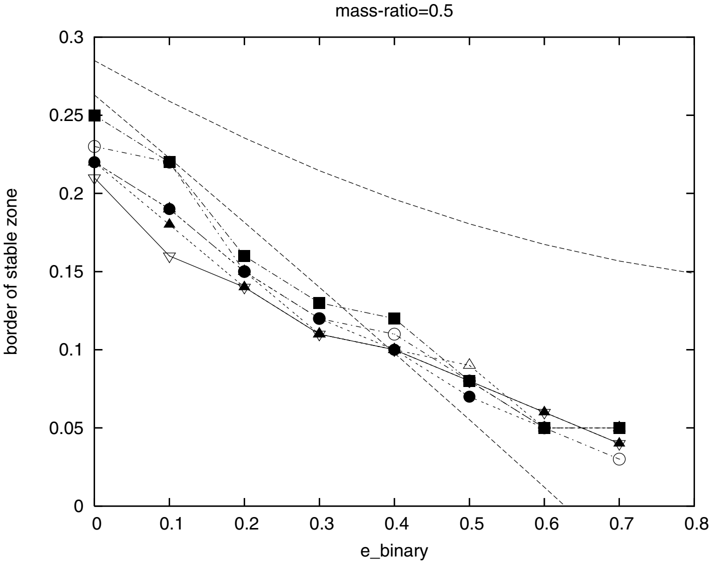


Fig. 8.3 The same as Fig. 8.2 but for all eccentricities of the planet from 0.3 to 0.9. The results are taken from [22]. One has to note that all border-lines lie outside if $e_{binary} < 0.4$.

orbit) where we took the lower value of the two studies [21] and [22]. Even if the results of these studies are in good agreement, minor variations will be caused by the different methods used to determine the stable region. In some cases the FLI results gave a slightly larger stable region due to the fact that only four starting positions were used, whereas Holman and Wiegert used eight. Furthermore, the variation in the stable region due to an increase of e_{binary} or e_{planet} is shown in Table 8.2 for all mass ratios.

In principle, one can see that the reduction in the stable zone, due to an increase in the binary's eccentricity, is between 0.07 [UL] (i.e., for the initially circular motion in a binary with $\mu = 0.9$) and 0.28 [UL] (i.e., for the initially circular motion in a binary with $\mu = 0.1$). Even if the dependence of the size of the stable region on the eccentricity of the planet is not so strong, it is not negligible, especially if a planet is close to the border of chaotic motion and moves in a highly eccentric orbit.

A presentation of the 3D stability plots of all mass ratios and a detailed discussion of this work is given in [22].

Table 8.1 Stable zone (in units of length) of S-type motion for all computed mass ratios and eccentricities of the binary. The given size for each μ, e_{binary} pair is the lower value of the studies in [21] and [22].

	Mass ratio (μ)								
e_{binary}	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	0.45	0.38	0.37	0.30	0.26	0.23	0.20	0.16	0.13
0.1	0.37	0.32	0.29	0.27	0.24	0.20	0.18	0.15	0.11
0.2	0.32	0.27	0.25	0.22	0.19	0.18	0.16	0.13	0.10
0.3	0.28	0.24	0.21	0.18	0.16	0.15	0.13	0.11	0.09
0.4	0.21	0.20	0.18	0.16	0.15	0.12	0.11	0.10	0.07
0.5	0.17	0.16	0.13	0.12	0.12	0.09	0.09	0.07	0.06
0.6	0.13	0.12	0.11	0.10	0.08	0.08	0.07	0.06	0.045
0.7	0.09	0.08	0.07	0.07	0.05	0.05	0.05	0.045	0.035
0.8	0.05	0.05	0.04	0.04	0.03	0.035	0.03	0.025	0.02

Table 8.2 Stable zone (in units of length) of S-type motion.

Mass ratio $\mu = m_2/(m_1 + m_2)$	e_{binary}	**Stable zone**	
		$e_{planet} = 0$	$e_{planet} = 0.5$
0.1	0	0.45	0.36
	0.5	0.17	0.13
0.2	0	0.38	0.31
	0.5	0.16	0.12
0.3	0	0.37	0.28
	0.5	0.13	0.11
0.4	0	0.30	0.25
	0.5	0.12	0.07
0.5	0	0.26	0.22
	0.5	0.12	0.07
0.6	0	0.23	0.21
	0.5	0.09	0.07
0.7	0	0.20	0.18
	0.5	0.09	0.07
0.8	0	0.16	0.16
	0.5	0.07	0.05
0.9	0	0.13	0.12
	0.5	0.06	0.04

In one of the following sections we show the results for $\mu = 0.2$ (see Fig. 8.8) as an application to the real binary γ Cephei that hosts a giant planet of 1.76 Jupiter-masses .

8.3.2
P-type Motion

According to previous results – and contrary to the S-type orbits – the stability limits for P-types are almost independent of the mass ratio (see Table 8.3) of the primaries. As a consequence we present the case of equal massive primary bodies only; i.e., with mass ratio $\mu = 0.5$.

Table 8.3 The inner stability border (in units of length) of P-type motion for various mass ratios and eccentricities of the binary. (Taken from [21].)

	Mass ratio (μ)				
e_{binary}	0.1	0.2	0.3	0.4	0.5
0.0	2.0	2.2	2.3	2.3	2.3
0.1	2.4	2.7	2.7	2.8	2.8
0.2	2.7	3.1	3.1	3.1	3.1
0.3	3.1	3.5	3.5	3.3	3.2
0.4	3.5	3.5	3.6	3.5	3.6
0.5	3.8	3.9	3.9	3.6	3.7
0.6	3.9	3.9	3.9	3.8	3.7
0.7	4.2	4.3	4.3	4.1	4.1

For such a system several studies of periodic orbits exist in the circular restricted three-body problem. The first investigations for the three-dimensional case date back to Goudas [43] (1961, 1963), followed by the works of Markellos [44, 45] (1977, 1978), Michalodimitrakis [46] (1979), Perdios and Markellos [47] (1988) and Zagouras and Markellos [48] (1977) and more recently Broucke [49] (2001).

In the study by Pilat-Lohinger et al. [23] the phase space around such a binary was examined with the following variation in the initial conditions:

- For the orbits of the binary:
  1. The eccentricity e_{binary} was increased from 0 to 0.5 with $\Delta e_{binary} = 0.05$.
  2. For the starting positions of the primaries we used the apo- and periastron position (for $e_{binary} > 0$).
- For the planetary orbits:
  1. The initial distance (**A**) of the planet from the barycenter of the binary was varied from 1.8 to 3.5 AU (for $e_{binary} \leq 0.15$) and to 4.5 AU (for $e_{binary} \geq 0.2$) with $\Delta \mathbf{A} = 0.01$ AU.
  2. For each **A** we used four starting positions for the planets which are defined by the following angles of the planetary orbit to the connecting line of the primaries: $\alpha = 0^\circ, 45^\circ, 90^\circ$ and 135°;

3. The inclination i_{planet} of the planetary orbit to the plane of motion of the binary was increased from 0° to 50° with a step size of $\Delta i_{planet} = 2.5°$.
4. All fictitious planets were started in circular orbits.

The *stable motion* was determined by means of the FLIs, where the dependency on two parameters, namely on the eccentricity of the binary and on the inclination of the planet was studied. From all computed orbits one can define the following borders:

- An *upper critical orbit* (UCO) [7], which builds the border of the stable zone and is therefore defined as the innermost orbit where all eight (or four if $e_{binary} = 0$) different starting positions were found to be stable.
- A *lower critical orbit* (LCO) which builds the border of the chaotic zone, that is the outermost orbit where all eight (or four if $e_{binary} = 0$) different starting positions were found to be unstable.
- An escape time for the unstable orbits, which is given by the time when a close encounter with one of the primaries occurred (within the Hill's sphere).

Three examples of the 22 computed stability maps are given in Figs. 8.4 and 8.5 (a) and (b), where the FLI results are shown for (1) circular binary motion (Fig. 8.4), and (2) two eccentric binary motions, i.e., $e_{binary} = 0.1$ (Fig. 8.5a) and $e_{binary} = 0.5$ (Fig. 8.5b). All maps show three types of motion: black for the stable, grey for the mixed and light grey for the chaotic one. In the plot of the circular case (Fig. 8.4) we see a nearly constant UCO (at $\mathbf{A} = 2.25$ AU) for inclinations up to 20°, and subsequently we observe an almost linear increase to smaller values of the initial distance down to $\mathbf{A} = 2.13$ AU for $i_{planet} = 50°$. For the LCO there is no significant change with the inclination, which can be seen by the almost-straight line at $\mathbf{A} = 1.98$ AU. The region between the LCO and UCO is the so-called mixed zone (where both motions can be found for a certain **A**), whose size is 0.3 AU for all inclinations from 0° to 15°. Thereafter, we recognize a separation of this zone due to a finger-like stable area from $i_{planet} = 15°$ to $i_{planet} = 45°$. A second finger-like separation due to another stable region, occurs at inclinations between 25° and 35° which finally results in a large stable region close to the unstable zone from $i_{planet} \geq 40°$ on. Furthermore, for $i_{planet} = 47.5°$ and $i_{planet} = 50°$ a small completely chaotic region appears at $\mathbf{A} = 2.1$ AU so that the mixed region disappears almost completely for such large values of the initial inclinations. Even if the result shows an increase in the stable zone for higher inclinations, this is not in contrast to the study by Kozai [50], who found more instability for highly inclined orbits in the asteroid belt, which is caused by a resonance that pumps up the initial small eccentricities to large ones. As all orbits were started in circular

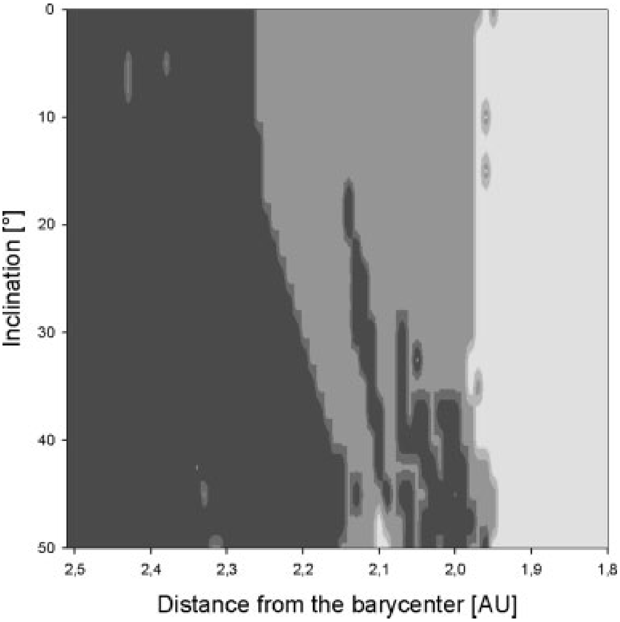
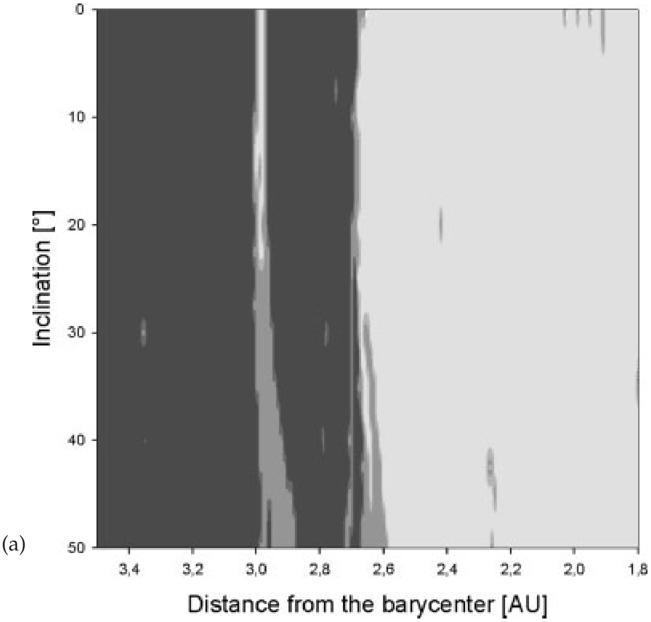


Fig. 8.4 FLI stability map for circular motion of the binary ($e_{binary} = 0$) – for details see the text. The different distances to the barycenter are plotted on the x-axis and the different inclinations denote the y-axis. Note that the scaling of both axes is from the largest to the lowest value.

motion, the Kozai resonance cannot be observed in the results. Moreover, it was already found, for the outer solar system, that the Kozai resonance plays a role only for orbits with high inclination and large eccentricity (see Thomas and Morbidelli [51]).

Additionally, some small chaotic islands appear inside the stable region. They may be caused by mean-motion resonances between the primaries and the planet, but that still has to be studied more closely.

For the elliptic study, ten different eccentricities for the binary's motion were used: $0.05 \leq e_{binary} \leq 0.5$ (with $\Delta e_{binary} = 0.05$); where eight initial positions for each orbit were computed (as the primaries were started in two positions – the peri- and apoastron). As an example we show the FLI results for a low eccentricity (0.1) and a high eccentricity (0.5) motion of the binary (see Figs. 8.5a,b). One can see that the border between chaotic and mixed motion (i.e., the LCO) is shifted from $\mathbf{A} = 1.95$ AU (for $e_{binary} = 0$ (see Fig. 8.4) to 2.6 or 2.67 AU depending on the inclination for $e_{binary} = 0.1$ (see Fig. 8.5a). Thereafter the limit remained at this position up to $e_{binary} = 0.3$. For an eccentricity of $e_{binary} = 0.35$ another shift can be observed to about $\mathbf{A} = 3$ AU, which corresponds to the position of the chaotic strip in the stable region for



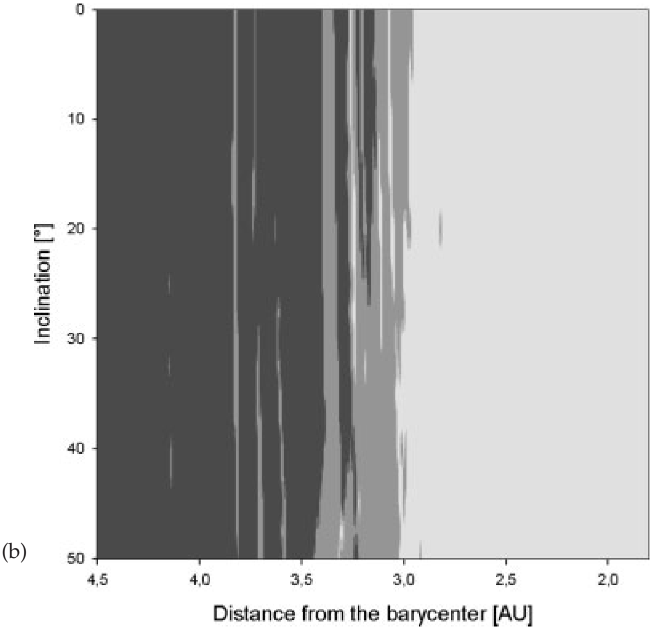


Fig. 8.5 FLI stability maps for different eccentricities of the binary: (a) $e_{binary} = 0.1$ and (b) $e_{binary} = 0.5$; for details see the text. The different distances to the barycenter are plotted on the x-axis and the different inclinations are denoted by the y-axis. Note that the scaling of both axes is from the largest to the lowest value.

$e_{binary} = 0.1$ (see Fig. 8.5a). A further increase in the eccentricity causes the stable zone between the LCO and UCO to become smaller and smaller, which finally leads to a shift of the two borders. Moreover, another chaotic island appears at about 3.4 AU from the barycenter – first indications for this can already be seen in Fig. 8.5a for $i_{planet} = 30°$, while in the results of the orbital computations, where we defined the different state of motion by means of the escape time of an orbit, this chaotic island appears for the first time for $e_{binary} = 0.2$. More details can be found in [23].

The second example for the elliptic case ($e = 0.5$) shows the border between unstable and mixed motion at about 3 AU. The border of the stable zone begins beyond 3.8 AU from the barycenter, which is about 0.3 AU further outside than in the result of the orbital computations – except for high inclinations ($i_{planet} \geq 47.5°$), where the UCO is the same in both studies. So we can summarize that:

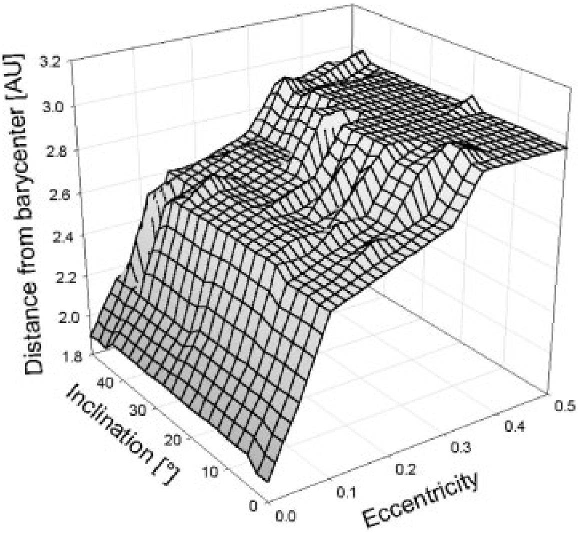
(1) the stability limit varies between 2.15 AU and 3.85 AU depending on the eccentricity of the binary;

(2) only for some eccentricities do we find a clear dependence of the stability limits on the inclination of the planetary orbit; and

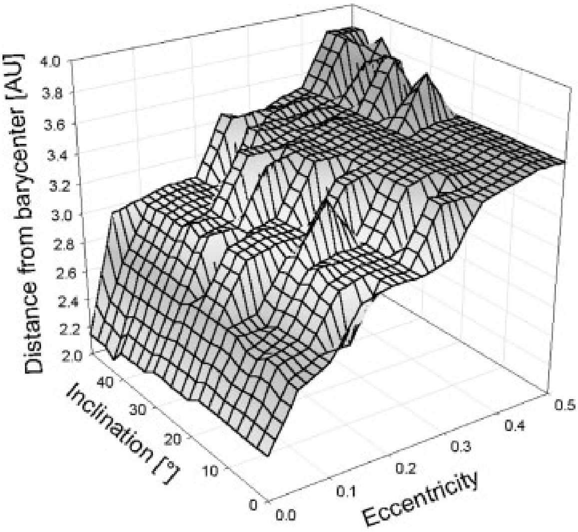
(3) the regions of stable, mixed and chaotic orbits show the following behavior: for certain eccentricities a small finger-like unstable island appears at high inclinations and evolves to a small strip of instability in the $(\mathbf{A}, i_{planet})$ plots if we increase the binaries' eccentricity. This development finally leads to a shift of the stable zone to larger distances from the barycenter.

In Fig. 8.6 we summarize the results for the LCO and UCO. The global result of the LCO (Fig. 8.6a) shows an increase for an eccentricity up to 0.1 with a more or less constant behavior with respect to the inclination (from $\mathbf{A} = 1.9$ to 2.5 AU). Between $e_{binary} = 0.1$ and 0.3 we find a slightly inclined plateau to larger $\mathbf{A}$ values (from 2.5 to 2.75 AU). A further increase of e_{binary} up to 0.35 yields a jump in $\mathbf{A}$ and subsequently $\mathbf{A}$ depends neither on e_{binary} nor on i_P ($\mathbf{A} = 2.95$ AU).

Figure 8.6b shows the global result of the UCO in the (e_{binary}, i_{planet}) plane, which is the border of the stable zone and therefore defined by the distance to the barycenter where all starting positions indicate stable motion. One can see that an increase of the binary's eccentricity shifts the stable zone to larger distances from the barycenter while the inclination of the planet has, in general, no significant influence on the UCO; the most visible exception is for $e_{binary} = 0.5$ and high inclinations. A more detailed study can be found in [23].



(a)



(b)

Fig. 8.6 The size of the stable zone of P-type motion in a binary with mass ratio $\mu = 0.5$ depending on the eccentricity of the binary and on the inclination of the planet. (a) summarizes the results for the LCO and (b) for the UCO.

8.3.3
L-type Motion

The third type of motion – according to Dvorak [4] – where the planet librates around one of the two Lagrangian triangular points of one of the stars, is not so interesting for planetary motion in double stars due to a limitation in the

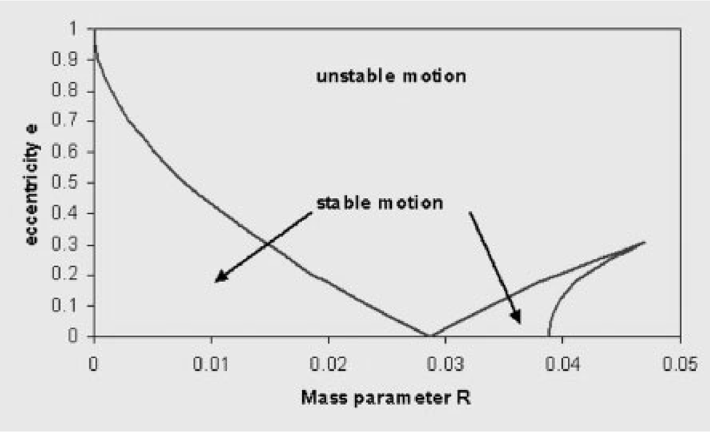


Fig. 8.7 Stable and unstable region of L-type motion in the ER3BP for different mass ratio R (see x-axis) and different eccentricities (y-axis) (from Marchal [52]).

mass ratio of the two stars:

$$\mu = m_2/(m_1 + m_2) < 1/26$$

Figure 8.7 shows quite clearly the restriction of the stable region to certain mass ratios. Even if an eccentricity of the binary between 0.1 and 0.3 extends the range for the mass ratio to nearly 0.05, this motion is more interesting for single-star – giant planet – systems, which fulfill the condition of a certain mass ratio without problems. Therefore, we do not show details of the L-type motion in this chapter, but refer the reader to the chapter on terrestrial planets (see Chapter 2), since this motion allows for planetary systems with a giant planet moving in the HZ, a special configuration for the existence of additional terrestrial-like planets in the HZ.

8.4
Application to Real Systems

As already mentioned above, these general stability studies can easily be applied to real binary systems that host a planet. In this context, γ Cephei is one of the most interesting double-star systems, where a planet was found (see Cochran et al. [53]). This binary is about 11 pc away from our solar system and consists of a K1 IV star (of 1.6 solar masses), around which the detected planet moves, and a M4 V star (of 0.4 solar masses). Thus the mass ratio (μ) of this system is 0.2. An overview of the size of the stable region for this mass ratio is given in Fig. 8.8, that shows the respective extension of the stable zone

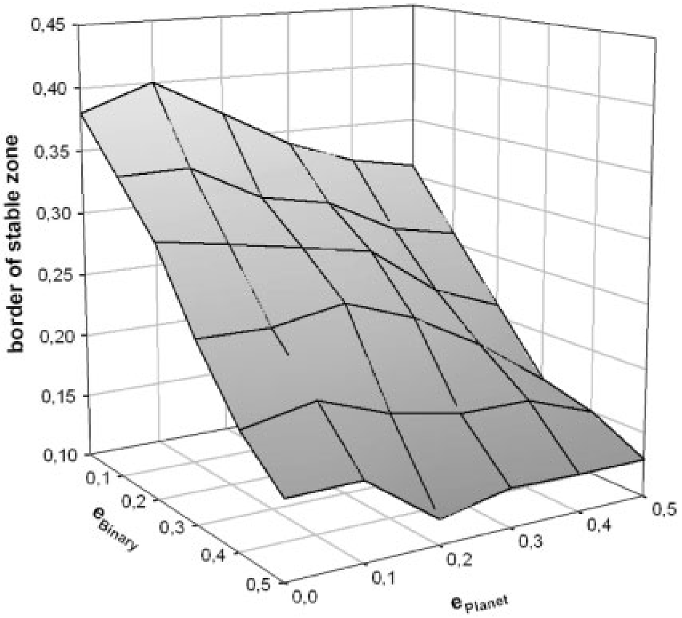


Fig. 8.8 The size of the stable zone of S-type motion in a binary with mass ratio $\mu = 0.2$ depending on the eccentricity of the binary (x-axis) and of the planet (y-axis). It is clearly seen that the variation of e_{binary} influences the extension of the stable zone stronger than the variation of e_{planet}.

(z-axis) for each (e_{binary}, e_{planet}) pair on the (x,y) plane, which is defined by the semi-major axis of the last stable orbit (corresponding to the largest distance of the planet to its host star, where all computed orbits remained stable for 1000 time units[2]). Taking the orbital parameters: the semi-major axis of the binary (a_{binary}) and the eccentricities of the binary (e_{binary}) and of the planet (e_{planet}), we obtain for the set of old orbital parameters ($a_{binary} \sim 22$ AU, $e_{binary} = 0.44$ and $e_{planet} = 0.209$) an extension of the stable zone up to 3.6 AU and for the new set of parameters ($a_{binary} \sim 23$ AU, $e_{binary} = 0.36$ and $e_{planet} = 0.12$) up to nearly 4 AU. As the work by [21] is often used to confirm the stability of a detected planet in a binary, we show in Fig. 8.9b that one has to be careful, especially in the case of eccentric motion of the planet, since the study in [21] gives a larger stable zone due to the limitation of circular planetary motion. This is clearly visible in Fig. 8.9b, where the two dashed lines show the zone for the stability borders defined by the relation given in [21]. The result for each eccentricity of a planet (from 0 to 0.8) is given by a line with a certain

---

2) Test-computations for $\mu = 0.3, 0.5$ and 0.7 up to 100 000 time units showed the same qualitative results.

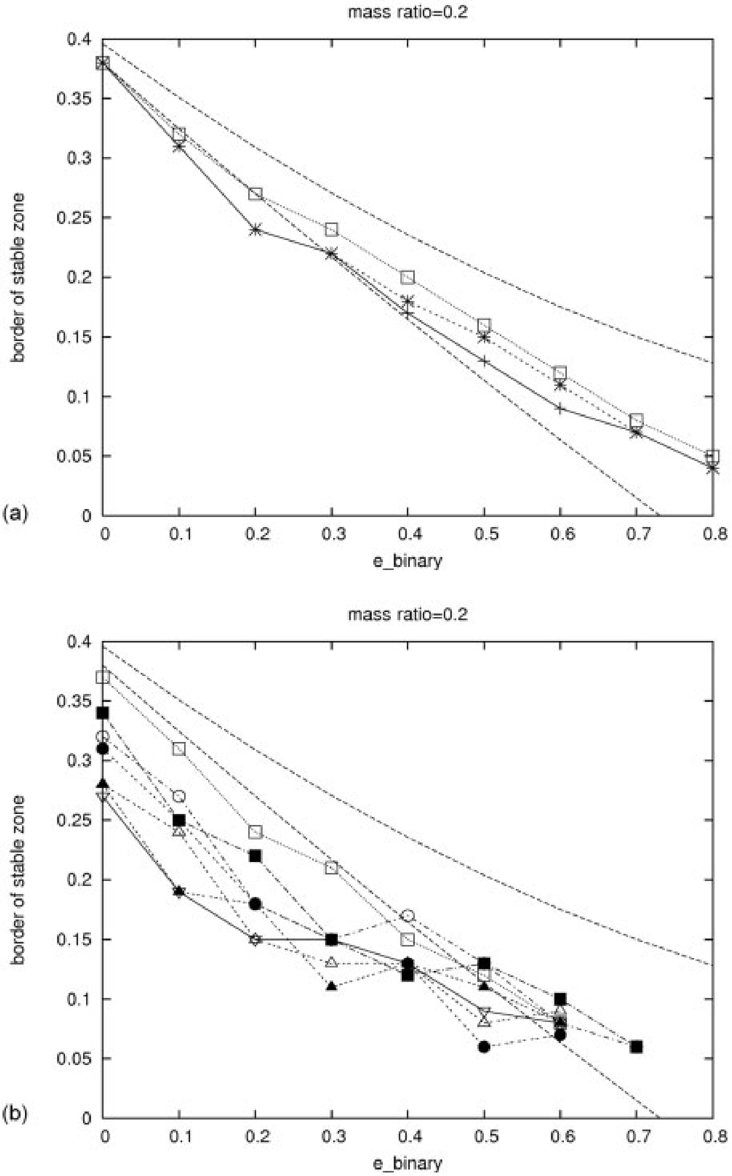


Fig. 8.9 A comparison of our results with those of [21]. The area between the two dashed lines defines the zone for the stability border according to the relation given in [21]. (a) shows the results for $e_{planet} = 0$ (full line with crosses) and $e_{planet} = 0.1$ (dashed line with stars) of [22] and those of [21] (dotted line with white squares). (b) shows the results for e_{planet} (from 0.3 to 0.9) in comparison with the theoretical zone for the borderline of stability.

symbol. It is clearly seen that the majority of the stability borders are outside the "border zone" defined by Holman and Wiegert only for high eccentric binary motion ($e_{binary} \geq 0.6$) all border lines are inside the dashed lines, since this zone is quite large. Fig. 8.9a shows the results for circular orbits (full line with crosses) and for $e_{planet} = 0.1$ (dashed line with stars). Furthermore, the dotted line with white squares shows the result obtained in [21]. It is clearly seen that the results for the circular problem ($e_{binary} = 0$) are the same for the three cases; but for the elliptic problem (e_{binary} from 0.1 to 0.8) the new determined stability borders are closer to m_1. Moreover, the results for $e_{planet} = 0$ and 0.1 are the same up to $e_{binary} = 0.3$ and again for $e_{binary} \geq 0.7$; and two cases ($e_{binary} = 0.1$ and 0.2) are outside the zone determined using Eq. (3) of [21].

Therefore, it is advisable to verify the stability by numerical computations especially if the detected planet is quite close to the border of the stable zone.

The application to the other two "close" binary systems, Gliese 86 and HD41004 AB clearly shows the stability for the detected planet Gl86 Ab, since it is a close-in planet with a semi-major axis of 0.11 AU. For the detected planet of HD41004 A we have three parameter sets, where the position of the planet varies between 1.31 and 1.7 AU and the eccentricity of the planet seems to be quite high (between 0.39 and 0.74). Since we have no indication of the binary's eccentricity, we have determined the stable zone for the different parameter sets as a function of e_{binary}. Depending on the planet's eccentricity, e_{binary} has to be < 0.6 in all cases, and even < 0.15 when $e_{planet} = 0.74$ (see Fig. 8.10, otherwise the detected planet would not be in the stable region. A detailed study of this system can be found in [20].

8.5
Influence of the Secondary

Detailed studies of two close binary systems, γ Cephei and HD41004 AB, where the influence of the secondary star was examined for the region[3] between the host star and the detected planet, are shown in Fig. 8.11(a–d). A comparison of the two results indicates significant differences. In the system γ Cephei, the presence of the perturbing star (see Fig. 8.11a) decreases the stable region (i.e., the faint region in the panels) and shows an arc-like chaotic path with a stable island around 1 AU (which corresponds to the 3:1 mean motion resonance (MMR)). The same study for HD41004 AB does not show a significant difference between the two models for eccentricities of the binary < 0.3 (compare (c) and (d) in Fig. 8.11).

[3] We must point out that this region corresponds to the so-called habitable zone (see Kaltenegger and Friedlund or Dvorak in this volume) in the case of HD41004AB

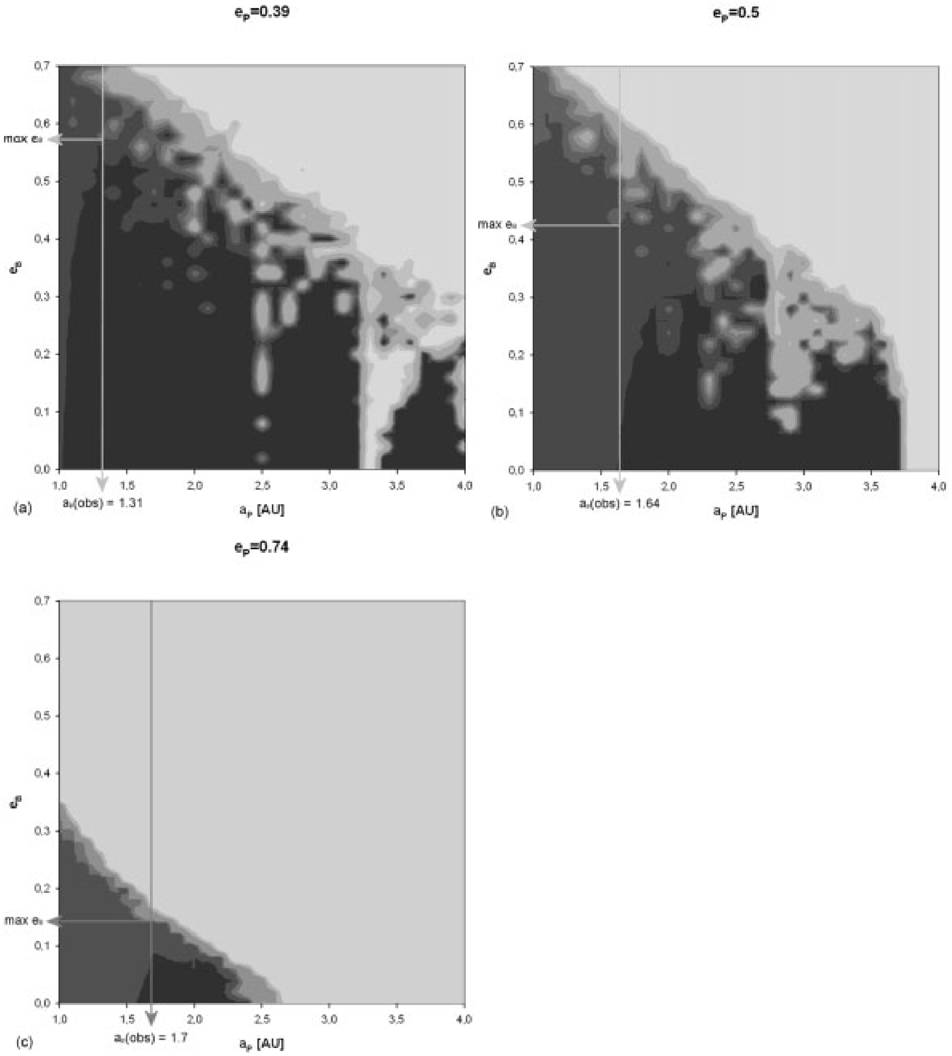


Fig. 8.10 FLI-stability maps for the different parameter sets, and a fixed value for the eccentricity of the detected giant planet: (a) the old HD41004 A system with $e_P = 0.39$, (b) the new HD41004 A system with $e_P = 0.5$ and (c) the new HD41004 A system with $e_P = 0.74$. A variation of the planet's semi-major axis a_P (x-axis) and the binary's eccentricity e_{binary} (y-axis) allows us to determine the maximum e_{binary} for which we have found long-term stability of the detected giant planet (see the horizontal line). The vertical line in each panel labels the observed position of the detected planet. Long-term stability can be expected in the black and dark gray regions.

A first investigation about this different behavior is given in [35], where a variation in the semi-major axis of the detected giant planet in both systems shows the following result.

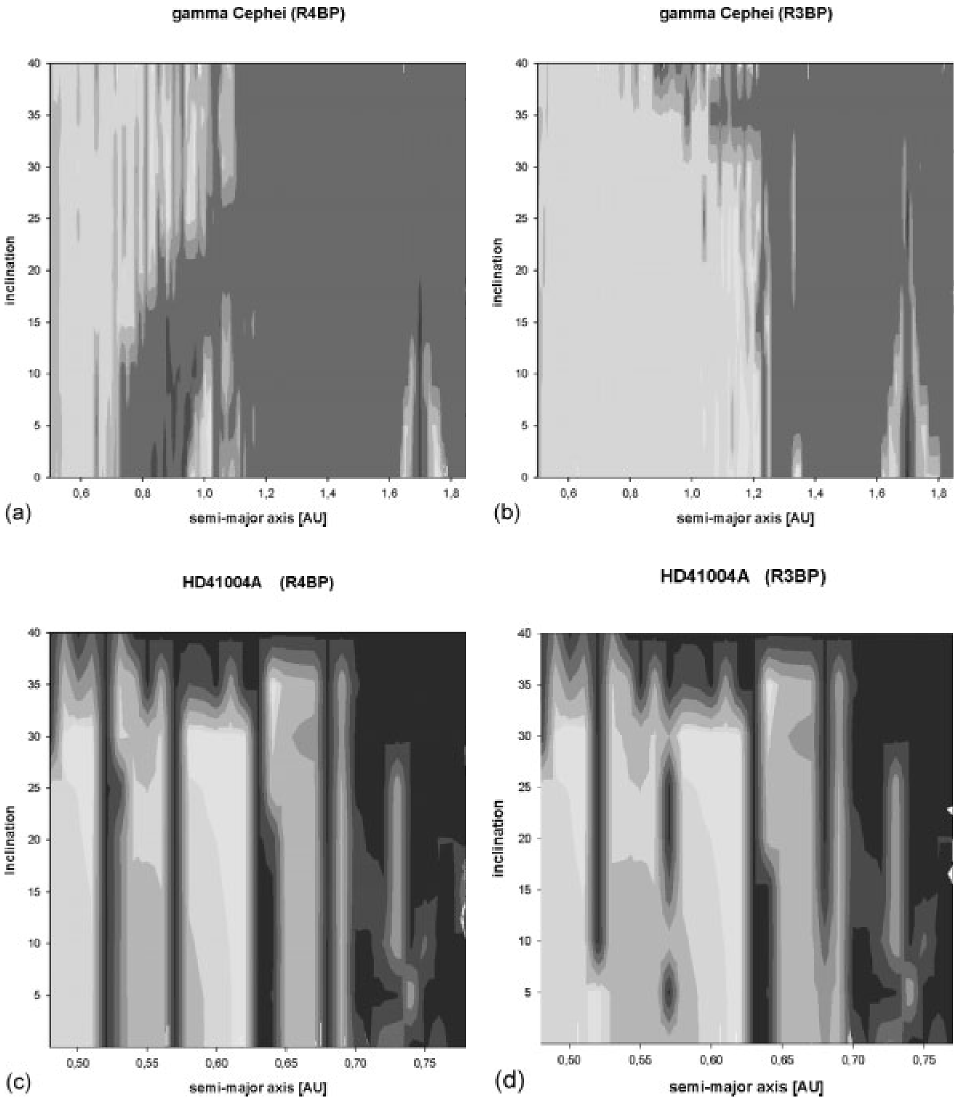


Fig. 8.11 (a) and (b): Stability maps for a fictitious planet in the vicinity of γ Cephei: (a) shows the FLI result in the restricted four-body problem (R4BP) (i.e., γ Cephei + secondary + detected planet + fictitious planet) and (b) shows the result in the restricted three-body problem (R3BP) (i.e., γ Cephei + detected planet + fictitious planet). The dark region labels the chaotic zone and the white area the stable one. (c) and (d): Stability maps (i.e., max-e maps) for a fictitious planet in the vicinity of HD41004 A: (c) shows the result in the R4BP (i.e., binary HD41004 AB + detected planet + fictitious planet) and right panel (d) shows the result in the R3BP (i.e., HD41004 A + detected planet + fictitious planet). The different grey scales indicate the zones of different eccentricities: $e < 0.2$ (white), $0.2 < e < 0.3$ (light grey), $\ldots$, $e > 0.8$ (black, i.e., unstable region).

When the giant planet is close enough to the host star (i.e., around 1.3 AU for the two binaries) the region is mainly perturbed by MMRs with respect to the giant planet. A curved chaotic structure appears if the giant planet is shifted towards the secondary star, so that a secular perturbation occurs, since the secondary star will cause a precession of the perihelion of the giant planet (as all massive bodies started in the same plane, a precession of the ascending node cannot be modeled). Figure 8.12 shows the region between 0.5 and 1.2 AU in the two binary systems, where the state of motion was determined by FLI computation. The stable zone is given by the white region and the chaotic one by the dark region. One can clearly see that an increase in the semi-major axis of the detected giant planet in the HD41004A systems to 2.1 AU (see Fig. 8.12d) leads to the same arc-like structure in the stability map as was found for γ Cephei. Moreover, when we started γ Cephei system's giant planet closer to the host star at $a_{planet} = 1.3$ AU (see Fig. 8.12a) the arc-like structure disappeared and we could observe a splitting of the stable zone into strips due to mean-motion resonances with respect to the giant planet.

Of course the corresponding dynamical maps for the two systems are not exactly alike for a certain semi-major axis, therefore, it is obvious that other parameters, like the eccentricity of the binary, the mass ratio of the binary or the mass of the giant planet and its eccentricity also have an important influence.

8.6
Detected Planets in Binary Systems

Following the observations about planets in double stars (see Table 8.4), one can clearly see that they have to move in S-type orbits due to the fact that most of these systems are wide binaries, where the distance between the two stars is more than 100 AU.

Table 8.4: The detected extrasolar planets in double stars [2].

Star	a_{binary} [AU]	a_{planet} [AU]	$M_{pl} \sin i$ [M_{Jup}]	e_{planet}
HD38529	12042	0.129	0.78	0.29
		3.68	12.7	0.36
HD40979	6394	0.811	3.32	0.23
HD222582	4746	1.35	5.11	0.76
HD147513	4451	1.26	1.00	0.52
HD213240	3909	2.03	4.5	0.45
Gl 777 A	2846	0.128	0.057	0.1
		3.92	1.502	0.36
HD89744	2456	0.89	7.99	0.67
GJ 893.2	2248	0.3	2.9	—

Table 8.4: (continued)

Star	a_{binary} [AU]	a_{planet} [AU]	$M_{pl} \sin i \, [M_{Jup}]$	e_{planet}
HD80606	1203	0.439	3.41	0.927
55 Cnc	1050	0.038	0.045	0.174
		0.115	0.784	0.02
		0.24	0.217	0.44
		5.25	3.92	0.327
GJ 81.1	1010	0.229	0.11	0.15
		3.167	0.7	0.3
16 Cyg B	860	1.66	1.69	0.67
HD142022	794	2.8	4.4	0.57
HD178911	789	0.32	6.292	0.124
Ups And	702	0.059	0.69	0.012
		0.83	1.89	0.28
		2.53	3.75	0.27
HD188015	684	1.19	1.26	0.15
HD178911	640	0.32	6.29	0.124
HD75289	621	0.046	0.42	0.054
GJ 429	515	0.119	0.122	0.05
HD196050	510	2.5	3.00	0.28
HD46375	314	0.041	0.249	0.04
HD114729	282	2.08	0.82	0.31
ϵ Ret	251	1.18	1.28	0.07
HD142	138	0.98	1.00	0.38
HD114762	132	0.3	11.02	0.25
HD195019	131	0.14	3.43	0.05
GJ 128	56	1.30	2.00	0.2
HD120136	45	0.05	4.13	0.01
γ Cep	20.3	2.03	1.59	0.2
Gl 86	21	0.11	4.01	0.046
HD41004 AB	23	1.7	2.64	0.5

8.7
Summary

In this chapter we discussed the stability of planetary motion in double-star systems, which is very important since observations have shown that most of the stars in the solar neighborhood build double or multiple-star systems. As the accompanying star will limit the planetary motion to a certain region around the host star, depending on the mass ratio of the two stars and on the eccentricity of the binary, it is important to verify the stability of a detected

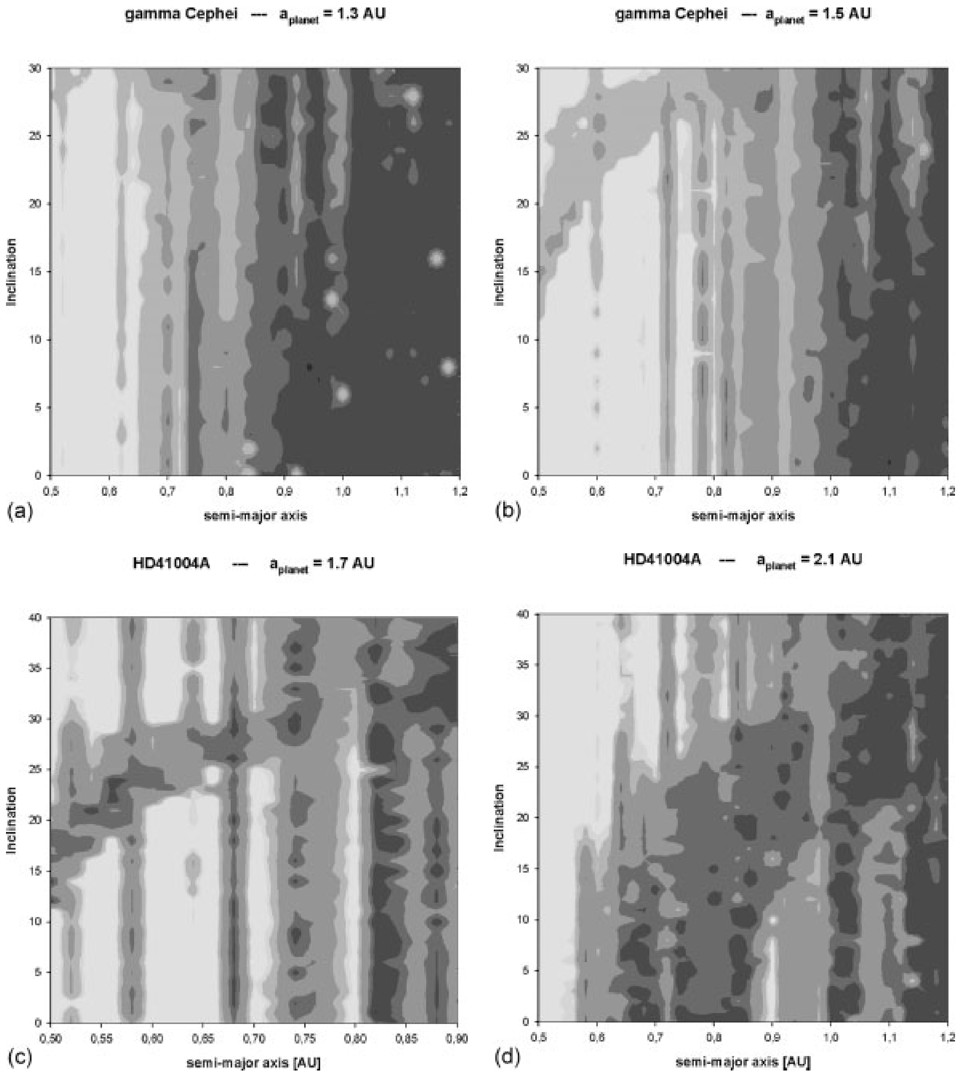


Fig. 8.12 Stability maps for γ Cephei (upper panels) and HD41004A (lower panels) where the semi-major axis of the detected giant planet was varied in the different plots. The dynamical behavior was determined by means of the FLIs, where white zones denote stable motion and the dark region corresponds to chaotic motion.

planet, especially if the two stellar components are quite close. However, the first dynamical studies on this had already been carried out more than 10 years before the discovery of the first planet near the sun-like star 51 Peg by Dvorak [4]. The detection of extrasolar planets in double-star systems – we know about 30 (see, e.g., [2]) – raised the importance of such studies, since they were no longer purely theoretical, but applicable, as was shown for γ Cephei (see Section 8.4).

From the three types of motion that can occur in binary systems, we have seen that only two are interesting, namely the S-type and P-type motions. For these motions, we discussed the sense and the results of general stability studies, where we considered mainly the works by Holman and Wiegert [21] and Pilat-Lohinger and Dvorak [22] for the S-type motion and Pilat-Lohinger et al. [23] for the P-types. All these studies can be regarded as continuations of the work by Dvorak [4, 5], Rabl and Dvorak [6] and Dvorak et al. [7]. Comparing the different studies, one finds a good agreement in the results.

For *S-type motion* – i.e., when the planet moves around one stellar component, we stressed the influence of the planet's eccentricity on the extension of the stable zone, as was studied in [22] for the first time. Even if this study showed that the reduction in the stable zone is stronger when the binary's eccentricity is increased, namely between 2.5 and nearly 8 times, because:

- the increase in e_{binary} from 0 to 0.8 reduces the stable zone from between 0.20 to 0.24 [units of length] in the different models, while
- the same increase in e_{planet} leads to a reduction of between 0.025 and 0.09 [units of length].

This should not be overlooked, especially when a detected planet is moving in a high eccentric orbit close to the border of stable motion. The application to the real binary γ Cephei emphasizes this.

Moreover, we gave an impression about the influence of the perturbing star, which induces secular perturbations depending on the distance from the giant planet. And we recognized that the binary's eccentricity and the planet's eccentricity also play an important role.

For *P-type motion* – i.e., when the planet moves around both stars, we followed the work by Pilat-Lohinger et al. [23], where the stability of inclined planetary motion was studied for the first time in a binary with mass ratio 0.5. Depending on the eccentricity of the binary the inner border of stable motion varies between 2.1 (for circular binary motion) and 3.85 [units of length] (for $e_{binary} = 0.5$) measured from the center of mass of the system. The P-type motion is mainly interesting for very close binaries. Currently all detected planets in binary systems have S-type motion, which is therefore of particular interest.

However, we expect an increasing number of planet detections in binary systems in the future, due to current (CoRoT) and future space missions (Darwin, Kepler, TPF and others), since such systems are more frequent in the solar neighborhood than in the single-star planetary systems. For these future discoveries, the presented general stability studies will be very helpful to determine whether or not a planet is to be found in the stable region.

Acknowledgments

EP-L wishes to acknowledge the support by the Austrian FWF (Hertha Firnberg Project T122). This study was also supported by the International Space Science Institute (ISSI) and benefits from the team "Evolution of Habitable Planets".

References

1 Mayor, M. and Queloz, D.: 1995, *Nature*, **378**, 355.
2 Raghavan, D., Henry, T.J., Mason, B.D., Subasavage, J.P., Jao, W.C., Beaulieu, T.D. and Hambly, N.C.: 2006, *ApJ*, **646**, 523.
3 Duquennoy, A. and Mayor, M.: 1991, *A&A*, **248**, 485.
4 Dvorak, R.: 1984, *CMDA*, **34**, 369.
5 Dvorak, R.: 1986, *A&A*, **167**, 379.
6 Rabl, G. and Dvorak, R.: 1988, *A&A*, **191**, 385.
7 Dvorak, R., Froeschlé, Ch., and Froeschlé, C.: 1989, *A&A*, **226**, 335.
8 Dvorak, R. and Lohinger, E.: 1991, *Proceedings of the NATO ASI Series*, Roy A.E. (Ed.), Plenum Press, New York and London.
9 Lohinger, E. and Dvorak, R.: 1993, *A&A*.
10 Benest, D.: 1988a, *A&A*, **206**, 143.
11 Benest, D.: 1988b, *CMDA*, **43**, 47.
12 Benest, D.: 1989, *A&A*, **223**, 361.
13 Benest, D.: 1993, *CMDA*, **56**, 45.
14 Benest, D.: 1996, *A&A*, **314**, 983.
15 Benest, D.: 1998, *A&A*, **332**, 1147.
16 Holman, M.J., Touma, J. and Tremaine, S.: 1997, *Nature*, **386**, 254.
17 Dvorak, R., Pilat-Lohinger E., Funk B. and Freistetter F.: 2003, *A&A*, **398**, L1–L4.
18 Dvorak, R., Pilat-Lohinger E., Funk B. and Freistetter F.: 2003, *A&A*, **410**, L13–L16.
19 Dvorak, R., Pilat-Lohinger E., Schwarz R. and Freistetter F.: 2004, *A&A*, **426**, L37–L40.
20 Pilat-Lohinger, E. and Funk, B.: 2007, Planetary motion in the binary system HD41004 AB, *A&A*, **submitted**.
21 Holman M.J. and Wiegert P.A.: 1999, *AJ*, **117**, 621.
22 Pilat-Lohinger, E. and Dvorak, R.: 2002, *CMDA*, **82**, 143.
23 Pilat-Lohinger, E., Funk, B., and Dvorak, R. 2003, *A&A*, **400**, 1085.
24 Harrington, R.S.: 1977, *AJ*, **82**, 753.
25 Graziani, F. and Black, D.C.: 1981, *ApJ*, **251**, 337.
26 Black, D.C.: 1982, *AJ*, **87**, 1333.
27 Innanen, K.A., Zheng, J.Q., Mikkola, S. and Valtonen, M.J.: 1997, *AJ*, **113**, 1915.
28 Musielak, Z.E., Cuntz, M., Marshall, E.A., and Stuit, T.D.: 2005, *A&A*, **434**, 355.
29 Stoer, J. and Bulirsch, R.: 1980, Introduction to Numerical Analysis, Springer Verlag, New York.
30 Lichtenegger, H. 1984, *CMDA*, **34**, 357.
31 Hanslmeier, A. and Dvorak, R. 1984, *A&A*, **132**, 203.
32 Froeschlé, C., Lega, E., and Gonczi, R.: 1997, *CMDA*, **67**, 41.
33 Bois, E., Kiseleva-Eggleton, L., Rambaux, N. and Pilat-Lohinger, E.: 2003, *ApJ*, **598**, 1312.
34 B. Érdi and A. Pál: Dynamics of resonant exoplanetary systems, *Proceedings of the 3rd Austrian-Hungarian Workshop on Trojans and related Topic* ed by F. Freistetter, R. Dvorak. B.Érdi (Eötvös University Press 2003), 3.
35 Pilat-Lohinger, E.: 2005, Dynamics of Populations of Planetary Systems, *Proceedings of IAU Coll. 197*, eds. Knezevic Z. and Milani A., Cambridge University Press, 71.
36 Sandor, Zs., SÃijli, A., Erdi, B., Pilat-Lohinger, E. and Dvorak, R.: 2007, *MNRAS*, **375**, 1495.
37 Laskar, J.: 1990, *Icarus*, **88**, 266.
38 Froeschlé, C. and Lega, E.: 1996, *Cel. Mech.*, **64**, 21.
39 Laskar, J.: 1994, *A&A*, **287**, L9.
40 Funk, B., Pilat-Lohinger, E., Dvorak, R., Freistetter, F. and Erdi, B.: 2004, *CMDA*, **90**, 43.
41 B. Érdi, R. Dvorak, Zs. Sándor and E. Pilat-Lohinger: 2004, *MNRAS*, **351**,1943.
42 Asghari, N., Broeg, C., and Carone, L. et al.: 2004, *A&A*, **426**, 31.
43 Goudas, C.L.: 1963, *Icarus*, **2**, 1.

44 Markellos, V.V.: 1977, *MNRAS*, **180**, 103.
45 Markellos, V.V.: 1978, *MNRAS*, **184**, 273.
46 Michalodimitrakis, M.: 1979, *A&A*, **76**, 6.
47 Perdios, E. and Markellos, V.V.: 1988, *Celest. Mech.*, **42**, 187.
48 Zagouras, C.G. and Markellos, V.V.: 1977, *A&A*, **59**, 79.
49 Broucke, R.A.: 2001, *CMDA*, **81**, 321.
50 Kozai, Y.: 1962, *AJ*, **67**, 591.
51 Thomas, F. and Morbidelli, A.: 1996, *CMDA*, **64**, 209.
52 Marchal, C.: 1990, *The Three-Body Problem*, Elsevier, 49.
53 Cochran, W., Hatzes, A., Endl, M., Paulson, D., Walker, G., Campbell, B. and Yang, S.: 2002, *BAAS*, **34**, 42.02.

9
The Transit Method

Heike Rauer and Anders Erikson

Abstract

The number of planets detected by the transit method has increased steadily in recent years as ground-based surveys expand their observational baseline. After the launch of the CoRoT satellite (CNES), however, we expect a significant increase in the number of detected transiting planets, including new classes of objects, like “hot-Earth” planets. This chapter describes the transit method as well as its limitations. The various elements of an effective transit search program are discussed, e.g., selection of the target field, data reduction and noise analysis, transit signal detection, and the required follow-up measurements. Finally, we provide an overview of the, already existing, potential to characterize further the transiting planets. Transiting planets do not only expand the statistics of known planets, but they also allow us to gain further insight into the physical nature of extrasolar planets.

9.1
Introduction

The observation of planetary transits has a long history. They are regularly observed when the moon transits in front of the sun, causing a solar eclipse, and when Mercury or Venus transit the solar disc. As such, transits have already had a great impact on astronomy. Consider, for example, the first attempts to determine the astronomical unit during the 18th century using transits of Venus, and solar eclipses which allowed not only observation of the solar corona, but also gave the opportunity to confirm Einstein’s general relativity theory in 1919. The first transiting planet, HD 209458 b, was observed in 1999 and since then thirteen more transiting planets have been detected. In fact, during the last five years, the observational search for planetary transits has started to become one of the most powerful methods of detecting extrasolar planets.

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

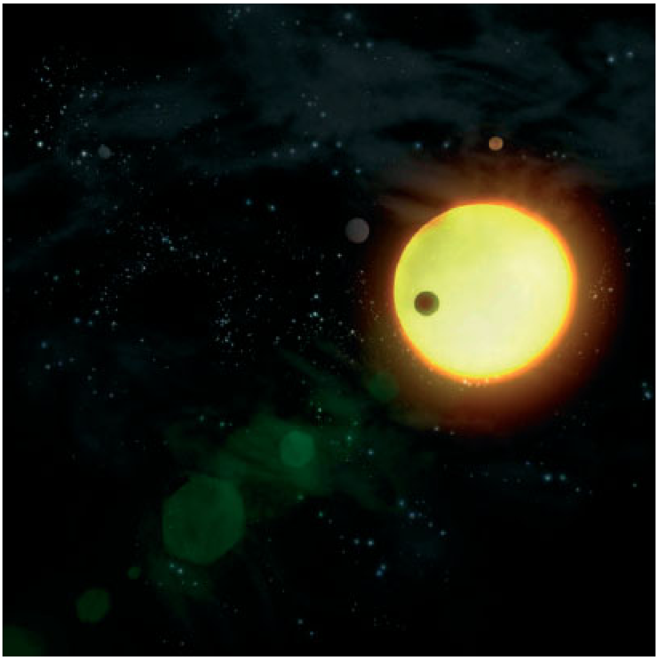


Fig. 9.1 Artist's view of a transiting planet (ESA).

The principle of the transit method for extrasolar planet detection is very simple. Planets with an orbital plane aligned with the line-of-sight from Earth will regularly pass in front of their central star according to the orbital periodicity of the planet. During the transit, the planet dims part of the stellar light. This dimming is proportional to the ratio of the projected areas of the planetary and the stellar surface: e.g., for a Jupiter-sized planet around a solar-like star, the signal is of the order of 1%, and less than 0.01% for an Earth-sized one. To detect planetary transits, photometric stellar light curves are acquired and searched for such dimming events, from which the basic planetary parameters can then be derived.

The transit method allows the detection of planets with favorable alignment of their orbital plane only (Fig. 9.1). Thus, the "geometrical" probability to find such planets is only around 0.5% for Earth-like – and about 10% for "hot Jupiter" – planets. Therefore, transit surveys monitor a large number of stars to reach sufficient detection probabilities.

The duration of the transit event is proportional to the orbital distance, the stellar and planetary radii and the inclination of the orbital plane. For typical "hot Jupiter" planets with orbital distances of 0.03–0.04 AU, the length of a transit is around 3–4 h. For an Earth-sized planet at 1 AU, the transit would last 13.1 h, and for Jupiter (5.2 AU) a transit would last as long as 32.4 h. Thus, very long time sequences are needed to detect distant planets and, consequently, photometric surveys operate over many years.

From the transit signal, the orbit and the radius of the planet can be derived. Complementary measurements are required to derive the planetary mass, for example, by the radial velocity method. Combining the transit method with follow-up measurements, therefore, provides us with the primary planetary parameters: radius, mass, mean density and orbit. In addition, the transit method has already demonstrated its potential for future follow-up observations needed to characterize further the detected planets by, for example, transit spectroscopy and reflected light measurements.

In the following, we first give an overview on the current status of extrasolar planet transit detections in Section 9.2. The transit signal and the derivation of planet parameters are discussed in Section 9.3. In Section 9.4 we discuss the search for transit signals in the acquired light curves and the related follow-up observations which finally lead to confirmed transiting planets. Transit surveys, however, have to deal with a number of limitations which affect their detection efficiency. Some of these, the effective duty cycle, the target field selection and photometric noise sources, are discussed in Section 9.5. Then, Section 9.6 summarizes the factors influencing the detection probabilities of transit surveys. The attempts made to further characterize the nature of transiting planets are summarized in Section 9.7. Finally, Section 9.8 discusses the future prospects of transit surveys via space missions.

9.2
Overview of Detected Transiting Planets

9.2.1
From Predictions to Regular Detections

All the fourteen presently known transiting planets have been found by radial velocity and photometric transit surveys conducted during the last decade and, for a majority of the detections, within the last four years. Despite this, the idea of discovering planets by the transit method is not new and dates back at least 50 years (e.g., [1–3]). The first actual transit search program, the TEP network (Transits of Extrasolar Planets), started in 1994 and focused on an eclipsing binary system over an extended time period [4]. No transiting planet was detected during the survey but the search nevertheless showed that the required photometric accuracy could be reached, thereby demonstrating the potential of the transit method.

The triggering factor for extensive transit surveys of extrasolar planets was the discovery of 51 Peg b and subsequent hot Jupiters by radial velocity

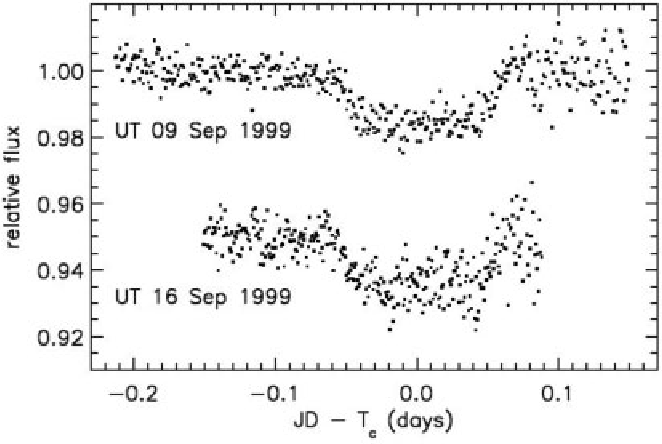


Fig. 9.2 The first observed transit of an extrasolar planet (HD209458 b). The observations were made with the STARE instrument on a planet previously discovered by the radial velocity method [6].

searches in the latter half of the nineties. Based on the properties of these discoveries it was realized that the orbital periods (in the order of days) and the sizes (Jupiter-sized) of the planets found, put them within reach of what could be detected by a ground-based photometric transit survey.

The first transit observed was by the planet HD209458 b. From radial velocity measurements [5] an orbital period of 3.5 days was derived. Photometric observations of the parent star were initiated and the first transit of an extrasolar planet was detected [6, 7] in the year 1999 (see Fig. 9.2). However, it should be noted that, for this first case, the transit method was used as a means of follow-up characterization for a planet discovered by a radial velocity search program. Later observations with the Hubble Space Telescope (HST) yielded a high accuracy characterization of the transit event [8] as displayed in Fig. 9.3.

Since the detection of a transiting planet around HD 209458, a number of new ground-based transit survey programs have been established (see e.g., the overview in [9]). The goal of most of these surveys is to detect planets by the transit method alone, i.e., photometric observations as a means of discovery, followed by radial velocity measurements to confirm and fully characterize the nature of the discovered planets. Generally, the different surveys can be divided into two main groups with respect to the instrumentation used and the adopted strategy:

- Wide-angle surveys with small aperture telescopes (10–30 cm) designed to monitor bright stars (8–15 mag).

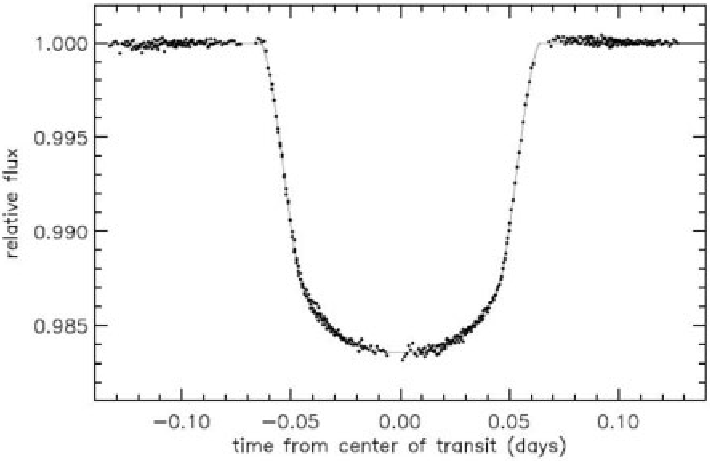


Fig. 9.3 Transit of the planet HD209458b observed with the Hubble Space Telescope [8].

- Narrower surveys with larger aperture telescopes (1–2 m) to go deeper in magnitude and thereby to reach fainter stars (14–20 mag).

For both cases transiting planets have been detected even if their numbers are lower than the initial expectations after the transit discovery of HD 209458 b (see Sections 9.5 and 9.6 for a discussion of the limiting factors).

In addition, due to the fainter magnitudes of the parent stars, surveys using larger telescopes face a more difficult task for the radial velocity follow-up observations needed to secure and characterize transiting planets .

The first extrasolar planet to be detected by the transit method and confirmed by radial velocity measurements was OGLE-TR-56 b in the year 2003 [10]. The OGLE (Optical Gravitational Lensing Experiment) program [11–15] starting in 2001 and using a 1.3 m telescope at the Las Campanas Observatory, Chile, has provided a large number of transiting planet candidates which have been followed by radial velocity measurements by other teams. Since OGLE-TR-56 b four additional transit candidates (see Table 9.1) were confirmed and OGLE, therefore, is presently the most successful transit survey.

Since the first transit detections, the search for extrasolar transit events has gained momentum. In addition to the five planets observed by OGLE, nine additional planets have been found. The majority of the recently detected exoplanets have been discovered by small aperture photometric transit surveys. They include HAT-P-1 b by the HAT (Hungarian Automated Telescope) survey [16], TrES-1 b and TrES-2 b by the Trans-Atlantic Exoplanet Survey [17, 18], WASP-1 b and WASP-2 b by the Super-WASP (Wide Angle Search for Planets) survey [19] and X0-1 b [20]. In addition, ongoing radial velocity surveys have

Table 9.1 Central star magnitude, mag_{star}, and radius, radius_s, transit depth, A, and discovery publication of the fourteen know transiting extrasolar planets.

Planet	mag_{star} (mag)	radius_s (Rsun)	A (%)	References
HAT-P-1	10.4 (V)	1.15±0.1	0.6	[16]
HD 149026b	8.15 (V)	1.45±0.10	0.5	[22]
HD 189733b	7.67 (V)	0.76±0.01	3	[23]
HD 209458b	7.65 (V)	1.146±0.050	1.7	[6, 7]
OGLE-TR-10	14.93 (I)	1.16±0.06	1.2	[40]
OGLE-TR- 56	16.6 (V)	1.32±0.06	1.1	[10]
OGLE-TR-111	15.55 (I)	0.831±0.031	1.8	[41]
OGLE-TR-113	14.42 (I)	0.77±0.02	2.1	[42]
OGLE-TR-132	15.72 (I)	1.43±0.10	0.8	[42]
TrES-1	11.79 (V)	0.811±0.02	1.9	[17]
TrES-2	11.41 (V)	1.00±0.06	1.6	[18]
WASP-1	11.79 (V)	1.24±0.20	1.1	[19]
WASP-2	11.98 (V)	0.78±0.06	1.8	[19]
XO-1	11.3 (V)	0.928±0.03	2	[20]

found two more transiting planets around bright stars. The N2K survey [21] discovered a transiting Saturn-sized planet orbiting the star HD 149026 [22] and a planet orbiting around HD 189733 was found by Bouchy et al. [23] as part of the "ELODIE metallicity-biased search for transiting Hot Jupiters" survey. Additional ongoing surveys like, for example, Vulcan [25], KELT [26], BEST [27], or EXPLORE [28] may provide successful transit discoveries in the near future. A recent survey of a star-rich field near the galactic bulge, performed with HST by the SWEEPS project (Sagittarius Window Eclipsing Extrasolar Planet Search) provided 16 planet candidates [29]. However, the central stars are very faint ($m < 18$ mag) and only two of the candidates are supported by radial velocity follow-up observations, allowing mass estimates. Clearly, follow-up of transit candidates around faint target stars, as detected by large-aperture space missions, will be difficult.

Many attempts have also been made to discover transiting planets in open stellar clusters. An early observation of 47 Tuc by HST [30] showed no transiting planets. At this point, many surveys of clusters have recently been made or are ongoing, like UStAPS (University of St. Andrews Search) [31–34], the EXPLORE-OC survey [32], STEPSS (Survey for Transiting Extrasolar Planets in Stellar Systems) [35, 36], MONITOR [37] and PISCES (Planets in Stellar Clusters Extensive Search) [38, 39].

All fourteen presently detected extrasolar planets are listed in Table 9.1 providing for each planet the central star magnitude and radius, and the observed transit depth. In total, eleven planets were discovered by the photometric transit observations while the remaining three planets were discovered by radial velocity surveys and their transits found during subsequent photometric follow-up measurements. Altogether, this shows the maturing of ground-based photometric transit surveys where the increased number of fully operating surveys with expanding total duty cycles are expected to give an even larger yield of detected extrasolar planets in the years to come.

9.2.2
Basic Parameters of the Known Transiting Planets

Extrasolar planets with detected transit events offer the possibility of a basic characterization of their physical and dynamical nature (as described in Section 9.7). This includes physical properties like the radius and mass and thereby the mean density of the planet, as well as the orbital parameters, like the period and inclination. Based on these properties further modeling can be performed of planetary formation and evolutionary scenarios. In Table 9.2 the basic properties (semi-major axis, orbital period, radius, mass and density) of the 14 know transiting extrasolar planets together with the type, effective temperature (T_{eff}) and metallicity (Fe/H) of their respective central stars, are listed.

The orbital periods of the known planets range from 1.21 days (OGLE-TR-56 b) to 4.47 days (HAT-P-1 b), corresponding to a semi-major axis range between 0.02 and 0.05 AU. This means all known transiting planets are on rather close orbits around their parent stars compared to the total set of known exoplanets and compared to the location of the giant planets within the Solar System. This fact is due to the limited total observational duty cycle of the surveys involved. The time needed to detect 3–4 transit events of the planets in Table 9.2 from an ideal ground-based observation site and under good photometric conditions is of the order of weeks or months and well within reach of the surveys involved. Transit surveys operating for many years are needed for planets in Earth- or Jupiter-like orbits. However, unfortunately the geometrical probability for a distant exoplanet to transit through the line-of-sight decreases with orbital distance (see Section 9.6). Consequently, most likely it will take a very long time before transiting planets, e.g., on Earth-like orbits, are detected from the ground. This prospect looks significantly better, however, for the surveys to be conducted by the space missions CoRoT and Kepler (see Section 9.8).

The known transiting planets have radii in the range 0.72–1.44 R_J and masses between 0.33 and 1.35 M_J. This range is a consequence of the photometric accuracy reached from ground due to the influence of the atmosphere.

Table 9.2 Basic properties of the 14 known transiting extrasolar planets, together with the type, effective temperature (T_{eff}) and metallicity (Fe/H) of their respective central stars. Here, a denotes the orbital semi-major axis of the planet, P the orbital period and r_p and m_p the planet radius and mass, respectively. Up-to-date values are available at obswww.unige.ch/~pont/TRANSITS.htm.

Planet	a (AU)	P (days)	r_p (R_J)	m_p (M_J)	Star	T_{eff} (K)	Fe/H
HAT-P-1b	0.0551	4.46529	$1.36^{+0.11}_{-0.09}$	0.53 ± 0.04	G0 V	5975	0.13
HD 149026b	0.042	2.8766	0.726 ± 0.064	0.330 ± 0.02	G0 IV	6147	0.36
HD 189733b	0.0313	2.218573	1.156 ± 0.046	1.15 ± 0.04	K1-K2	5050	-0.03
HD 209458b	0.045	3.52474859	$1.320^{+0.024}_{-0.025}$	0.657 ± 0.006	G0 V	6117	0.02
OGLE-TR-10	0.04162	3.10129	1.26 ± 0.07	0.63 ± 0.14	G or K	6075	0.28
OGLE-TR-56	0.0225	1.211909	1.30 ± 0.05	1.29 ± 0.12	G	6119	0.25
OGLE-TR-111	0.047	4.0144479	1.067 ± 0.054	0.52 ± 0.13	G or K	5044	0.19
OGLE-TR-113	0.0229	1.4324757	1.09 ± 0.03	1.32 ± 0.19	K	4804	0.15
OGLE-TR-132	0.0306	1.689857	1.14 ± 0.12	1.18 ± 0.07	F	6210	0.37
TrES-1	0.0393	3.030065	1.081 ± 0.029	0.729 ± 0.036	K0 V	5214	0.001
TrES-2	0.0367	2.47063	1.24	1.28	G0 V	5960	
WASP-1	0.0382	2.51997	1.443 ± 0.039	0.867 ± 0.073	F7 V	6200	
WASP-2	0.0307	2.152226	1.038 ± 0.050	(0.81- 0.95)	K1 V	5200	
XO-1	0.0488	3.94134	1.184	0.90 ± 0.07	G1 V	5750	0.015

References: Hat1P: [16]; HD149025b: [22, 43]; HD 189733b: [16, 23, 44]; HD209458: [45–47]; OGLE10: [48]; OGLE56: [48]; OGLE111: [41, 49, 50]; OGLE113: [51]; Tres1: [17, 46]; Tres2: [18], WASP-1: [19, 52, 53]; WASP-2b: [19, 52], X0-1: [20, 54].

Thus, the Earth's atmosphere puts a rather firm lower limit on the sizes of detectable planets.

In Fig. 9.4 the planet mass is plotted against the semi-major axis for all known exoplanets. The parameter range of the fourteen known transiting planets is marked by a box. Both selection effects mentioned above are clearly visible in the figure. The detected transiting planets are located in a rather constrained area compared to all known exoplanets, all with rather small semi-major axis and with Jupiter-sized masses.

In Fig. 9.5 the radius versus the planetary mass of the fourteen transiting exoplanets, as well as Jupiter and Saturn are displayed. The mean densities for the transiting exoplanets range between about 0.3 and 1.5 g cm^{-3}, partly comparable to the densities of Jupiter and Saturn in the Solar System, and clearly identify them as giant gas planets. For four planets (HAT-P-1, HD 209458b and OGLE-TR-10 and WASP-1) radii have been found that are larger than expected from structural models of this type of planet (see, e.g., [43] for a detailed discussion of the phenomena and possible causes).

Based on the first six transiting planets discovered, a correlation between the orbital periods and masses was originally proposed Mazeh et al. [55]. In

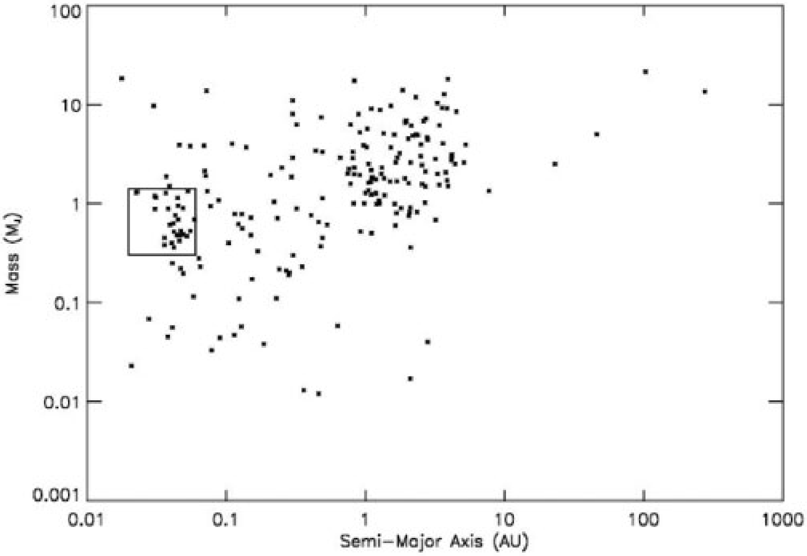


Fig. 9.4 Planet mass over semi-major axis for all known exoplanets. All fourteen known transiting planets are located within the marked box.

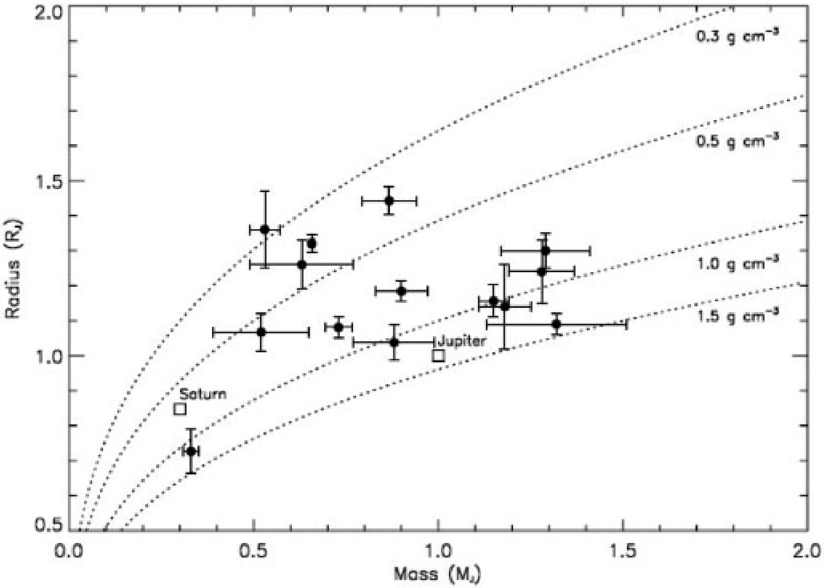


Fig. 9.5 Planetary radius over the mass for the fourteen transiting exoplanets, together with Jupiter and Saturn.

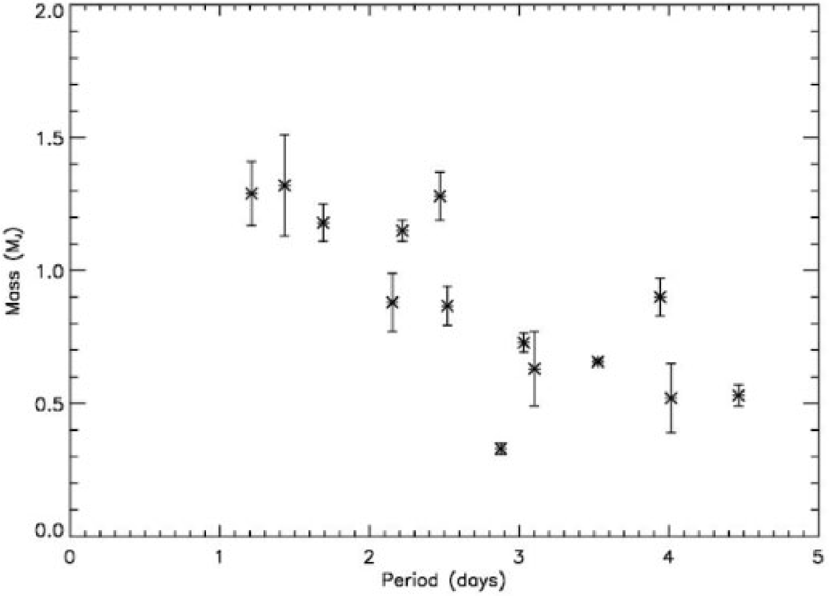


Fig. 9.6 Orbital period vs. mass for the fourteen transiting exoplanets.

Fig. 9.6 the masses are plotted for the present fourteen planets against their orbital period, apparently confirming the originally proposed correlation. A possible explanation for this is a critical mass limit depending upon the orbital distance of the planet below which the atmosphere is quickly evaporated [56] coupled with a decreased detection probability at larger orbital distances. A larger sample of transiting exoplanets is needed to verify the present correlation and fully determine its origin.

9.3
The Transit Signal

Based on the detected transit signal some basic planetary parameters can be determined. The geometry of a planetary transit is shown in Fig. 9.7. The periodicity of the measured transit directly provides the orbital period of the planet, P. With the help of Kepler's third law, the orbital distance of the planet to its star, the semi-major axis a_p, can be obtained:

$$a_p^3 = \frac{GM_sP^2}{4\pi^2} \tag{9.1}$$

where G is the gravitational constant and M_s the stellar mass. The planetary mass is small compared to the stellar mass and has been neglected.

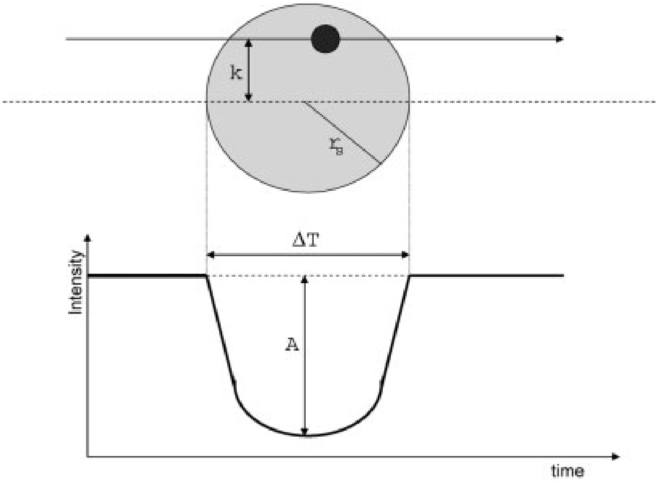


Fig. 9.7 Illustration of the transit geometry and signal.

The amplitude, A, of the transit signal is proportional to the area on the star occulted by the transiting planet and is, therefore, proportional to the ratio of the stellar and planetary radii, r_s and r_p, respectively:

$$A = \frac{\triangle F}{F} = \left(\frac{r_p}{r_s}\right)^2 \tag{9.2}$$

Here, F is the measured stellar flux.

For a solar-type star (G2V) and a Jupiter-sized planet (1 Jupiter radius (R_J)), the amplitude of the intensity dimming during transit is about 1%. This is close to the detection limit of most ground-based transit surveys, but nevertheless within reach from excellent observation sites with high accuracy photometry. For small dwarf stars, however, the dimming of the stellar intensity can increase up to 5% (see Fig. 9.8), a signal easily detected with good confidence by most transit surveys located even at moderate sites. Good detection limits would even allow the detection of smaller planets for such small central stars.

The duration of a transit event, $\triangle T$, can be expressed as [57]:

$$\triangle T = 2Z(r_p + r_s)\frac{\sqrt{1-e^2}}{1+e\cos\phi}\left(\frac{P}{2\pi G M_s}\right) \tag{9.3}$$

with

$$Z = \sqrt{1 - \frac{a_t^2 \cos^2 i}{(r_p + r_s)^2}} \tag{9.4}$$

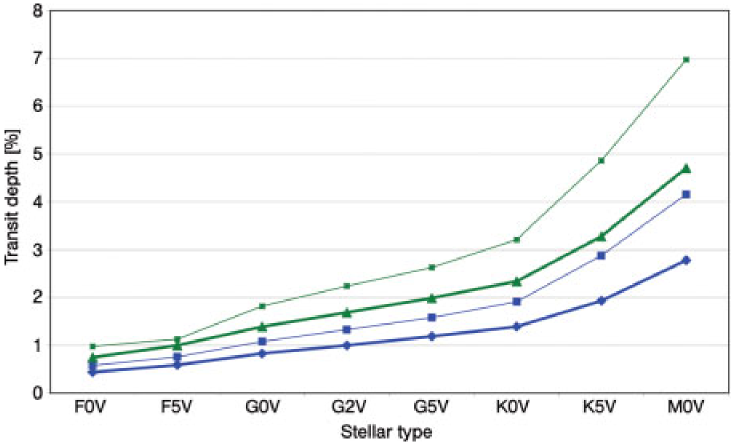


Fig. 9.8 Central transit depths for different stellar types. Thick blue line: $R = 1R_J$, without limb-darkening; thin blue line: same, but with quadratic limb-darkening; green lines: same as blue, but for $R = 1.3R_J$.

Here, a_t^2 is the distance of planet and star during the transit, i the inclination of the orbital plane, e the eccentricity, and ϕ the eccentric viewing angle of the transit (see [57]).

The shape of the transit light curve depends on several parameters: the radius of the star and planet, the orbital inclination, and the stellar limb darkening. For noisy data (e.g., Fig. 9.2), which do not allow us to sample the shape of the transit light curve in detail, the star transited can be approximated by a uniformly radiating surface. Then, the transit has a fast ingress and egress with a flat bottom of the light curve during the event. Thus, when no limb-darkening of the central star is taken into account the signal depth during transit is simply given by [58]:

$$F(p,z) = 1 - L(p,z) \tag{9.5}$$

With

$$L(p,z) = \begin{cases} 0 & 1+p<z \\ \frac{1}{\pi}\left[p^2\kappa_0 + \kappa_1 - \sqrt{\frac{4z^2-(1+z^2-p^2)^2}{4}}\right] & \mid 1-p \mid < z \leq \mid 1+p \mid \\ p^2 & z \leq 1-p \\ 1 & z \leq p-1 \end{cases}$$

Here, d is the distance of the centres of star and planet. The light curve is expressed in terms of the normalized separation of star and planet, $z = d/r_s$, and their size ratio, $p = r_p/r_s$. The coefficients are given by:

$$\kappa_0 = \cos^{-1}[(p^2 + z^2 - 1)/2pz]; \ \kappa_1 = \cos^{-1}[(1 - p^2 + z^2)/2z] \tag{9.6}$$

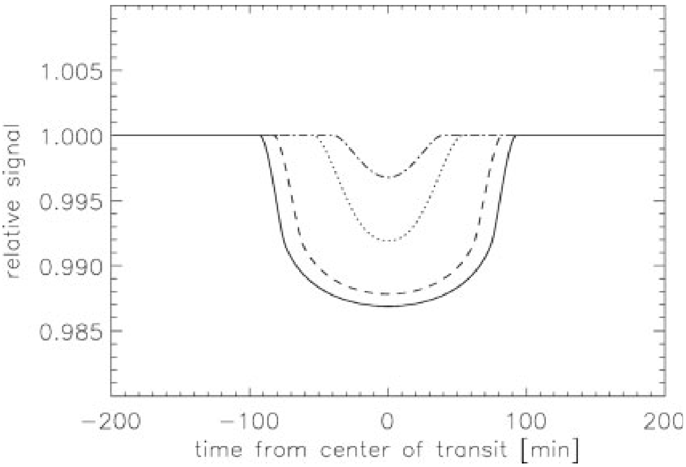


Fig. 9.9 Transit curve with limb-darkening for a central transit (solid line) up to a just-grazing transit (dashed-dotted line) [60].

The planetary radius and the orbital inclination can therefore be derived by fitting this relation to a measured transit signal, if the stellar radius is known. Obviously, the accuracy of the stellar parameters places a strict limit on the accuracy which can be obtained for the planet parameters. It has been suggested that stellar mass estimates are uncertain by up to 5% [59]. If neither the stellar mass, radius and limb-darkening law are known, it is difficult to disentangle the parameters entering the derivation of r_p (see discussion in Charbonneau et al. [43]).

The true shape of the transit light curve is of course affected by stellar limb-darkening, and this effect has to be taken into account for high-quality, high S/N light curves (e.g. Fig. 9.3). An expression for the flux of an occulted star considering nonlinear limb-darkening is also given in [58] together with an algorithm which can be used to compute the transit shape for various stellar types. With limb-darkening present, the stellar flux is concentrated more strongly towards the center of the star leading to a gradual egress and ingress of the transit light curve and a deeper central dimming. Thus, the amplitude of the transit signal is larger if one considers limb darkening (Fig. 9.9). A large planet, e.g. 1.3 R_J, would cause a relatively deep signal of more than 3% if orbiting a K0V star. Transiting planets around M dwarfs will cause even stronger signals.

It is often thought that the relatively flat bottom of a transit light curve can provide a first indication of the need to separate a planet detection from eclipsing binaries. Due to limb-darkening, however, planetary transits result in curved light curves, appearing similar to light curves of eclipsing binaries. Furthermore, noncentral transits are clearly v-shaped (Fig. 9.9). Therefore, the shape alone is only a poor indicator to separate firm planet detections from other causes of the dimming of light (see Section 9.4.2).

The mean density of a planet is obtained when the planetary radius derived from transit data is combined with $m_p \sin i$ determined by radial velocity (rv) measurements. The orbit inclination, i, which is an unknown in rv-observations, can be determined from transit data as described above. Therefore, the combination of transit observations with rv-measurements allows the determination of the basic planetary parameters: mass, radius, mean density and orbit.

9.4
From Light Curves to Planets

9.4.1
Transit Signal Detection

Highly accurate photometric light curves are produced by relative photometry, comparing the intensity of the stars in the field-of-view to a set of reference stars in the same field. After stellar light curves have been extracted from a transit survey data set, they are scanned for transit-like signals. Several algorithms to detect periodic transit events in light curves have been published, such as box-fitting least-square algorithms (BLS) [19, 61], matched filter approaches , and the Bayesian technique [62]. In a matched-filter approach, a transit-shaped signal of assumed length and depth is correlated with the measured light curves to detect such a signal in the data. Most transit surveys apply BLS or matched-filter algorithms for the initial transit search. Comparisons on the performance of these algorithms have not shown large differences between BLS and match-filter approaches in general [63, 64]. The Bayesian technique also allows us to reconstruct the shape of the transit signal, but requires very good S/N input data. As an exercise to test the performance of different methods, several transit-search algorithms have been applied to simulated light curves as expected from the CoRoT satellite mission. A comparison of their performance [65] favored matched-filter algorithms, although the difference in the various approaches was not as large as one might expect, based on the range of simplicity/complexity of the methods used.

Transit surveys usually obtain light curves for several tens of thousands of stars and the resulting list of planet-transit candidates can still include a large number of objects. Naturally, individual inspection of all of these is impossible and, therefore, some automatic pre-selection algorithm has to be applied. Usually a limit is set on the transit amplitude searched for (e.g., less than 10%) and the duration of the event like, e.g., <5 h for ground-based transit surveys which concentrate on hot Jupiter planets. The resulting list of candidates can then be further reduced visually before entering the follow-up process (see Section 9.4.2).

The result of the search for transit signals is an initial list of candidates to be confirmed or rejected by various types of follow-up observations (see Section 9.4.2). Criteria for a good candidate are:

- The transit depth must be in agreement with the planet for the spectral type of the star (see Fig. 9.8).
- The transit duration must agree with the planet for the radius of the central star.
- The ellipsoidal and sinusoidal amplitude of the light curve out of transit must be low (such variations would indicate binaries).
- The transit must be observed at least three times to provide a period and a confirmed signal. In practice, for most surveys, more events are common.
- A flat-bottomed shape of the transit signal favors a small companion, in comparison to a clearly v-shaped binary light curve, although it is not final proof for a planet.

9.4.2
Confirmation or Rejection

A planet candidate list resulting from transit signal search does not indicate in itself the detection of extrasolar planets. The measured periodic dimming of a stellar light curve can have many causes, like eclipsing binaries, variable stars and even instrumental effects. While instrumental causes can usually be quickly ruled out by checking the data set, astronomical causes of the detected signal require further follow-up observations. Therefore, such observations are an integral part of most transit surveys.

There can be several causes of confusing events (see, e.g., [66]), the major ones being:

1. Nonplanetary eclipses, like:
    (a) grazing eclipsing binaries of two main-sequence stars,
    (b) main-sequence stars transiting giant stars,
    (c) M-stars or brown dwarfs transiting main sequence stars.

2. Diluted signals by, e.g.:
   (a) a third, unresolved star within the instrument point-spread-function (PSF) diluting the transit signal,
   (b) a background eclipsing binary system or an unresolved hierarchical triple causing the observed transit signal.
3. Stellar spots, mimicking the transit of a small transiting planet.

Nonplanetary transit signals should be identified as soon as possible from a light curve analysis and available catalogue information to minimize observing time intensive follow-up measurements. This results in a number of tests which have to be passed before a candidate is confirmed as a planet:

- *Carefully inspect the measured light curve* (resolve cases 1(a) and 1(b), some 2(a) and 2(b) for:
  - instrumental effects,
  - secondary transits, indicating eclipsing binaries,
  - modulation of the light curve, either by ellipsoidal modulation as expected for eclipsing binaries when tidal forces elongate a star towards its companion [67], or sinusoidal modulation caused by the heated hemisphere of the orbiting secondary; and
  - a v-shaped transit signal, showing long egress and ingress. However, a v-shaped light curve may well be caused by a planet (see Fig. 9.9) and a candidate should not be rejected, based on the light curve shape only.
- *Check* whether *the central star* is a binary, variable or a giant star (resolve cases 1(a) and 1(b) by:
  - checking information in available catalogues, and/or
  - performing additional photometric and spectroscopic observations to determine or refine the stellar type.
- Perform *high spatial resolution photometry*to identify diluted cases (resolve many 2(a) and (2b), e.g., as time-series observations covering the transit event, or as time-efficient on-off transit snapshots.
- Perform *multi-colour photometry* to identify binaries (especially for cases 2(a), 2(b) and 3, also 1(a) and 1(b)). Since transits are to first order achromatic, multi-colour observations help to identify eclipsing stars of different colour (temperature) or identical eclipsing stars with a diluting third star of different colour. Colour photometry also helps in identifying stellar spots that show a deeper transit signal in the blue than in the red wavelength range.

- Perform *low-resolution spectroscopy* (resolve cases 1(a) and 1(b)) to reject eclipsing binary systems.
- Perform *high-resolution spectroscopic measurements* (resolve all cases) of the radial-velocity Doppler-shift to determine the mass of the companion and bi-sector analysis to identify binaries.

The first step in preparing a planet candidate list should therefore be a careful inspection of the light curves, combined with knowledge on the central star from available stellar catalogues or preparatory observations. Many candidates can usually already be ruled out at this level. For example, inspection of light curves of the OGLE survey for ellipsoidal and sinusoidal modulations identified these effects in several of the published candidates [68, 69]. Additional tools proposed to eliminate false candidates solely from light curve inspections are proposed by, e.g., [70] and [57]. They do, however, require very good S/N light curves. Finally, dedicated photometric and low-resolution follow-up observations help to reduce further the number of candidates on the list, to minimize the effort on high-spectral resolution Doppler-shift measurements. These will eventually allow the confirmation of a planet candidate and the determination of its mass.

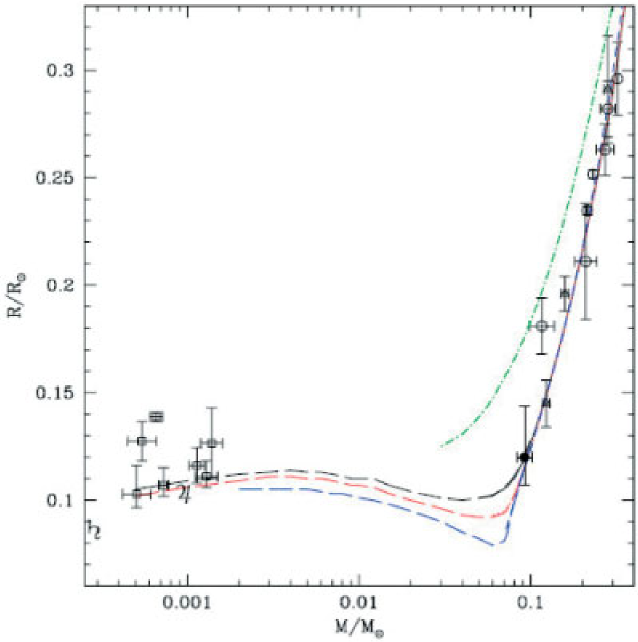


Fig. 9.10 Mass versus radius for planets and low-mass stars [71].

The most difficult cases to solve are those diluted eclipsing binaries (case 2) that cannot be resolved by high-spatial resolution imaging. It may however be possible to solve some of these cases using high-S/N data, e.g., by an analysis of the light curve for secondary eclipses, a multi-color analysis solving systems with different temperature, or Doppler measurements. Another very difficult case to solve is small brown dwarfs orbiting main-sequence stars (case 1(c)). An example was recently published (Fig. 9.10) by [71]. In such cases the secondary eclipse caused by the dwarf is most likely not visible and only Doppler measurements can resolve these issues. At the limit of small terrestrial planets and faint central stars there will be, however, cases where the candidates cannot be resolved with existing instrumentation.

9.5
Characterizing Factors of Transit Searches

The efficiency of transit surveys is limited by several factors: instrument and observation performance, photometric accuracy, transit-signal detection and follow-up performance. The latter have already been discussed in previous Sections. Here, we present a discussion on the effective survey duty cycle, target-field selection and the photometric accuracy achieved.

9.5.1
The Effective Duty Cycle

A good planet transit candidate must show at least three transit signals in the observational data set. Therefore, the minimum-time baseline for a transit survey is at least three times the orbital period of the targeted planets: $\triangle t_{min} \geq 3P$. However, the minimum baseline is sufficient only if continuous observations of high quality are available over the whole time period (100% duty cycle). The duty cycle of an observation is given as the ratio of the observing time on the target field, t_{target}, to the total observing period, t_{obs}, over which the survey has been performed: duty cycle = t_{target}/t_{obs}.

Space missions can reach very high duty cycles of more than 90%. Together with the higher photometric precision, this is one of the great advantages of space-based observations for transit search. Ground-based transit surveys, however, are severely limited in their observing efficiency. Transits occurring during day time and bad weather periods remain undetected. Even at a good observing site the effective duty cycle of a survey cannot be more than about 50% on average, simply due to the day/night variation. In addition, during full moon periods the photometric accuracy is reduced and usually below the limit for the weak transit signals. The effective duty cycle of most surveys is even lower, because periods of poor weather conditions or occasional instrumental problems further reduce the efficiency. As a result, ground-based transit surveys have to span long time periods on a given target field, $\triangle t \gg \triangle t_{min}$, to cover at least three events in their data. The effective duty cycle of ground-based transit surveys results in high detection probabilities only for short-period planets.

Figure 9.11 shows the probability of detection for three transit events, based on the observational coverage of the orbits of planets. We assumed an observational campaign stretching over seven months and an effective duty cycle of around 40% (i.e., excluding bad weather and full-moon time) at an excellent observing site (Cerro Armazones Observatory, Chile). In the figure, the observational coverage of three planetary orbits is plotted against the semi-major axis. A coverage of one corresponds to 100% probability to detect three transits (taking into account only the observing sequence). In Fig. 9.11 the probability of detecting three transits is reduced beyond about 0.08 AU because of the limitations imposed by the effective duty cycle over the observed time period. In addition, a characteristic pattern can be seen. At regular frequencies the probability drops, corresponding to transit events occurring dur-

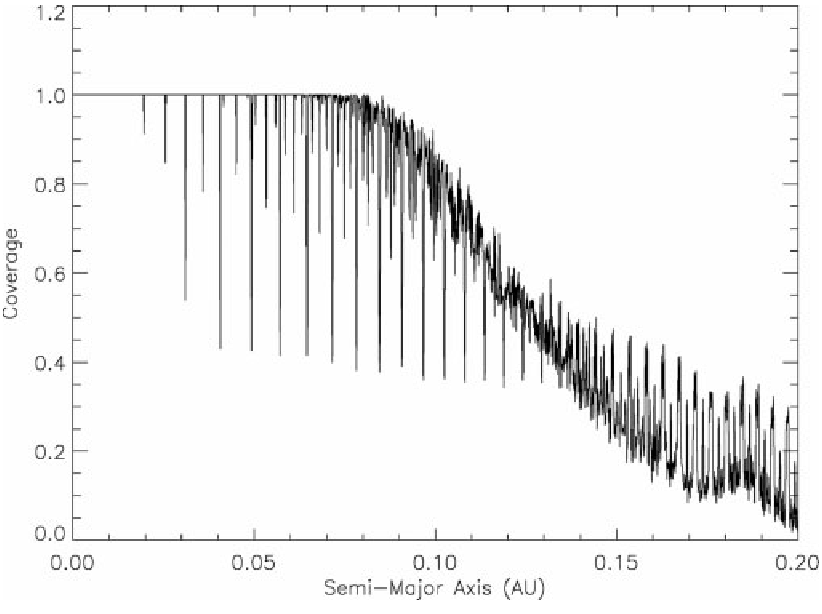


Fig. 9.11 The coverage of the planetary orbits versus semi-major axis for three full planetary orbits. A coverage of one means that three transits can be detected. A duty cycle of 40% was assumed for a target field observed from a very good site (Cerro Armazones Observatory, Chile), over seven months.

ing day time. In fact transit surveys with very poor observational coverage can encounter a reverse stroboscopic effect which always favors the detection of multiples of days, because only transits always fully covered during the observing nights can be detected with high significance in poor sampling surveys (see large distances in Fig. 9.11).

9.5.2
Target Field Selection

The transit method does not require an input catalogue of target stars, and the detections are therefore less biased by observational selection effects than, e.g., rv-measurements. Nevertheless, an intelligent selection of the target field can significantly increase the detection probability of any given survey. It is known that planets around giant stars are extremely difficult to detect, even by space missions, due to their low amplitude. In areas of the sky with a large number of overlapping stellar PSFs it is difficult to obtain high-precision photometric data. Furthermore, background galaxies or interstellar clouds affect the photometric precision of foreground stars. To maximize the efficiency of a transit survey, a careful selection of the target field is therefore important.

Criteria for the target-field selection are:

(a) A large number of dwarf stars, with radii resulting in detectable signals for a given transit survey threshold.

(b) Little crowding, optimized for the spatial resolution of a survey.

Since a large number of light curves is the fundamental condition for achieving a sufficient probability to detect transit signals, it is crucial to optimize the target fields for dwarf stars which provide reasonably deep transit signals when orbited by a planet. It is also useful to avoid fields in the galactic plane, which are excluded by criterion (b). As a result of criteria (a) and (b), good target fields are near to, but not, in the galactic plane.

9.5.3
Noise Sources Affecting the Photometry

Noise in the stellar light curves is caused by the instrumental system used and related to the incoming radiative flux through the atmosphere. To define an optimal transit search system therefore requires the right combination of instrument choice and telescope site. CCD calibration and photometric data reduction then aim at minimizing the effects of the various noise sources in the final light curves. However, some noise sources, like photon noise, are always present and even the most sophisticated data-reduction procedures never work perfectly.

Noise present in the final stellar light curves divide into different types:

1. *Random noise:* photon noise, scintillation.
2. *Detector defects:* imperfect corrections for: pixel sensitivity variations, hot pixels, dead pixels, etc.
3. *Systematic temporal variations:* imperfect corrections for: varying dark and/or bias level, extinction, moving clouds, seeing variations, stellar activity, etc.

In the following, we discuss these noise sources in more detail:

Type 1: Random Noise

An ideal system is limited only by noise of type 1. A basic and unavoidable noise is the photon noise of the target stars and the sky background. Both noise sources are independent and obviously the choice of a good dark site will significantly improve the detection capabilities.

Another source of noise results from scintillation effects. An approximation of scintillation noise can be given as [72]:

$$\sigma_{sc} = 0.09 D^{-\frac{2}{3}} A_z^x e^{-\frac{h}{h_0}} (2t)^{-\frac{1}{2}} \tag{9.7}$$

Here, D is the telescope aperture in cm, h the height of the telescope site, h_0 the scale height of the atmosphere (8000 m), t the exposure time and A_z the air mass. The exponent x depends on the angle of the wind direction with the observing direction and an average value of 1.75 is often assumed. The scintillation noise is independent of the stellar magnitude. It depends, however, on the aperture used for the observations. Scintillation noise is large for small apertures, in particular below 10 cm. Thus, for many transit surveys using small lenses, scintillation can be a significant source of noise.

Type 2: Detector Effects

CCD detectors vary in sensitivity from pixel to pixel. Depending on the quality of the CCD used, the detector may also show "dead" pixel or columns. In addition, some pixels may show continuous high levels of thermal electrons ("hot" pixels, that might also vary with time.). All these effects always apply to the same fixed pixels on the CCD. Let us first assume that we always image the same stars on the same pixels (perfect guiding). In this case, some light curves will always be lost when stars are imaged on "dead" or "hot" pixels. However, pixel-to-pixel sensitivity variations obviously do not affect the final light curves to a high degree when the stellar images remain fixed on the CCD.

High-accuracy guiding, therefore, is the observing strategy chosen when no flat fields can be obtained (e.g., on space satellites). To obtain very low noise light curves (e.g., to detect Earth-sized planets) will require guiding even on the level of sub-pixel accuracy, to avoid intra-pixel sensitivity variations.

In ground-based observations such guiding accuracies are usually not achieved. The level of noise in the light curves resulting from "bad" pixels and imperfect corrections of pixel sensitivity variations then strongly depends on the quality of the guiding system. In the case of imperfect guiding, for example, stars will "move" across the CCD and are imaged on a "dead" or "hot" pixel for a certain time. In the final light curves, this can mimic stellar intensity variations and leads to an enhanced noise level over the night as the stars move across different pixels. The same is true if no proper flat field can be obtained and stars move across pixels with different sensitivities. Also intra-pixel sensitivity variations can be important in this respect for highly accurate photometry. Telescope systems without guiding will, of course, suffer most from this type of noise.

In a similar way, additional noise sources related to the detector can result from variations of the bias and dark levels across the CCD frame. Temporal variations of bias and dark levels affect the residual systematic noise in the final light curves and are discussed below.

Type 3: Systematic Temporal Variations

Noise of type 3 is detected in the final light curves as residuals showing small, but correlated variations of the stellar intensity. Such residuals are indeed found [48]. It is difficult to identify clearly the cause of such residuals, because most likely they are a combination of the various effects contributing to type-3 noise. For example, variations in atmospheric transparency are an additional source of "noise" in the light curves. Atmospheric transparency varies over the night with changing airmass (extinction). In addition, it can vary on short time scales, for example, related to cirrus clouds. While extinction correction is an integrated part of the data reduction procedures, it can be a problem for very wide-field surveys where the extinction can vary over the field of view, and working on sub-frames may be the only choice, leading to an effective correction. In the ideal case, also short-term transparency variations are corrected in the pipeline during extinction correction. In practice, most surveys use reference stars in the same image to monitor transparency variations, thus performing relative photometry. However, the correction is often not perfect, leading to systematic residuals. Colour-dependent effects can further lead to residuals varying from star to star. Another source of systematic noise in the light curves is related again to the instrument used. Small variations of dark current, bias level and read-out noise may all contribute to the systematics. And finally, the central star itself can be variable and provide a source of correlated noise in the light curves.

Since it is difficult to find the physical causes of residuals, correction methods have been developed that require no knowledge of the nature of the residual noise [73]. These correction algorithms help to achieve the required low noise level in the light curves of many transit surveys.

9.6
Detection Probabilities

Up to now, only 14 transiting planets have been found, although several hundred thousand stars are monitored in several survey programs. What is the expected number of transits in such surveys, and why is the observed number so low?

The most uncertain parameter when performing estimates of expected transit detections is the fraction of stars which harbor planets, f_{planet}. So far, we have no information on the true mass-distribution function of planets, because detections of radial velocity surveys are biased towards large gas-giants. However, the ongoing rv surveys have provided more than 200 planet detections and allow us to obtain a first estimate of f_{planet} at least for the close-in hot Jupiters. These are also the target of the ground-based transit searches. Numbers for f_{planet} in the literature are around 1% [74].

The "geometric" probability, P_g, considers the probability of a planet to pass in front of its star while viewed along a line-of-sight, L. Thus, P_g depends on the radii of the star and planet, the orbital planet distance and eccentricity, e, and the longitude of periastron, ω [43]:

$$P_g = 0.0045 \left(\frac{1AU}{a}\right)\left(\frac{r_s + r_p}{r_{sun}}\right)\left[\frac{1 + e\cos\left(\frac{\pi}{2}\right)}{1 - e^2}\right] \tag{9.8}$$

Figure 9.12 shows P_g over orbital period for a Jupiter-sized planet around three different types of stars. For a solar type central star and a "hot Jupiter"-like planet, P_g, is around 10%. It drops to about 0.5% for a planet at 1 AU stellar distance. The geometric probability has two implications for the transit method: (a) it is biased towards short-period planets; (b) it requires the monitoring of a large number of stars.

Since the detection limit of most ground-based surveys is around 1 % transit depth, only large planets around dwarf stars have significant detection probabilities. Thus, what is relevant when predicting transit detections is not the total number of stars in a target field, but only the number of dwarf stars, N_{dwarf} (see Section 9.5.2 on the selection of target fields). For example, only 50% of the stars in fields of the SuperWASP survey are late-type dwarfs [75].

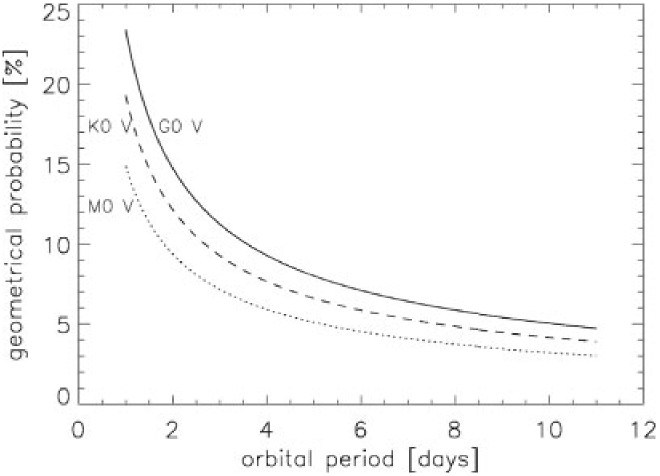


Fig. 9.12 The "geometrical" probability of viewing a planetary system edge-on for G, K and M stars.

Only light curves with sufficient S/N allow for the detecting of transits and therefore the number of light curves for which the required accuracy can be obtained, $N_{S/N}$, is crucial for the transit discovery rate of an operating survey. For example, in a transit survey like SuperWASP [75] about 50% of the detected stars in the survey have sufficient S/N.

Taking into account only the parameters discussed so far, leads to relatively high detection probabilities. For example, a search program monitoring 100 000 dwarf stars, of which 1% have a hot Jupiter planet, would lead to 25 detected planets. Obviously, this is not in agreement with real detection rates. There are, however, additional factors further reducing the detection probability of any given transit survey.

As discussed in Section 9.5.1, planetary transits often remain unobserved, unless the effective duty cycle of a survey is very high. Thus, for space missions we expect a relatively large number of detected transits, whereas in ground-based surveys the transit detection probability strongly depends on the duty cycle, P_{duty}. A transit survey has to be performed until at least three transits of a planet in the field-of-view are observed. If the duty cycle is high, this can be achieved quickly. If the duty cycle is low this may require observations of a target field over several seasons. For observations during one season, we estimate P_{duty} to be around 30–50%.

Putting these numbers together ($f_{planet} \cdot P_g \cdot N_{dwarf} \cdot N_{S/N} \cdot P_{duty}$), we further reduce our expected detection rate to 2–3 planets per 100 000 stars surveyed.

So far we assumed that all planetary transit signals will be detected by the transit search method applied. Of course, this will not be the case if a transit signal is close to the noise level, especially in the presence of residual systematic noise and high stellar variability [48]. For example, tests on the performance of transit search algorithms for the CoRoT space mission Moutou et al. [65] gave a detection rate of about 80% of the transit signals implemented in simulated light curves. The detection rate is of course higher for large planets, and detection remains difficult for small amplitude signals, in particular for variable central stars.

The actual number of expected planets may be reduced further if planets predominantly form around metal-rich host stars, as suggested by studies of the known planetary systems [76, 77].

Several estimates of the efficiency of transit searches have been made. A prediction made by Horne [9] for ongoing and planned surveys was shown to be overly optimistic compared with real detection rates. Other predictions find detection rates of one to a few "hot Jupiters" per few ten thousand stars (e.g. [66, 78]). Many surveys do not yet reach the duty cycle required for high detection statistics, in addition to relatively high noise levels. However, the number of detected transiting planets will certainly increase significantly in the near future.

9.7
Further Characterization of Transiting Planets

The transit detection and basic characterization of exoplanets as described in the previous Sections form a first step in the investigation of their nature. Once a transiting planet is securely detected a number of further investigation techniques will allow a more precise understanding of the planets and their environment. Most obvious are further observations that improve the accuracy of the transit epoch and/or the information on the light curve shape during the event. Below, some of the methods used or proposed for further investigations of transiting planets are described.

Both improved photometric precision and multi-band photometry during the transiting event will allow a better constraint of the planetary radius (as discussed in Section 9.3; see also [79]). Improved modeling of the observed light curve during the transiting event decreases the uncertainty in the planetary radius determined, since the uncertainties introduced by the limited knowledge of radius, mass and limb-darkening parameters of the parent star will be minimized. A recent example is the planet HD 209458 b [45].

High-resolution spectroscopic observations during the transit allow the determination of the sense of stellar rotation as well as the projected angle of the orbital plane of the exoplanet to the spin axis of the star through the Rossiter–McLaughlin effect [80, 81] and, e.g., [82, 83]. During the transit, different parts of the stellar disk are obscured by the planet, resulting in an asymmetric distortion of the line profiles of the stellar spectrum. Therefore, based on radial velocity measurements, from the resulting time-dependent shift of the stellar absorption lines, the above properties of the stellar system can be modeled. This type of investigation has been performed for HD 209458 [47] and HD 149026.

Precise measurements of the timing of a transiting exoplanet can be used to detect additional satellites or planet companions in the planetary system due to the timing anomalies introduced by gravitational perturbations of additional bodies [84–86].

Transiting planets provide the opportunity to probe the composition of close-in extrasolar planet atmospheres as well as the possible presence of clouds and the general weather pattern. It should be noted that the detections are not due to direct imaging of the planet, but rather through unresolved observations at different eclipsing geometries (i.e., observation both in and outside the transiting event).

Although extremely challenging at visible wavelengths it is, in principle, possible to detect the reflected light of exoplanets through both spectroscopy and photometry. The main problem is to reach the S/N ratio needed to detect the planet and therefore the method is better suited for observations from space. In the same way as for the planets in the Solar System, the flux observed will depend on the phase function and the geometric albedo of the planet. Ground-based studies have been able to put upper limits on the geometric albedo for some exoplanets (e.g. [87, 88]). Rowe et al. [89] have put an upper limit on the albedo of the transiting planet HD 209458 b based on observations from the MOST satellite mission.

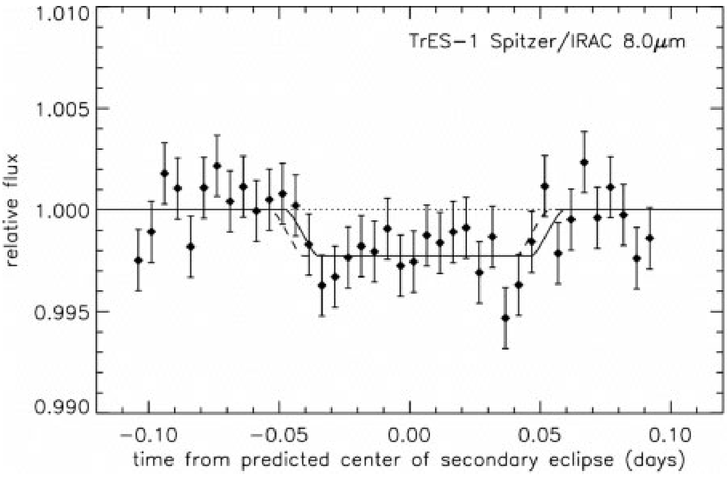


Fig. 9.13 The relative flux during the secondary eclipse of TrES-1 b. (From [105].).

Transmission spectroscopy compares stellar spectra acquired during the transit event to spectra outside the event to detect absorption features in the planets upper atmosphere, thereby giving insights into the atmospheric composition. After many unsuccessful attempts and predictions made for the detection of atmospheric absorptions [90–97], the first such detection was made for absorptions due to atomic sodium in the atmosphere of HD 209458 observed with HST [98]. In the following, H I, O I and C II have been detected, and upper limits put on other species, again using HST [99–104].

The infrared emission of an eclipsing planetary system can be observed to determine the planet-to-star brightness ratio during the secondary eclipse when the planet is behind the parent star. Based on such observations planetary temperatures and albedos can be estimated. Using the Spitzer Space Telescope Charbonneau et al. [105] and Deming et al. [106, 107]) have detected infrared emission of TrES-1 b (see Fig. 9.13), HD 209458 b and HD189733 b.

Together all these various methods provide a palette of tools that can be applied to further characterize the detected transiting exoplanets. A good example is the planet HD 209458 b, where in principle, all the above-mentioned methods have been applied.

9.8
Future Prospects

The statistics of known transiting extrasolar planets is heavily biased by observational constraints as outlined in the previous Sections. As shown in Fig 9.14 the present detection capacity of ground-based transit surveys operates in a rather limited sub-set of the total known population of extrasolar planets, and is focussed on hot Jupiters. One can expect that the number of planets detected from the ground in this size range will increase in the years to come, and potentially, around low-mass M stars the detection of terrestrial planets may even already be possible from the ground. In particular, steadily accumulating total duty cycles for the operating surveys, as well as the potential of telescope systems operating within networks distributed all over the globe, will push the transit detection region outwards, as shown in Fig. 9.14.

However, the expansion of transit detections towards the range of small terrestrial planets around solar-like stars will be in the detection domain of dedicated space missions. Here the great advantages are a significantly improved duty cycles and the absence of the constraining influence from the Earth’s atmosphere. Both the CoRoT [108] and Kepler [109] missions will search for exoplanets using the transit method with high-accuracy photometric measurements for a large number of stars over a substantial period of time. Predictions for transit detections of CoRoT include several tens of “hot Jupiter” planets

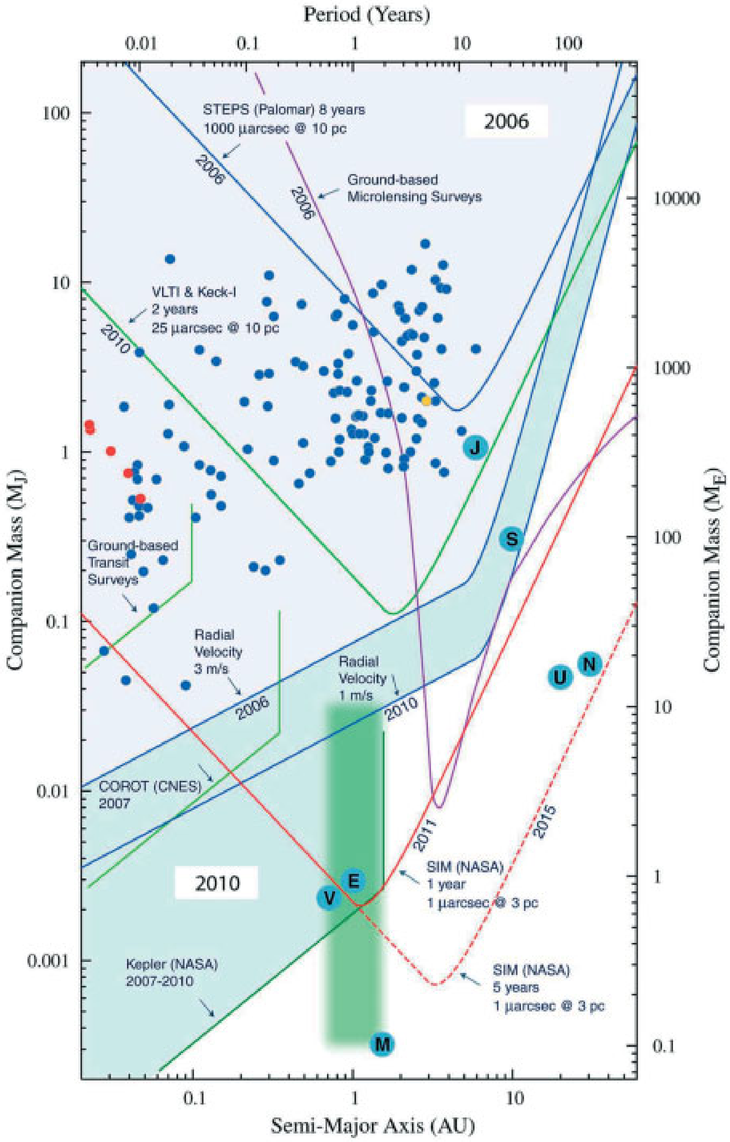


Fig. 9.14 Planetary detection probabilities using the radial velocity and transit methods, together with the potential for space missions like CoRoT and Kepler. In the figure the detection limits in mass and distance from the parent star (Semi-Major Axis/Period) are marked. Note also the location of the planets (blue circles) and the habitable zone (green area) within the Solar System. Figure from the JPL publication "Precursor Science for the Terrestrial Planet Finder" edited by P.R. Lawson, S.C. Unwin, and C.A. Beichman (JPL Pub 04-014, page 21, Fig. 5, 5th Oct 2004). In recent years, the detection of 1 $\mathrm{m\,s^{-1}}$ radial velocity, and even less, became possible, thus expanding the detection range of this method towards even smaller planets than predicted only a few years ago.

and the first "hot Earth" planets [24, 78] while Kepler is expected to reach the region of true Earth-like planets [109]. Consequently, the transit method will play a crucial role in exploring terrestrial planets outside the Solar System and provide the sample of planetary systems to be studied in even greater detail by the next generation space missions.

References

1 Struve, O.: 1952, Proposal for a project of high-precision stellar radial velocity work, *The Observatory*, **72**, 199–200.

2 Rosenblatt, F.: 1971, A two-color photometric method for detection of extra solar planetary systems, *Icarus*, **14**, 71.

3 Borucki, W. J. and Summers, A. L.: 1984, The photometric method of detecting other planetary systems, *Icarus*, **58**, 121–134.

4 Deeg, H. J., Doyle, L. R., Kozhevnikov, V. P., Martin, E. L., Oetiker, B., Palaiologou, E., Schneider, J., Afonso, C., Dunham, E. W., Jenkins, J. M., Ninkov, Z., Stone, R. P. S., and Zakharova, P. E.: 1998, Near-term detectability of terrestrial extrasolar planets: TEP network observations of CM Draconis, *A&A*, **338**, 479–490.

5 Mazeh, T., Naef, D., Torres, G., Latham, D. W., Mayor, M., Beuzit, J.-L., Brown, T. M., Buchhave, L., Burnet, M., Carney, B. W., Charbonneau, D., Drukier, G. A., Laird, J. B., Pepe, F., Perrier, C., Queloz, D., Santos, N. C., Sivan, J.-P., Udry, S., and Zucker, S: 2000, The spectroscopic orbit of the planetary companion transiting HD 209458, *ApJ L*, **532**, L55–L58.

6 Charbonneau, D., Brown, T. M., Latham, D. W., and Mayor, M.: 2000, Detection of planetary transits across a sun-like star, *ApJ L*, **529**, L45–L48.

7 Henry, G. W., Marcy, G. W., Butler, R. P., and Vogt, S. S.: 2000, A transiting 51 Peg-like planet, *ApJ L*, **529**, L41–L44.

8 Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W., and Burrows, A.: 2001, Hubble space telescope time-series photometry of the transiting planet of HD 209458, *ApJ*, **552**, 699–709.

9 Horne, K.: 2003, Status and Prospects of Planetary Transit Searches: Hot Jupiters Galore, *book ASP Conf. Ser. 294: Scientific Frontiers in Research on Extrasolar Planets*, 361–370.

10 Konacki, M., Torres, G., Sasselov, D. D., and Jha, S.: 2003, High-resolution spectroscopic follow-up of OGLE planetary transit candidates in the galactic bulge: Two possible Jupiter-mass planets and two blends, *ApJ*, **597**, 1076–1091.

11 Udalski, A., Szewczyk, O., Zebrun, K., Pietrzynski, G., Szymanski, M., Kubiak, M., Soszynski, I., and Wyrzykowski, L.: 2002, The optical gravitational lensing experiment. Planetary and low-luminosity object transits in the Carina fields of the galactic disk, *Acta Astronomica*, **52**, 317–359.

12 Udalski, A., Zebrun, K., Szymanski, M., Kubiak, M., Soszynski, I., Szewczyk, O., Wyrzykowski, L., and Pietrzynski, G.: 2002, The optical gravitational lensing experiment. Search for planetary and low-luminosity object transits in the galactic disk. Results of 2001 campaign – supplement, *Acta Astronomica*, **52**, 115–128.

13 Udalski, A., Paczynski, B., Zebrun, K., Szymanski, M., Kubiak, M., Soszynski, I., Szewczyk, O., Wyrzykowski, L., and Pietrzynski, G.: 2002, The Optical Gravitational Lensing Experiment. Search for Planetary and Low-Luminosity Object Transits in the Galactic Disk. Results of 2001 Campaign, *Acta Astronomica*, **52**, 1–37.

14 Udalski, A., Pietrzynski, G., Szymanski, M., Kubiak, M., Zebrun, K., Soszynski, I., Szewczyk, O., and Wyrzykowski, L.: 2003, The optical gravitational lensing experiment. Additional planetary and low-luminosity object transits from the OGLE 2001 and 2002 observational campaigns, *Acta Astronomica*, **53**, 133–149.

15 Udalski, A., Szymanski, M. K., Kubiak, M., Pietrzynski, G., Soszynski, I., Zebrun, K., Szewczyk, O., and Wyrzykowski, L.: 2004, The optical gravitational lensing experiment. Planetary and low-luminosity object transits in the fields of galactic disk.

Results of the 2003 OGLE observing campaigns, *Acta Astronomica*, **54**, 313–345.

16 Bakos, G. A., Noyes, R. W., Kovacs, G., Latham, D. W., Sasselov, D. D., Torres, G., Fischer, D. A., Stefanik, R. P., Sato, B., Johnson, J. A., Pal, A., Marcy, G. W., Butler, R. P., Esquerdo, G. A., Stanek, K. Z., Lazar, J., Papp, I., Sari, P., and Sipocz, B.: 2006, HAT-P-1b: A large-radius, low-density exoplanet transiting one member of a stellar binary, 2007 *ApJ*, **656**, 552–559.

17 Alonso, R., Brown, T. M., Torres, G., Latham, D. W., Sozzetti, A., Mandushev, G., Belmonte, J. A., Charbonneau, D., Deeg, H. J., Dunham, E. W., O'Donovan, F. T., and Stefanik, R. P.: 2004, TrES-1: The transiting planet of a bright K0 V star, *ApJ L*, **613**, L153–L156.

18 O'Donovan, F. T., Charbonneau, D., Mandushev, G., Dunham, E. W., Latham, D. W., Torres, G., Sozzetti, A., Brown, T. M., Trauger, J. T., Belmonte, J. A., and Rabus, M., Almenara, J. M., Alonso, R., Deeg, H. J., Esquerdo, G. A., Falco, E. E., Hillenbrand, L. A., Roussanova, A., Stefanik, R. P., and Winn, J. N.: 2006, TrES-2: The first transiting planet in the Kepler field, *ApJ L*, **651**, L61–L64.

19 Collier Cameron, A., Bouchy, F., Hebrard, G., Maxted, P., Pollacco, D., Pont, F., Skillen, I., Smalley, B., Street, R. A., West, R. G., Wilson, D. M., Aigrain, S., Christian, D. J., Clarkson, W. I., Enoch, B., Evans, A., Fitzsimmons, A., Fleenor, M., Gillon, M., Haswell, C. A., Hebb, L., Hellier, C., Hodgkin, S. T., Horne, K., Irwin, J., Kane, S. R., Keenan, F. P., Loeillet, B., Lister, T. A., Mayor, M., Moutou, C., Norton, A. J., Osborne, J., Parley, N., Queloz, D., Ryans, R., Triaud, A. H. M. J., Udry, S., and Wheatley, P. J.: 2006, WASP-1b and WASP-2b: Two new transiting exoplanets detected with SuperWASP and SOPHIE, 2007 *MNRAS*, **375** 951–957.

20 McCullough, P. R., Stys, J. E., Valenti, J. A., Johns-Krull, C. M., Janes, K. A., Heasley, J. N., Bye, B. A., Dodd, C., Fleming, S. W., Pinnick, A., Bissinger, R., Gary, B. L., Howell, P. J., and Vanmunster, T.: 2006, A transiting planet of a sun-like star, *ApJ*, **648**, 1228–1238.

21 Fischer, D. A., Laughlin, G., Butler, P., Marcy, G., Johnson, J., Henry, G., Valenti, J., Vogt, S., Ammons, M., Robinson, S., Spear, G., Strader, J., Driscoll, P., Fuller, A., Johnson, T., Manrao, E., McCarthy, C., Muñoz, M., Tah, K. L., Wright, J., Ida, S., Sato, B., Toyota, E., and Minniti, D.: 2005, The N2K Consortium. I. A hot Saturn planet orbiting HD 88133, *ApJ*, **620**, 481–486.

22 Sato, B., Fischer, D. A., Henry, G. W., Laughlin, G., Butler, R. P., Marcy, G. W., Vogt, S. S., Bodenheimer, P., Ida, S., Toyota, E., Wolf, A., Valenti, J. A., Boyd, L. J., Johnson, J. A., Wright, J. T., Ammons, M., Robinson, S., Strader, J., McCarthy, C., Tah, K. L., and Minniti, D.: 2005, The N2K consortium. II. A transiting hot Saturn around HD 149026 with a large dense core, *ApJ*, **633**, 465–473.

23 Bouchy, F., Udry, S., Mayor, M., Moutou, C., Pont, F., Iribarne, N., da Silva, R., Ilovaisky, S., Queloz, D., Santos, N. C., Ségransan, D., and Zucker, S.: 2005, ELODIE metallicity-biased search for transiting Hot Jupiters. II. A very hot Jupiter transiting the bright K star HD 189733, *A&A*, **444**, L15–L19.

24 Bordé, P., Rouan, D., and Léger, A.: 2001, Exo-planet detection with the COROT space mission. I. A multi-transit detection criterion, *Academie des Sciences Paris Comptes Rendus Serie Physique Astrophysique*, **7**, 1049–1055.

25 Borucki, W. J., Caldwell, D., Koch, D. G., Webster, L. D., Jenkins, J. M., Ninkov, Z., and Showen, R.: 2001, The Vulcan photometer: A dedicated photometer for extrasolar planet searches, *PASP*, **113**, 439–451.

26 Pepper, J., Gould, A., and Depoy, D. L.: 2004, KELT: The Kilodegree Extremely Little Telescope, *book AIP Conf. Proc. 713: The Search for Other Worlds*, 185–188.

27 Rauer, H., Eislöffel, J., Erikson, A., Guenther, E., Hatzes, A. P., Michaelis, H., and Voss, H.,: 2004, The Berlin exoplanet search telescope system, *PASP*, **116**, 38–45.

28 Mallén-Ornelas, G., Seager, S., Yee, H. K. C., Minniti, D., Gladders, M. D., Mallén-Fullerton, G. M., and Brown, T. M.: 2003, The EXPLORE Project. I. A deep search for transiting extrasolar planets, *ApJ*, **582**, 1123–1140.

29 Sahu, K. C., Casertano, S., Bond, H. E., Valenti, J., Ed Smith, T., Minniti, D., Zoccali, M., Livio, M., Panagia, N., Piskunov, N., Brown, T. M., Brown, T., Renzini, A., Rich, R. M., Clarkson, W., and Lubow, S.: 2006, Transiting extrasolar planetary candidates in the Galactic bulge, *Nature*, **443**, 534–540.

30 Gilliland, R. L., Brown, T. M., Guhathakurta, P., Sarajedini, A., Milone, E. F., Albrow, M. D., Baliber, N. R., Bruntt, H., Burrows, A., Charbonneau, D., Choi, P., Cochran, W. D., Edmonds, P. D., Frandsen, S., Howell, J. H., Lin, D. N. C., Marcy, G. W., Mayor, M., Naef, D., Sigurdsson, S., Stagg, C. R., Vandenberg, D. A., Vogt, S. S., and Williams, M. D.: 2000, A Lack of planets in 47 Tucanae from a hubble space telescope search, *ApJ L*, **545**, L47–L51.

31 Street, R. A., Horne, K., Lister, T. A., Penny, A. J., Tsapras, Y., Quirrenbach, A., Safizadeh, N., Mitchell, D., Cooke, J., and Cameron, A. C.: 2003, Searching for planetary transits in the field of open cluster NGC 6819 - I, *MNRAS*, **340**, 1287–1297.

32 von Braun, K., Lee, B. L., Seager, S., Yee, H. K. C., Mallén-Ornelas, G., and Gladders, M. D.: 2005, Searching for planetary transits in galactic open clusters: EXPLORE/OC, *PASP*, **117**, 141–159.

33 Hood, B., Cameron, A. C., Kane, S. R., Bramich, D. M., Horne, K., Street, R. A., Bond, I. A., Penny, A. J., Tsapras, Y., Quirrenbach, A., Safizadeh, N., Mitchell, D., and Cooke, J.: 2005, A dearth of planetary transits in the direction of NGC 6940, *MNRAS*, **360**, 791–800.

34 Bramich, D. M., Horne, K., Bond, I. A., Street, R. A., Cameron, A. C., Hood, B., Cooke, J., James, D., Lister, T. A., Mitchell, D., Pearson, K., Penny, A., Quirrenbach, A., Safizadeh, N., and Tsapras, Y.: 2005, A survey for planetary transits in the field of NGC 7789, *MNRAS*, **359**, 1096–1116.

35 Marshall, J. L., Burke, C. J., DePoy, D. L., Gould, A., and Kollmeier, J. A.: 2005, Survey for transiting extrasolar planets in stellar systems. II. Spectrophotometry and Metallicities of open clusters, *AJ*, **130**, 1916–1928.

36 Burke, C. J., Gaudi, B. S., DePoy, D. L., and Pogge, R.W.:2006, Survey for transiting extrasolar planets in stellar systems. III. A limit on the fraction of stars with planets in the open cluster NGC 1245, *AJ*, **132**, 210–230.

37 Aigrain, S., Hodgkin, S., Irwin, J., Hebb, L., Irwin, M., Favata, F., Moraux, E., and Pont, F.: 2006, The Monitor project: Searching for occultations in young open clusters, 2007 *ASP*, **366**, 33.

38 Mochejska, B. J., Stanek, K. Z., Sasselov, D. D., Szentgyorgyi, A. H., Bakos, G. Á., Hradecky, J., Devor, V., Marrone, D. P., Winn, J. N., and Zaldarriaga, M.: 2005, Planets in stellar clusters extensive search. III. A search for transiting planets in the metal-rich open cluster NGC 6791, *AJ*, **129**, 2856–2868.

39 Mochejska, B. J., Stanek, K. Z., Sasselov, D. D., Szentgyorgyi, A. H., Adams, E., Cooper, R. L., Foster, J. B., Hartman, J. D., Hickox, R. C., Lai, K., Westover, M., and Winn, J. N.: 2006, Planets in Stellar Clusters Extensive Search. IV. A detection of a possible transiting planet candidate in the open cluster NGC 2158, *AJ*, **131**, 1090–1105.

40 Konacki, M., Torres, G., Sasselov, D. D., and Jha, S.: 2005, A transiting extrasolar giant planet around the star OGLE-TR-10, *ApJ*, **624**, 372–377.

41 Pont, F., Bouchy, F., Queloz, D., Santos, N. C., Melo, C., Mayor, M., and Udry, S.: 2004, The missing link: A 4-day period transiting exoplanet around OGLE-TR-111, *A&A*, **426**, L15–L18.

42 Bouchy, F., Pont, F., Santos, N. C., Melo, C., Mayor, M., Queloz, D., and Udry, S.: 2004, Two new "very hot Jupiters" among the OGLE transiting candidates, *A&A*, **421**, L13–L16.

43 Charbonneau, D., Winn, J. N., Everett, M. E., Latham, D. W., Holman, M. J., Esquerdo, G. A., and O'Donovan, F. T.: 2006,, Precise radius estimates for the exoplanets WASP-1b and WASP-2b, 2007 *AJ*, **658**, 1322–1327 .

44 Winn, J. N., Holman, M. J., Henry, G. W., Roussanova, A., Enya, K., Yoshii, Y., Shporer, A., Mazeh, T., Johnson, J. A., Narita, N., and Suto, Y.: 2006, The transit light curve project. V. System parameters and stellar rotation period of HD 189733, 2007 *AJ*, **133** 1828–1835.

45 Knutson, H. A., Charbonneau, D., Noyes, R. W., Brown, T. M. and Gilliland, R. L.: 2007, Using Stellar Limb-Darkening to Refine the Properties of HD 209458b, *ApJ*, **655**, 564–575.

46 Laughlin, G., Marcy, G. W., Vogt, S. S., Fischer, D. A., and Butler, R. P.: 2005, On the eccentricity of HD 209458b, *ApJ L*, **629**, L121–L124.

47 Winn, J. N., Noyes, R. W., Holman, M. J., Charbonneau, D., Ohta, Y., Taruya, A., Suto, Y., Narita, N., Turner, E. L., Johnson, J. A., Marcy, G. W., Butler, R. P., and Vogt, S. S.: 2005, Measurement of spin-orbit alignment in an extrasolar planetary system, *ApJ*, **631**, 1215–1226.

48 Pont, F., Zucker, S., and Queloz, D.: 2006, The effect of red noise on planetary transit detection, *MNRAS*, **373**, 231–242.

49 Santos, N. C., Pont, F., Melo, C., Israelian, G., Bouchy, F., Mayor, M., Moutou, C., Queloz, D., Udry, S., and Guillot, T.: 2006, High resolution spectroscopy of stars with transiting planets. The cases of OGLE-TR-10, 56, 111, 113, and TrES-1, *A&A*, **450**, 825–831.

50 Winn, J. N., Holman, M. J., and Fuentes, C. I.: 2007, The transit light curve project. II. Two transits of the exoplanet OGLE-TR-111b, *AJ*, **133**, 11–16.

51 Gillon, M., Pont, F., Moutou, C., Bouchy, F., Courbin, F., Sohy, S., and Magain, P.: 2006, High accuracy transit photometry of the planet OGLE-TR-113b with a new deconvolution-based method, *A&A*, **459**, 249–255.

52 Charbonneau, D., Brown, T. M., Burrows, A., and Laughlin, G.: 2007, When extrasolar planets transit their parent stars, Protostars and Planets Vp. 701–716.

53 Shporer, A., Tamuz, O., Zucker, S., and Mazeh, T.: 2006, Photometric follow-up of the transiting planet around WASP-1, *astro-ph/0610556*.

54 Holman, M. J., Winn, J. N., Latham, D. W., O'Donovan, F. T., Charbonneau, D., Bakos, G. A., Esquerdo, G. A., Hergenrother, C., Everett, M. E., and Pál, A.: 2006, The transit light curve project. I. Four consecutive transits of the exoplanet XO-1b, *ApJ*, **652**, 1715–1723.

55 Mazeh, T., Zucker, S., and Pont, F.: 2005, An intriguing correlation between the masses and periods of the transiting planets, *MNRAS*, **356**, 955–957.

56 Baraffe, I., Selsis, F., Chabrier, G., Barman, T. S., Allard, F., Hauschildt, P. H., and Lammer, H.: 2004, The effect of evaporation on the evolution of close-in giant planets, *A&A*, **419**, L13–L16.

57 Tingley, B. and Sackett, P. D.: 2005, A photometric diagnostic to aid in the identification of transiting extrasolar planets, *ApJ*, **627**, 1011–1018.

58 Mandel, K. and Agol, E.: 2002, Analytic light curves for planetary transit searches, *ApJ L*, **580**, L171–L175.

59 Girardi, L., Bertelli, G., Bressan, A., Chiosi, C., Groenewegen, M. A. T., Marigo, P., Salasnich, B., and Weiss, A.: 2002, Theoretical isochrones in several photometric systems. I. Johnson–Cousins–Glass, HST/WFPC2, HST/NICMOS, Washington, and ESO imaging survey filter sets, *A&A*, **391**, 195–212.

60 Voss, H.: 2006, Developing a ground-based search system for transits of extrasolar planets, *TU Berlin*, **PhD**.

61 Kovács, G., Zucker, S., and Mazeh, T.: 2002, A box-fitting algorithm in the search for periodic transits, *A&A*, **391**, 369–377.

62 Defaÿ, C., Deleuil M., and Barge, P.: 2001, A Bayesian method for the detection of planetary transits, *A&A*, **365**, 330–340.

63 Tingley, B.: 2003, A rigorous comparison of different planet detection algorithms, *A&A*, **403**, 329–337.

64 Tingley, B.: 2003, Improvements to existing transit detection algorithms and their comparison, *A&A*, **408**, L5–L7.

65 Moutou, C., Pont, F., Barge, P., Aigrain, S., Auvergne, M., Blouin, D., Cautain, R., Erikson, A. R., Guis, V., Guterman, P., Irwin, M., Lanza, A. F., Queloz, D., Rauer, H., Voss, H., and Zucker, S.: 2005, Comparative blind test of five planetary transit detection algorithms on realistic synthetic light curves, *A&A*, **437**, 355–368.

66 Brown, T. M.: 2003, Expected detection and false alarm rates for transiting Jovian planets, *ApJ L*, **593**, L125–L128.

67 Shobbrook, R. R., Herbison-Evans, D., Johnston, I. D., and Lomb, N. R.: 1969, Light variations in Spica, *MNRAS*, **145**, 131.

68 Drake, A. J.: 2003, On the selection of photometric planetary transits, *ApJ*, **589**, 1020–1026.

69 Sirko, E. and Paczyński, B.: 2003, Ellipsoidal variability in the OGLE planetary transit candidates, *ApJ*, **592**, 1217–1224.

70 Seager, S. and Mallén-Ornelas, G.: 2003, A unique solution of planet and star parameters from an extrasolar planet transit light curve, *ApJ*, **585**, 1038–1055.

71 Pont, F., Melo, C. H. F., Bouchy, F., Udry, S., Queloz, D., Mayor, M., and Santos, N. C.: 2005, A planet-sized transiting star around OGLE-TR-122. Accurate mass and radius near the hydrogen-burning limit, *A&A*, **433**, L21–L24.

72 Young, A. T.: 1993, Scintillation noise in CCD photometry, *The Observatory*, **113**, 41–42.

73 Tamuz, O., Mazeh, T., and Zucker, S.: 2005, Correcting systematic effects in a large set of photometric light curves, *MNRAS*, **356**, 1466–1470.

74 Lineweaver, C. H. and Grether, D.: 2003, What fraction of sun-like stars have planets?, *ApJ*, **598**, 1350–1360.

75 Christian, D. J., Pollacco, D. L., Skillen, I., Street, R. A., Keenan, F. P., Clarkson, W. I., Collier Cameron, A., Kane, S. R., Lister, T. A., West, R. G., Enoch, B., Evans, A., Fitzsimmons, A., Haswell, C. A., Hellier, C., Hodgkin, S. T., Horne, K., Irwin, J., Norton, A. J., Osborne, J., Ryans, R., Wheatley, P. J., and Wilson, D. M.: 2006, The SuperWASP wide-field exoplanetary transit survey: candidates from fields 23 h $<$ RA $<$ 03 h, *MNRAS*, **372**, 1117–1128.

76 Santos, N. C., Israelian, G., and Mayor, M.: 2003, Confirming the metal-rich nature of stars with giant planets, **book** *The Future of Cool-Star Astrophysics: 12th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun(2001 July 30 – August 3)*, 148–157.

77 Fischer, D. A. and Valenti, J.: The planet–metallicity correlation, *ApJ*, **622**, 1102–1117.

78 Bordé, P., Rouan, D., and Léger, A.: 2003, Exoplanet detection capability of the COROT space mission, *A&A*, **405**, 1137–1144.

79 Deeg, H. J., Garrido, R., and Claret, A.: 2001, Probing the stellar surface of HD 209458 from multicolor transit observations, *New Astronomy*, **6**, 51–60.

80 Rossiter, R. A.: 1924, On the detection of an effect of rotation during eclipse in the velocity of the brighter component of beta Lyrae, and on the constancy of velocity of this system., *ApJ*, **60**, 15–21.

81 McLaughlin, D. B.: 1924, On the long-period velocity variation and the absolute scale of the system of Algol(abstract), *Popular Astronomy*, **32**, 225.

82 Ohta, Y., Taruya, A., and Suto, Y.: 2005, The Rossiter–McLaughlin effect and analytic radial velocity curves for transiting extrasolar planetary systems, *ApJ*, **622**, 1118–1135.

83 Gaudi, B. S. and Winn, J. N.: 2007, Prospects for the characterization and confirmation of transiting exoplanets via the Rossiter-McLaughlin effect, *ApJ*, **655**, 550–563.

84 Sartoretti, P. and Schneider, J.: 1999, On the detection of satellites of extrasolar planets with the method of transits, *A&A*, **134**, 553–560.

85 Miralda-Escudé, J.: 2002, Orbital perturbations of transiting planets: A possible method to measure stellar quadrupoles and to detect earth-mass planets, *ApJ*, **564**, 1019–1023.

86 Agol, E., Steffen, J., Sari, R., and Clarkson, W.: 2005, On detecting terrestrial planets with timing of giant planet transits, *MNRAS*, **359**, 567–579.

87 Collier Cameron, A., Horne, K., Penny, A., and Leigh, C.: 2002, A search for starlight reflected from ypsilon And's innermost planet, *MNRAS*, **330**, 187–204.

88 Charbonneau, D., Noyes, R. W., Korzennik, S. G., Nisenson, P., Jha, S., Vogt, S. S., and Kibrick, R. I.: 1999, An upper limit on the reflected light from the planet orbiting the star tau Bootis, *ApJ L*, **522**, L145–L148.

89 Rowe, J. F., Matthews, J. M., Seager, S., Kuschnig, R., Guenther, D. B., Moffat, A. F. J., Rucinski, S. M., Sasselov, D., Walker, G. A. H., and Weiss, W. W.: 2006, An upper limit on the albedo of HD 209458b: Direct imaging photometry with the MOST satellite, *ApJ*, **646**, 1241–1251.

90 Rauer, H., Harris, A., Collier Cameron, A., and Export Team: 2000, Search for Signatures of Atmospheres/Exospheres in Extra-solar Planets, *IAU Symposium 109*.

91 Rauer, H., Bockelée-Morvan, D., Coustenis, A., Guillot, T., and Schneider, J.: 2000, Search for an exosphere around 51 Pegasi B with ISO, *A&A*, **355**, 573–580.

92 Moutou, C., Coustenis, A., Schneider, J., St Gilles, R., Mayor, M., Queloz, D., and Kaufer, A.: 2001, Search for spectroscopical signatures of transiting HD 209458b's exosphere, *A&A*, **371**, 260–266.

93 Moutou, C., Coustenis, A., Iro, N., Schneider, J., Mayor, M., and Queloz, D.: 2003, VLT Spectroscopy of HD209458b Transits, *book ASP Conf. Ser. 294: Scientific Frontiers in Research on Extrasolar Planets*, 475–478.

94 Seager, S. and Sasselov, D. D.: 2000, Theoretical transmission spectra during extrasolar giant planet transits, *ApJ*, **537**, 916–921.

95 Hubbard, W. B., Fortney, J. J., Lunine, J. I., Burrows, A., Sudarsky, D., and Pinto, P.: 2001, Theory of extrasolar giant planet transits, *ApJ*, **560**, 413–419.

96 Brown, T. M.: 2001, Transmission spectra as diagnostics of extrasolar giant planet atmospheres, *ApJ*, **553**, 1006–1026.

97 Brown, T. M., Butler, R. P., Charbonneau, D., Noyes, R. W., Sasselov, D., Libbrecht, K. G., Marcy, G. W., Seager, S., and Vogt, S. S.: 2000, Seeking the atmospheric transmission spectrum of HD209458b, *Bulletin of the American Astronomical Society*.

98 Charbonneau, D.: 2002, Characterizing the Atmosphere of an Extrasolar Planet, *book: HST Proposal*, 5794.

99 Vidal-Madjar, A.: 2004, Physical Parameters of the Upper Atmosphere of the Extrasolar Planet HD209458b, **book** *HST Proposal*, 6463.

100 Vidal-Madjar, A., Désert, J.-M., Lecavelier des Etangs, A., Hébrard, G., Ballester, G. E., Ehrenreich, D., Ferlet, R., McConnell, J. C., Mayor, M., and Parkinson, C. D.: 2004, Detection of oxygen and carbon in the hydrodynamically escaping atmosphere of the extrasolar planet HD 209458b, *ApJ L*, **604**, L69–L72.

101 Vidal-Madjar, A., Lecavelier des Etangs, A., Désert, J.-M., Ballester, G. E., Ferlet, R., Hébrard, G., and Mayor, M.: 2003, An extended upper atmosphere around the extrasolar planet HD209458b, *Nature*, **422**, 143–146.

102 Moutou, C., Coustenis, A., Schneider, J., Queloz, D., and Mayor, M.: 2003, Searching for helium in the exosphere of HD 209458b, *A&A*, **405**, 341–348.

103 Winn, J. N., Suto, Y., Turner, E. L., Narita, N., Frye, B. L., Aoki, W., Sato, B., and Yamada, T.: 2004, A search for Hα absorption in the exosphere of the transiting extrasolar planet HD 209458b, *PASJ*, **56**, 655–662.

104 Narita, N., Suto, Y., Winn, J. N., Turner, E. L., Aoki, W., Leigh, C. J., Sato, B., Tamura, M., and Yamada, T.: 2005, Errata: Subaru HDS transmission spectroscopy of the transiting extrasolar planet HD 209458b, *PASJ*, **57**, 705–705.

105 Charbonneau, D., Allen, L. E., Megeath, S. T., Torres, G., Alonso, R., Brown, T. M., Gilliland, R. L., Latham, D. W., Mandushev, G., O'Donovan, F. T., and Sozzetti, A.: 2005, Detection of thermal emission from an extrasolar planet, *ApJ*, **626**, 523–529.

106 Deming, D., Brown, T. M., Charbonneau, D., Harrington, J., and Richardson, L. J.: 2005, A new search for carbon monoxide absorption in the transmission spectrum of the extrasolar planet HD 209458b, *ApJ*, **622**, 1149–1159.

107 Deming, D., Harrington, J., Seager, S., and Richardson, L. J.: 2006, Strong infrared emission from the extrasolar planet HD 189733b, *ApJ*, **644**, 560–564.

108 Baglin, A.: 2003, COROT: A minisat for pionnier science, asteroseismology and planets finding, *Advances in Space Research*, **31**, 345–349.

109 Borucki, W. J., Koch, D., Basri, G., Brown, T., Caldwell, D., Devore, E., Dunham, E., Gautier, T., Geary, J., Gilliland, R., Gould, A., Howell, S., and Jenkins, J.: 2003, Kepler Mission: a mission to find earth-size planets in the habitable zone, *book ESA SP-539: Earths: DARWIN/TPF and the Search for Extrasolar Terrestrial Planets*, 69–81.

10
US Programs and Space Missions for Extrasolar Planet Research

P. Kenneth Seidelmann

Abstract

The search for extrasolar planets and planet-forming regions is the focus of a number of programs, including the NASA Navigator Program. The ultimate goal of these programs is finding earth-like planets to determine if they are habitable, and search for signs of life. This effort includes both ground and space-based missions, and techniques of astrometry, interferometry, photometry, nulling, and coronographic imaging in both visual and infrared wavelengths to detect transits, periodic motions, and direct images. The space missions include Kepler, SIM PlanetQuest, Terrestrial Planet Finder – Coronograph and Infrared, Spitzer Space Telescope, James Webb Space Telescope (JWST), and Microvariability and Oscillations of Stars (MOST). Stratospheric Observatory for Infrared Astronomy (SOFIA) is a high-flying airplane, and ground-based observations are made by the Keck Interferometer, the Large Binocular Telescope Interferometer (LBTI), Palomar Observatory Instruments, Pan-STARRS, and proposed Thirty Meter Class Telescopes.

10.1
Introduction

How many stars have planets around them? How common are solar systems in the universe? How do solar systems form? What are the masses and distributions of the planets? Is there a brown dwarf desert? Are there other habitable planets than the Earth? What are the chemical compositions of the planets and atmospheres? Can vegetation be detected? Is there other intelligent life in the universe? These are some of the questions that the missions and programs to discover extrasolar planets are seeking to answer.

The basic problem is that extrasolar planets are an enormous distance from us, they do not produce any light themselves, and they are near very bright stars. Hence, the techniques for discovering extrasolar planets are transits of the planets across or behind a star, periodic variations in the radial velocities of

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

stars, periodic variations in the proper motions of stars, micro-lensing observations indicating the presence of planetary size objects, and direct imaging of a planet moving around a star, either by nulling or coronographic techniques.

Radial velocity observations from the ground have been very successful in discovering planets within the accuracy and time periods of the observational data. Some ground-based transit observations have been made with small successes so far. Also micro-lensing detections of planets have been announced. Astrometric observations have not yet reached the accuracies necessary to detect planetary size objects.

The astrometric technique is complementary to the radial-velocity method. Astrometry is the most sensitive method for detecting planets in larger, slower orbits, while radial velocity is most sensitive for detecting planets in smaller, faster orbits. The astrometric technique is also important because it provides the planet mass without ambiguity attributable to the orbit inclination.

The search for habitable planets and life has strong support within the scientific community and is of great interest to the general public. The search has been advocated by NASA and National Academy of Science advisory committees, including reports from the TOPS (Toward Other Planetary Systems) and COMPLEX (Committee on Planetary Exploration) committees. The concept for the Keck Interferometer was described in the TOPS report, and restated in A Road Map for the Exploration of Neighboring Planetary Systems (“The ExNPS Report”, 1996). The TOPS and ExNPS reports recognized the unique potential of interferometry for the problem of planet detection. The TOPS report specifically identified the detection of "Hot Jupiters" using the differential phase technique. The ExNPS report identified the need to measure the brightness of the clouds of zodiacal dust around other stars. Since this emission can, in extreme cases, mask the signal of individual planets, its characterization is important to the design of NASA’s Terrestrial Planet Finder (TPF) space mission. Both the TOPS and ExNPS reports noted the capability of ground-based interferometry to detect Uranus-size planets through the astrometric signature of their parent stars.

The 2001 decadal review of astronomy and astrophysics, "Astronomy and Astrophysics in the New Millennium", prepared by the US National Research Council stated, "The discovery of life on another planet is potentially one of the most important scientific advances of this century, let alone this decade, and it would have enormous philosophical implications". The decadal committee recognized the importance and challenges of building TPF and recommended proceeding with TPF, but requested that technology must be demonstrated first and precursor missions should confirm the expectation that terrestrial planets are common around solar-type stars. The committee also recommended that TPF give equal attention to the broader science goals of studying planetary systems and the structure of astronomical sources imaged at high angular resolution.

There is a NASA Navigator Program with the goals of finding earth-like planets around nearby stars to determine if they are habitable, and search for signs of life. Hence, there are planned space missions to detect planets by transit observations, Kepler, and to detect planets by interferometry, SIM PlanetQuest. Terrestrial Planet Finder - Coronagraph (TPF-C) is to seek to detect and characterize planets that may be habitable. Terrestrial Planet Finder - Infrared (TPF-I) is to detect and characterize planets from infrared emission.

There is a trade-off between the capabilities of telescopes on the ground and in space. Ground based telescopes can be bigger and cheaper, but they are restricted by the atmosphere. However, adaptive optics and advanced detector instrumentation can compensate for some of these limitations. Light collection of more than an order of magnitude larger than HST can be obtained from the ground. So, for optical and infrared observations, new advanced mountaintop telescopes offer potential for future discovery, at a much smaller cost. SOFIA offers the infrared observing capability from an airborne platform above the atmosphere.

NASA is currently struggling with budget problems, which are leading to serious changes in program plans. So all the future schedules and launch dates should be considered as uncertain. Right now flight plans for SOFIA and the development of TPF are on hold. All NASA programs are suffering budget reductions that might reduce capabilities and delay schedules.

10.2
Kepler

Kepler will measure the depth, duration, and repeatability of the dimming of a star when a planet transits in front of the stellar disk. Kepler should detect hundreds of terrestrial planets in the habitable zone and determine whether terrestrial planets are common or rare. It will also determine the planets' albedos, periods, sizes, masses, and densities (the latter two with radial velocity data).

The Kepler instrument is a 0.95 meter telescope with a 105 square degree field of view. The detectors are 42 CCDs, each with 2200 × 1024 pixels. The photometric one sigma noise for a solar-like stellar variability and photon shot noise for an mv=12 star is $< 2 \times 10^{-5}$. The instrument has no mechanisms except for antenna gimbals and a one-time ejectable cover. The spacecraft will be launched in 2008 into an earth-trailing heliocentric orbit. It will continuously monitor 100 000 main-sequence stars in a single star field in the Cygnus-Lyra region for four years, searching for planets. To observe a transit of a star the planet orbit must be lined up edgewise to our solar system. This is more critical as the planet orbit gets larger, with a probability of proper alignment of the planet orbit of about 0.5%. The photometer in space has the advantage of precision and avoiding interruptions caused by day–night and seasonal cycles and atmospheric perturbations.

If planets are common around stars like our sun, then, depending on the average size of the planets, the expected terrestrial planet detections would be about 50, 185, and 640, if the planet sizes were 1.0, 1.3, or 2.2 times the earth radius, respectively. Two or more planets would be expected in about 12% of the cases. The detections for giant planets would be about 135 and 30 for inner and outer orbits, respectively. The sample size of the stars is large enough that the lack of detection of planets would still be very significant.

For this mission the Science Operations Center is at NASA Ames, the Mission Operations Center is at the University of Colorado LASP, and the Data Management Center is at the Space Telescope Science Center.

10.3
SIM PlanetQuest

Space Interferometry Mission (SIM) PlanetQuest measures stellar positions with respect to neighboring stars over the full sky over a period of years by means of interferometers. It will detect planets from the astrometric periodic motion of the stars. It will determine the planet's period, mass, orbital distance, shape and orientation. The mission will also determine the distances to stars throughout the galaxy and perform other fundamental astrophysics investigations.

In a narrow field of view (one degree), SIM will aim to achieve an accuracy of one microarcsecond in a single measurement and search for planetary companions to nearby stars by detecting the astrometric wobble relative to nearby stars. From wide-angle measurements (15 degrees) SIM has the precision goal of 4 microarcseconds over the five-year mission. The limiting magnitude is 20. These observations will be made relative to an astrometric grid of reference stars covering the whole sky. The astrometric grid will be made up of overlapping tiles of 15 degrees, observed by repeated relative observations. The 15 degree tiles are the area of the sky accessible to SIM by repointing the optics without changing the orientation of the spacecraft. Pointing the spacecraft will be accomplished using reaction wheels, with small thrusters used to desaturate the reaction wheels. To protect the optics from overheating, the nominal viewing axis will never be within 45 degrees of the sun.

SIM is a space-based 10 m baseline optical Michelson interferometer operating in the visible waveband (0.4–0.9 μm). The telescopes have 0.3 m diameter apertures, the detectors are CCDs, and the orbit is an earth-trailing solar orbit after a launch in 2011. The spacecraft will slowly drift away from the earth at a rate of about 0.1 AU per year, reaching a distance of about 95 million kilometers after 5.5 years. In this orbit the spacecraft will receive continuous solar illumination. The spacecraft velocity needs to be determined to an accuracy of 20 $\mathrm{mm\,s^{-1}}$, or better, to correct for relativistic stellar aberration. This will be achieved using ranging and Doppler data obtained from the Deep Space Network (DSN) ground stations. Normally, observation sequences will be uplinked every week or two. Observational data will be stored onboard and transmitted to earth several times a week. Special quick-turnaround procedures will be used to handle targets of opportunity.

In 2005 SIM PlanetQuest successfully completed eight technological milestones, which were rigorous tests of the ultra-precise technologies, required by NASA to prove that the mission could accomplish its scientific goals.

10.4
Terrestrial Planet Finder – Coronograph (TPF-C)

TPF-C uses a single large telescope with a coronograph to directly detect and characterize planets from a visible wavelength, direct image of a star–planet system. TPF-C will determine the planet's period, brightness, atmospheric gases (oxygen, water vapor, ozone, carbon dioxide, methane), and orbital distance, shape, and orientation. In good conditions, TPF-C will detect the planet's length of day, surface plant life, ocean/land ratio, and cloud variability. TPF-C will search for habitability and signs of life on extrasolar planets.

The visible-light concept uses a single telescope with an effective diameter of 6.5–8 meters. It must achieve a billion to one image contrast, so very precise control of the optical quality is required. Wavefront control is a critical issue for TPF. Imperfections in the optics, which scatter light and degrade image contrast, must be corrected. Internal imperfections can be corrected by active optics, similar to the technology used on ground-based telescopes to correct for wavefront distortion in the earth's atmosphere.

The coronograph instrumentation will use a central disc and specialized techniques to block the glare of a star, so that dimmer planets around the star can be detected and characterized. To study the area around a star requires dealing with the diffraction of light around the edges of the telescope, which detracts from the potential angular resolution of the image. The diffraction pattern from a simple round telescope is a series of concentric rings; Aries rings, with a bright central spot. In order to see an orbiting planet it is necessary to suppress several of the bright rings without blocking out the planet. Masks can simulate a telescope with a different shape giving a diffraction pattern such that the starlight is much dimmer, or brighter, closer to the center in some areas. Then by rotating the telescope about the line-of-sight the planet image passes in and out of dim and bright regions. There are a number of options available for managing the diffraction pattern.

Advantages of operating at visible wavelengths include that smaller telescopes can obtain the required resolution, optical detectors require less thermal control, which reduces the need for onboard cooling, and the telescope can operate at room temperatures. Various missions have been proposed to demonstrate the technologies needed for TPF-C, which is to launch in 2014.

10.5
Terrestrial Planet Finder – Infrared (TPF-I)

TPF-I will determine the planet's period, infrared brightness, amounts of life-related atmospheric gases (water vapor, ozone, carbon dioxide, methane, and nitrous oxide), and orbital distance, shape, and orientation. If the planet's atmosphere is transparent enough, TPF-I will determine the planet's length of day, ocean/land ratio, and cloud variability.

One TPF-I concept uses an array of four formation-flying, large, cooled telescopes (3 to 4 meters) flying in precise formation with interferometric nulling beam combination to detect and characterize planets from infrared emissions. Another TPF-I concept is placing the telescopes on a physical structure. SIM will demonstrate spaceborne interferometry and path length control. By combining infrared, or heat radiation, gathered by the multiple telescopes by means of interferometry, much larger telescopes will be simulated.

Interferometers allow smaller, widely separated mirrors to work together as a giant virtual telescope. The resolution obtained would be the same as a single telescope the size of the separation between the individual telescopes. To develop a good picture the interferometer must rotate around its line-of-sight to different relative positions and take repeated exposures. In addition to taking pictures, an interferometer can obtain spectra of the targets. The joint mission with the European Space Agency is to launch in 2020.

10.6
Spitzer Space Telescope

The Spitzer Space Telescope, formerly the Space Infrared Telescope Facility (SIRTF), has a 0.85 meter telescope and three cryogenically cooled science instruments. The three instruments are: an Infrared Array Camera (IRAC), which is a four-channel camera providing 5.12 × 5.12 arcminute simultaneous images in 3.6, 4.5, 5.8, and 8 microns; an Infrared spectrograph (IRS), which has four separate modules giving a low-resolution, short-wavelength mode covering 5.3–14 microns, a high-resolution, short-wavelength mode covering

10–19.5 microns, a low-resolution, long-wavelength mode covering 14–40 microns, and a high resolution, long wavelength mode for 19–37 microns; and a Multiband Imaging Photometer (MIPS), which has a 128 × 128 array for imaging at 24 microns, a 32 × 32 array for imaging at 70 microns and spectra from 50–100 microns, and a 2 × 20 array for imaging at 160 microns. It was launched in 2003 into an earth-trailing orbit with an expected minimum lifetime of 2.5 years. This orbit is far enough away from the earth to allow the telescope to cool rapidly without a large amount of coolant. To observe in the infrared the telescope needs to be cooled to about -273 degrees C, and be protected from the heat of the sun and infrared radiation from the earth.

The Spitzer Space Telescope is an infrared system so it can detect heat radiation in the infrared spectra from planets. For transiting planets, observations can be made when the planet is in front (primary transit) and behind (secondary transit) the star, and the observations subtracted to give the planet spectra itself. The advantages of this approach are that the entire surface of the planet emits infrared radiation and at infrared wavelengths the planet is only about 1000 times fainter than a sun-like host star. Using this technique, the temperature of the planet can be measured and the presence of water vapor, carbon monoxide, and methane can be inferred.

Since hot Jupiters should be tidally locked to the parent stars and have permanent day and night sides, the planets should experience uneven heating. This should lead to very fast winds to redistribute the heat from the day to night sides. Observations of orbital variations could reveal how absorbed energy circulates around the planet.

IRAS has discovered the presence of dust disks around some nearby stars, providing key information about the formation of extrasolar planetary systems. Spitzer can study hundreds of stars to determine the frequency of these disks and can use imaging and spectroscopy to characterize the structure and composition of the disks.

10.7
James Webb Space Telescope (JWST)

JWST is to have a 6.5 meter mirror and a sunshield the size of a tennis court, both of which will be folded for launch and open in orbit. JWST will be launched in 2014 into an L2 Lissajous orbit. JWST will be an infrared optimized large telescope with instruments primarily for the infrared range, but with some visual range capability.

JWST should be able to detect heat emission in the infrared from hot Neptunes and possibly hot earths. It will also study the earliest galaxies and some of the first stars formed after the big bang.

JWST is made up of three major elements; the Integrated Science Instrument Module (ISIM), the Optical Telescope Element (OTE), and the Space Support Module (SSM). The ISIM consists of a cryogenic instrument module, which is integrated with the OTE and science processors, software, and electronics in the warm SSM. The ISIM contains three science instruments, NIRCam, NIRSpec, and MIRI, and the Fine Guidance Sensor (FGS).

The Near Infrared Camera (NIRCam) is designed to detect the first light-emitting objects to form in the universe after the big bang, to study star formation in the Milky Way, and to discover and characterize planets around other stars.

The Near Infrared Spectrograph (NIRSpec) will observe in the wavelength range of 0.6 to 5 microns. It will study galaxy formation, clustering, chemical abundances, star formation and kinematics, active galactic nuclei, young stellar clusters, and measurements of the initial mass function of stars (IMF). The current NIRSpec design provides three observing modes: a low-resolution R 100 resolving power prism mode, a R 1000 multi-object mode, and a R 3000 integral field unit or long-slit spectroscopy mode. The field of view will be 3.4 × 3.4 arcminute in each case. The European Space Agency (ESA) will be providing the NIRSpec instrument (NASA will provide the detectors) as part of their contribution to JWST.

The Mid Infrared Instrument (MIRI) is for imaging and spectroscopy at wavelengths between 5 and 27 microns. MIRI is designed to discover "first light", investigate the history of star formation, growth of black holes, production of heavy elements, how stars and planetary systems form, evolution of planetary systems, and the conditions for life. It is a collaboration between NASA JPL and ESA.

The FGS will provide high-precision pointing-error signals to the observatory Attitude Control Subsystem (ACS) to enable stable pointing at the milliarcsecond level. The FGS will assure that an appropriate guide star is available with 95% probability at any point in the sky. The FGS will have three simultaneous fields of view of about 2.3 × 2.3 arcminutes. Two of those fields will feed guiding channels and the third will feed a R 100 tunable filter camera. The Fine Guidance Sensor (FGS) will be supplied by the Canadian Space Agency (CSA).

10.8
Microvariability and Oscillations of Stars (MOST)

MOST is an optical telescope with a 15 cm mirror, a CCD camera with twin side-by-side Marconi frame-transfer devices (1024 × 1024 pixels). One CCD is used for science observations and the other is read out every second to track guide stars for satellite attitude control. A broadband filter selects light in the 350–700 nm wavelength region. The camera has an array of Fabry microlenses to project a large stable image of the telescope pupil illuminated by the target starlight, which gives the photometric precision. The low-cost instrument has no moving parts. The structure automatically maintains the same focus for a wide range of temperatures. The CCDs are cooled by a passive radiator system, and exposure times are controlled by rapid frame transfer.

MOST was launched in 2003 by a Russian ROCKOT into a low-earth polar, sun synchronous orbit that remains over the terminator of the earth. Thus, it has a Continuous Viewing Zone from −19 to 36 degrees declination, in which a selected target star remains observable for up to 60 days without interruption.

The small Canadian space telescope, MOST, can measure a planet's reflected light from multiple observations of secondary eclipses, when a planet can be observed before, during, and after, an eclipse.

10.9
Stratospheric Observatory for Infrared Astronomy (SOFIA)

SOFIA, NASA's modified Boeing 747 airplane with a 2.5 meter infrared telescope and modern detectors, will make observations that cannot be made from ground-based infrared telescopes. It will be based at Ames Research Center at Moffet Federal Airfield. It is expected to fly into the stratosphere (to approximately 41 000 feet altitude), open the telescope cavity door, and point the telescope at selected targets three or four nights a week for at least twenty years. The scientific instruments are designed to make high-resolution measurement over 0.3 micron – 1.6 mm wavelengths. There are three categories of instrument, Facility Science Instruments (FSI), which are available for any observer, Principal Investigator Instruments (PSI), which are exclusively available to the PI team, and Special Purpose Principal Investigator Science Instruments (SSI), which are available to people collaborating with the instrument team.

For extrasolar planet observations it can observe transiting planets during primary and secondary eclipses. It will also observe proto-planetary disks and planet formation in nearby star systems.

10.10
Keck Interferometer

The Keck Interferometer connects the two 10-meter Keck telescopes on top of Mauna Kea, Hawaii, as an infrared interferometer with a baseline of 85 meters.

A Jupiter-sized planet at a temperature of 900 K is about 10 000 times dimmer than a solar-type star. Our own Jupiter is about a million times fainter than the sun in the thermal infrared, and a billion times fainter in the visible and near infrared. Direct detection of a "cold Jupiter" can only be done from space, but direct detection of a hot Jupiter can be done with the Keck Interferometer. The Keck Interferometer will be able to detect the radiated light from Jupiter-sized planets at a separation of 0.15 AU from a parent star, and at a distance of 10 pc, through the use of multi-color phase-difference interferometry.

While very accurate wide-angle astrometry is only possible from space with a mission like SIM, narrow-angle astrometry with an accuracy of tens of microarcseconds is possible from the ground with an optimized instrument. Utilizing the proposed, but currently unsupported, outrigger telescopes, the Keck Interferometer would be able to survey hundreds of nearby stars out to 25 pc for the presence of planets.

The Keck Interferometer would complement SIM for astrometric planet detection. SIM will provide ultra-high accuracy for detection of planets down to several earth masses, while the Keck Interferometer would provide a long time scale for detecting longer-period systems.

10.11
Large Binocular Telescope Interferometer (LBTI)

The LBTI will be able to directly detect extrasolar giant planets and study solar system formation. It will provide high-resolution images of faint objects over a wide field of view. The LBT, equipped with adaptive optics and fully exploiting the quality of its mirror surface, will permit formation of diffraction-limited images, such that a planet could be detected against a low surface brightness halo of residual scattered light. In this manner, a Jupiter-like planet could be detected, if present, around some fifty of the nearest stars. The interferometric mode will enhance the planet/background contrast even further, thus increasing the number of candidate stars and the sensitivity of the survey. The LBTI will be able to study emissions for faint dust clouds around stars by means of nulling techniques. The dust clouds reflect light and give off heat, which interferes with planet searches. LBTI will characterize the emissions.

Observations with existing telescopes have shown that star (and presumably planet) formation is a continuing process within dense interstellar clouds in our own and other galaxies. However, very little is known about the nature of the process, especially the role of angular momentum, magnetic fields and initial turbulence on fragmentation (the final mass spectrum), the formation of pre-planetary disks, and the formation of binary stars. The difficulties arise, both due to the obscuration of visible light by interstellar dust in star-forming regions, and because, during the stages of interest, the condensations are sufficiently cool that they are faint and radiate mainly in the infrared, where the resolution and sensitivity of existing telescopes is too low to solve the problems. The LBTI will provide almost an order of magnitude gain in both resolution and sensitivity and should permit major advances.

LBTI is two 8 meter telescopes connected as an infrared interferometer with a maximum baseline of 22.8 meters. The configuration allows essentially complete sampling of all spatial frequencies in the image up to 22.8 meters using interferometric imaging between the two 8.4-m pupils. This provides unique capabilities for high-resolution imaging of faint objects. When combined with adaptive optics, the LBT interferometric mode offers high signal-to-noise imaging on even the faintest objects, over a relatively wide field. The revolutionary optical design, especially the f/1 borosilicate honeycomb primary mirrors, permit the entire LBT to be housed in a compact, low-cost structure. The rigid, honeycomb primaries also permit simple mirror-support systems. The optical configuration also permits incorporation of a one degree field of view, which is well matched to array detectors at f/4, for wide field imaging and multi-object spectroscopy at visible wavelengths. The binocular structure will also provide a natural way of storing secondary mirrors onboard the telescope, and interchanging them in 10–15 minutes to take advantage of changing observation conditions.

10.12
Palomar Observatory

There are some planet-finding projects at Palomar Observatory. These include Sleuth, Single Telescope Extrasolar Planet Survey (STEPS), and the Palomar Testbed Interferometer.

Sleuth is a 10 cm, robotic telescope dedicated to search for gas-giant planets transiting a star. It nightly monitors about 10 000 stars brighter than 12th visual magnitude in a six degree field of view. Sleuth is part of the Transatlantic Exoplanet Survey (TrES) network that includes Stellar Astrophysics and Research of Exoplanets (STARE) located on Tenerife in the Canary Islands and Planet Search Survey Telescope (PSST) located in northern Arizona. Candidates detected by Sleuth will be followed up to eliminate false positives by Sherlock, currently under construction.

STEPS uses a precisely calibrated CCD detector to try to detect extrasolar planets astrometrically. The Palomar 5 meter telescope has been used for a feasibility study for this project.

The Palomar Testbed Interferometer (PTI) was designed to test the hypothesis that the limits of ground-based astrometry due to atmospheric turbulence for very narrow-angle astrometry with long-baseline interferometers, could achieve an astrometric precision that was perhaps 100 times better than what was generally accepted in the 1980's. In addition, PTI is designed to test several key technologies for the Keck Interferometer and SIM.

10.13
Pan-STARRS

The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) is four 1.8 meter diameter mirror telescopes simultaneously observing the same region of the sky. Each telescope will have a three degree field of view and a CCD camera with one billion pixels. For the principal observing program of searching for near earth asteroids, 6000 square degrees of the sky will be covered each night. The whole available sky from Hawaii will be observed three times during the dark time of each lunation. Pan-STARRS is expected to reach 24th magnitude with exposures ranging between 30 and 60 seconds. Orthogonal Transfer CCDs (OTCCDs) allow shifting the charge along both rows and columns to achieve on-chip image motion compensation, similar to tip-tilt adaptive optics. Thus, Pan-STARRS will be able to observe large areas of the sky with great sensitivity and detect moving and variable objects.

A single mirror prototype is under construction at Haleakala on Maui and expected to have first light in 2007. The complete system will be located at Haleakala or Mauna Kea. The construction of Pan-STARRS is funded by the US Air Force. The University of Hawaii is seeking funding for the operations of Pan-STARRS.

Pan-STARRS can be used for a photometric survey that lasts for a month in a given region, to search for planet transits. With its field of view it should either detect a few habitable planets or put a constraint on their frequency. It should also detect eclipsing binaries, flares and spots, rotational rates, and weather patterns. It is expected that extrasolar planets should have neutral colors, while time-variable phenomena on stars should have strong color effects. The filters on the four telescopes will help separate the effects and provide preferred observations for looking for hot Jupiters and planet transits around brown dwarfs.

10.14
Thirty Meter Class Telescopes

In response to the decadal committee recommendation and the Astronomy and Astrophysics Advisory Committee, there are a number of concept studies for telescopes in the thirty meter size range. These have different consortiums of backers and names, like the Giant Segmented Mirror Telescope (GSMT), High Dynamic Range Telescope (HDRT), and Thirty Meter Telescope (TMT). They feature different design characteristics and offer different advantages for searching for and characterizing extrasolar planets. They all are seen as complementing the James Webb Space Telescope (JWST).

The GSMT is conceived to have 10 times the collecting area of the 10m Keck and 28 times that of the JWST. It will be sensitive to light throughout the blue and visible spectrum, while JWST operates mostly in the near to mid-infrared. GSMT will have sophisticated adaptive optics to provide superb image quality along with the great sensitivity. It will have high spectral resolution (100 000) in the optical and IR. Hence, GMST is expected to separate the light from Jupiter-like planets from parent stars out to 500 light years and to reveal the structure and chemical composition of the atmospheres of these planets. In the region shortward of 5 microns, GSMT will have high resolution imaging and spectroscopy with possible coronography and adaptive optics to probe planets and disk separations with better spectral resolution than JWST.

The Thirty Meter Telescope (TMT) is to be a 30 meter facility operating in the wavelength range of 0.3–30 microns. TMT will use adaptive optics to achieve diffraction-limited performance. TMT will explore planet-formation processes and characterization of extrasolar planets. TMT will seek spectroscopic discovery observations in the terrestrial planet regime, kinematics of proto-planetary disks, spectroscopic detection and analysis of extrasolar planet atmospheres, and direct detection of extrasolar planets in reflected and emitted light.

The High Dynamic Range Telescope (HDRT) is to have a light-collecting area equivalent to a 15.9 m unobstructed aperture. It will have an off-axis optical design of 6.5m mirrors with an unobstructed pupil for imaging of faint objects near bright sources. Wide and narrow field, low-scattered light properties will be excellent for direct imaging and coronographic applications. An optical resolution of 0.005 arcsecond is planned. Operating modes will provide wide (2 degrees) and narrow (1 minute) fields of view. High spatial resolution will be achieved from adaptive optics. The science drivers for this telescope include extrasolar planets, star formation, Kuiper belt objects, weak lensing surveys, and wide-field galaxy and redshift surveys.

10.15
Scientific Results

The different missions and ground-based programs seek to detect and discover extrasolar planets, determine their orbits and characteristics, determine the frequency of planets and solar systems around stars, investigate the form and methods of developing solar systems, and explore the possibilities of habitable planets and living creatures in the universe. The different programs can contribute information to the collection of knowledge about planets.

The programs can all determine orbits of discovered planets and establish if they are in the habitable regions. However, the more capable missions will want to investigate the possible presence of solar systems where individual planets have been detected. Also, there will be marginal detections of planets that should be pursued by the more capable missions to verify or reject the detections.

The planet's temperature and its variability can be estimated from the orbital size and eccentricity, and from an estimate of the albedo. By combining infrared and visible spectra, the effective temperature and model of the atmosphere can be established and the surface temperature estimated.

If the planet's mass and radius can be determined, the density can be determined which will infer its composition (ice, gas, rocks, or a combination), its habitability (solid surface and water), and its history (where formed). The combination of observational methods can contribute to the knowledge of the mass, radius, and brightness, along with spectral information. Surface gravity, which can be determined from the planet's mass and radius, is a good indicator of habitability, because a massive, dense planet is more likely to have plate tectonics and retain an atmosphere, which are crucial for the earth.

The combination of observations in the infrared and visible wavelengths can determine the compositions of the atmosphere and surface of planets. TPF-C can detect the presence of water, oxygen, ozone, carbon dioxide, methane, and clouds in the atmosphere. It can measure vegetation from the "red edge" effect. TPF-I can measure carbon dioxide, ozone, water, methane, and nitrous oxide from spectral features. The two missions will be able to determine the relative abundances of the elements and thermal structure of the atmosphere.

The presence of water vapor can be detected from the spectral observations of TPF- I and TPF-C. However, habitability requires the presence of liquid water, which in turn requires the solid surface and a temperature in the necessary range. So, the mass and radius of the planet must be determined by SIM, or another such mission.

Spectral observations can also detect the changes in color or spectral features as the planet rotates. From that data the length of the day, the presence of oceans or land masses, and changes in weather patterns might be determined.

10.16
Conclusions

The large number of missions and the different techniques being used almost assures that there will be significant discoveries and progress in our knowledge of planetary systems. The knowledge of their frequency and their characterization will indicate the formation methods for planets and the probability of solar systems. The capability to discover earth-like planets, their probability in habitable regions, and their chemical characteristics should bring new knowledge concerning the possibilities for living creatures elsewhere in the universe.

References

1 Marcy, G.W., D.A. Fischer, C. McCarthy, and E.B. Ford: 2005, “Exoplanets and the Space Interferometry Mission”, Astrometry in the Age of the Age of the Next Generation of Large Telescopes, P. Kenneth Seidelmann and Alice K.B. Monet eds, *Astronomical Society of the Pacific Conference Series, 338,191–200*.

2 Matthews, J. : 2005, “A suitcase full of astrophysics: the MOST microsat and opportunities for low cost space astrometry”, Astrometry in the Age of the Age of the Next Generation of Large Telescopes, P. Kenneth Seidelmann and Alice K.B. Monet eds, *Astronomical Society of the Pacific Conference Series, 338, 37–45*.

3 Nolan, E.: 2005, “Astrometry with the James Webb Space Telescope”, Astrometry in the Age of the Age of the Next Generation of Large Telescopes, P. Kenneth Seidelmann and Alice K.B. Monet eds, *Astronomical Society of the Pacific Conference Series, 338,59*.

4 Traub, W.A., S.T. Ridgeway, C. A. Beichman, K.J. Johnston, J. Kasting, and M. Shao: 2005, Transit, Astrometric, Coronagraphic and Interferometric Exo-planet Studies- Synergy and Complementarity, *Proceedings IAU Colloquium No 200, Claude Aime ed*.

5 Unwin, S.C.: 2005, “Precision Astrometry with the Space Interferometry Mission-PlanetQuest”, Astrometry in the Age of the Age of the Next Generation of Large Telescopes, P. Kenneth Seidelmann and Alice K.B. Monet eds, *Astronomical Society of the Pacific Conference Series, 338, 37–45*.

Websites

http://kepler.nasa.gov
http://planetquest.jpl.nasa.gov
http://tpf.jpl.nasa.gov
http://sim.jpl.nasa.gov
www.spitzer.caltech.edu
www.jwst.nasa.gov
www.sofia.usra.edu
www.astro.ubc.ca/MOST/
http://medusa.as.arizona.edu/lbto/index.html

http://planetquest.jpl.nasa.gov/keck/keck_index.cfm
http://www.tmt.org/tmt/science
http://pan-starrs.ifa.hawaii.edu/public
http://www2.ifa.hawaii.edu/newsletters/article.cfm?a=26&n=1
http://www.aura-nio.noao.edu/book

11
Habitable Zones in Extrasolar Planetary Systems

Werner von Bloh, Christine Bounama, and Siegfried Franck

Abstract

The habitable zone, (HZ) around a given central star is defined as the region within which an earth-like planet might enjoy the moderate surface temperatures required for advanced life forms. At present, there are several models determining the HZ. One class of models utilizes climate constraints for the existence of liquid water on a planetary surface. Another approach is based on an integrated earth system analysis that relates the boundaries of the HZ to the limits of photosynthetic processes. Within the latter approach, the evolution of the HZ for our solar system over geological time scales is calculated straightforwardly, and a convenient filter can be constructed that picks the candidates for photosynthesis-based life from all the extrasolar planets discovered by novel observational methods. Using this approach for determining the HZ it is possible to estimate the likelihood of finding habitable earth-like planets on stable orbits for 86 selected extrasolar planetary systems, where luminosity, effective temperature and stellar age are known. We find that habitability strongly depends on the age of the stellar system and the characteristics of a virtual earth-like planet. In particular the portion of land/ocean coverages plays an important role. We approximated the conditions for orbital stability using a method based on the Hill radius. Almost 60% of the investigated systems could harbor habitable earth-like planets on stable orbits. In 18 extrasolar systems we find even better prerequisites for dynamic habitability than in our own solar system.

11.1
Introduction

Is there life beyond planet Earth? This is one of the fundamental questions which humankind tries to solve through scientific research. The extraterrestrial life debate spans from the ancient Greek world of Democritus over the 18th century European world of Immanuel Kant to the recent discoveries

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

of extrasolar planets. Concerning the search for life in our planetary system, Schiaparelli's observation of a system of "canali" on the Martian surface in 1877 was the beginning of an epochal effort to reveal planetary conditions relevant to life with the refinement of observational techniques. The detection by McKay et al. [1] of the chemical biomarkers and possible microfossils in a meteorite from Mars called ALH 84001 (a meteorite found in 1984 in Antarctica) has stimulated research in the newly emerging field of astrobiology. Mars holds great interest for the latter, and presently stands centre-stage in the plans to explore the inner solar system for signs of past or present life. Even now, it can be stated that the search for extra-terrestrial life will be one of the predominant themes of science in the 21st century.

The definition of habitability is closely related to the very definition of life. Up to now, we only know terrestrial life, and therefore the search for extra-terrestrial life is the search for life, as we know it, from our home planet. Life can be defined as a self-sustained system of organic molecules in liquid water immersed in a source of free energy. It is well known that organic molecules are rather common in the solar system and even in interstellar clouds. There is also no problem in finding any source of free energy for extra-terrestrial life. Therefore, the existence of liquid water is the central point in the search for extra-terrestrial life and in the definition of habitability. Nevertheless, it is evident that liquid water and basic nutrients are essential but not sufficient requirements for life. Another important factor is the evolution of the solar radiation and particle flux. In particular high XUV fluxes might lead to a atmospheric loss of planets in close-in habitable zones defining an additional constraint on habitability.

The histories and fates of the three "terrestrial planets" – Venus, Earth, and Mars – suggest that a combination of factors like distance from the sun, planetary size, as well as geological and perhaps biological evolution, will control the existence of liquid water at a planetary surface. Earth-like planets cannot have liquid surface water if they are much closer to the sun than one astronomical unit (as defined by earth's orbit), because unfavorably high temperatures and loss of water by photo-dissociation would be unavoidable in this case. On the other hand, an earth-like planet, which is quite distant from the sun, would have permanent surface temperatures below the freezing point of water, and therefore would not be habitable either.

The first detection of a Jupiter-sized planet around 51 Peg in 1995 changed our knowledge about extrasolar planetary systems from fiction to fact. Up to now, more than 200 extrasolar planets have been discovered. Questions that scientists once considered beyond the reach of observation may soon have at least partial answers. In particular, is there life on planets found beyond the solar system, or are they at least habitable? Fundamental works in these fields have been performed by Lovelock [2, 3], and Hitchcock and Lovelock [4].

They discuss general interactions between life and its planetary environment. Later on, discussing the possibility of self-regulation by and for the biosphere, Lovelock and Margulis [5] (see also [6, 7]) put forward the Gaia hypothesis.

Studies conducted by the authors of this review (see, e.g., [8, 9]) have generated a rather general characterization of habitability, based on the possibility of photosynthetic biomass production under large-scale geodynamic conditions. Thus, not only is the availability of liquid water on a planetary surface taken into account but also the suitability of CO_2 partial pressure. As a matter of fact, the same type of stability calculations sketched above for the solar system, with the sun as the central star, can also be performed for other stars. The basic results for the habitable zone (HZ) around these other central bodies are relatively simple. In order to have a surface temperature in the terrestrial range, a planet orbiting a central star with lower mass would have to be closer to the star than 1 AU, whereas a planet orbiting a brighter star having more mass than our sun would have to be farther away than 1 AU from the star. But the problem is a bit more complicated. One also has to take into account the different times that stars spend on the so-called main sequence. Such stars receive their energy mainly from hydrogen burning, i.e., the fusion of hydrogen to helium [10, 11].

In the following, the possibilities of the existence of life on an earth-like planet at various distances from the sun are investigated. The method is based on the earth system science approach [12] that calculates the past and future evolution of a dynamic earth under the influence of an increasing solar luminosity [8, 13]. Such a long-term climate regulation is thought to be governed by the global carbon cycle and its effect on atmospheric CO_2 content and biological productivity. Our methodological apparatus is sketched in the next section, and results for the habitable zone for our solar system are presented afterwards. The model is extended to extrasolar planetary systems with central stars different from our sun, and the insights obtained there are used to estimate the likelihood to find habitable worlds in known extrasolar planetary systems. Orbital stability of a virtual earth-like planet in an extrasolar planetary system has to be taken into account as an additional constraint for habitability in the case of the presence of a giant planet. In this way we can define a dynamic habitable zone.

11.2
Modeling the Earth System

The earth system model employed here is a stylized geosphere–biosphere model to analyze the evolution of this complex on a global scale from the geological past to the planetary future in 1.5 billion years. The model consists of the components solid earth, hydrosphere, atmosphere, and biosphere. It couples the increasing solar luminosity, the silicate-rock weathering rate, and the global energy balance to estimate the partial pressure of atmospheric and soil carbon dioxide, the mean global surface temperature, and the biological productivity as a function of time. The crucial point is the long-term balance between the CO_2 sink in the atmosphere–ocean system and the metamorphic (plate-tectonic) sources. In the following, the respective features of our earth system model are described in some detail.

11.2.1
Planetary Climate

The climate of a planet is governed by the energy balance equation between incoming and outgoing radiation. The pertinent equation is based on the simple zero-dimensional model of climate that was already introduced by Arrhenius [14]. In this model, the surface temperature of planet earth is the sum of the effective (black body) radiation temperature depending on insolation, and the greenhouse warming, depending on atmospheric carbon dioxide content. The increase of insolation results from the increase of the main-sequence hydrogen-burning rate, and can be well described by rather simple formulas (see, e.g., [15]). During earth history, the luminosity of the sun has increased to the present level with a rate of about 10% per Gyr, and will increase in the future (up to the next 5 Gyr) at approximately the same rate.

The connection between the stellar parameters and the planetary climate can be formulated by using a radiation balance equation [16]

$$\frac{L}{4\pi R^2}\left[1-\alpha(T_{\text{surf}}, P_{\text{atm}})\right] = 4I_R(T_{\text{surf}}, P_{\text{atm}}) \qquad (11.1)$$

Here α denotes the planetary albedo, I_R the outgoing infrared flux, and R the distance from the sun or central star.

11.2.2
The Global Carbon Cycle

The carbon cycle is the main process for the regulation of the atmospheric composition and climate with respect to increasing insolation. Walker et al. [17], Berner [18], Berner et al. [19], Lasaga et al. [20], von Bloh et al. [21], and Franck et al. [22] as well as many others, investigated the so-called carbonate–silicate geochemical cycle between the atmosphere, the ocean, and the continents. The main feedback loop stabilizing the planetary climate is given by the silicate rock weathering. An increase in the luminosity leads to a higher mean global temperature causing an increase in weathering. Then more CO_2

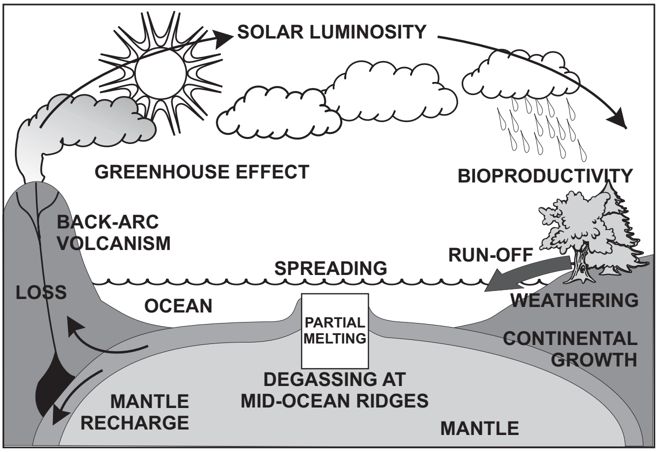


Fig. 11.1 The global carbon cycle as part of the general volatile exchange between mantle and surface reservoirs.

is extracted from the atmosphere, weakening the greenhouse effect. Overall the temperature is lowered and homeostasis is achieved.

On geological time scales, however, the deeper parts of the earth are considerable sinks and sources for carbon and the tectonic activity as well as the continental area have change markedly. Therefore, Tajika and Matsui [23] have favored the so-called "global carbon cycle". In addition to the usual carbonate–silicate geochemical cycle, it also contains the subduction of large amounts of carbon into the mantle with descending slabs and the degassing of carbon from the mantle at mid-ocean ridges. The global carbon cycle is sketched in Fig. 11.1.

11.2.3
Silicate-Rock Weathering Rate

Weathering plays an important role in the earth's climate because it provides the main sink for atmospheric carbon dioxide. The overall chemical chemical reactions for the weathering process are

$$\begin{aligned} CO_2 + CaSiO_3 &\rightarrow CaCO_3 + SiO_2 \\ CO_2 + MgSiO_3 &\rightarrow MgCO_3 + SiO_2 \end{aligned}$$

The question of to what extent the biota are actually able to play a key role in stimulating this sink is crucial for an understanding of the dynamic properties of the overall earth system. The total process of weathering embraces (i) the reaction of silicate minerals with carbon dioxide, (ii) the transport of weathering products, and (iii) the deposition of carbonate minerals in sediments. Combining all these effects, the global mean weathering rate can be formulated via an equation introduced by Walker et al. [17].

$$\frac{F_{\mathrm{wr}}}{F_{\mathrm{wr},0}} = \left(\frac{a_{H^+}}{a_{H^+,0}}\right)^{0.5} \exp\left(\frac{T_{\mathrm{surf}} - T_{\mathrm{surf},0}}{13.7\ K}\right) \tag{11.2}$$

Here the pre-factor outlines the role of the CO_2 concentration in the soil, P_{soil}; a_{H^+} is the activity of H^+ in fresh soil-water and depends on P_{soil} and the global mean surface temperature T_{surf}. The quantities $F_{\mathrm{wr},0}$, $a_{H^+,0}$, and $T_{\mathrm{surf},0}$ are the present-day values for the weathering rate, the H^+ activity, and the surface temperature, respectively. The activity a_{H^+} is itself a function of the temperature and the CO_2 concentration in the soil. The equilibrium constants for the chemical activities of the carbon and sulfur systems involved have been taken from Stumm and Morgan [24]. Note that the sulfur content in the soil also contributes to the global weathering rate, but its influence does not depend on temperature. It can be regarded as an overall weathering bias, which has to be taken into account for the estimation of the present-day value.

For any given weathering rate the surface temperature and the CO_2 concentration in the soil can be calculated self-consistently, as will be shown below. P_{soil} can be assumed to be linearly related to the terrestrial biological productivity Π (see [25]) and the atmospheric CO_2 concentration P_{atm}. Thus we have

$$\frac{P_{\mathrm{soil}}}{P_{\mathrm{soil},0}} = \frac{\Pi}{\Pi_0}\left(1 - \frac{P_{\mathrm{a\ tm},0}}{P_{\mathrm{soil},0}}\right) + \frac{P_{\mathrm{atm}}}{P_{\mathrm{soil},0}} \tag{11.3}$$

where $P_{\mathrm{soil},0}$, Π_0 and $P_{\mathrm{atm},0}$ are again present-day values.

11.2.4

Biological Productivity

The biological productivity, Π, is the amount of biomass that is produced by photosynthesis per unit time and per unit of continental area. In reality, Π itself is a function of various parameters such as water supply, photosynthetic active radiation, nutrients, atmospheric carbon dioxide content, and surface temperature. In our model, biological productivity is considered to be solely a function of the surface temperature and the CO_2 partial pressure in the atmosphere

$$\begin{aligned}\frac{\Pi}{\Pi_{\max}} &= \Pi_T(T_{\mathrm{surf}}) \cdot \Pi_P(P_{\mathrm{atm}}) \\ &= \max\left(\left(1 - \left(\frac{T_{\mathrm{surf}} - 50°\mathrm{C}}{50°\mathrm{C}}\right)^2\right)\left(\frac{P_{\mathrm{atm}} - P_{\min}}{P_{1/2} + (P_{\mathrm{atm}} - P_{\min})}\right), 0\right)\end{aligned} \tag{11.4}$$

Here $\Pi_{\max}$ denotes the maximum biological productivity, which is assumed to amount to twice the present value Π_0 [25]. $P_{1/2} + P_{\min}$ is the value at which the pressure-dependent factor is equal to $1/2$, and $P_{\min}$ is fixed at 10^{-5} bar, the presumed minimum value for C_4-photosynthesis [26, 27]. The evolution of the biosphere and its adaption to even lower CO_2 partial pressures are not taken into account in our model. For a given P_{atm}, Eq. (11.4) yields maximum productivity at $T_{\mathrm{surf}} = 50°\mathrm{C}$ and zero productivity for $T_{\mathrm{surf}} \leq 0°\mathrm{C}$ and $T_{\mathrm{surf}} \geq 100°\mathrm{C}$. There exist hyperthermophilic life forms with a temperature tolerance well above 100°C. In general, these are chemoautotrophic organisms, not included in this study.

11.2.5
Geodynamics

Caldeira and Kasting [28] have investigated an earth-system model using Eq. (11.2) under the assumption that the weathering rate is always equal to the present value:

$$\frac{F_{\mathrm{wr}}}{F_{\mathrm{wr},0}} \equiv 1 \tag{11.5}$$

This is clearly a rather rough approximation, and we call this approach the geostatic model. In the framework of a geodynamic-equilibrium approach for the global carbon cycle at longer time scales of about hundred thousands of years, Walker et al. [17] first proposed a balance between the CO_2 sink in the atmosphere–ocean system and the metamorphic source. The main forcings and feedbacks acting in the earth system are depicted in Fig. 11.2. Using the balance, one can find a relation between the weathering rate on the one side, and continental area and spreading rate on the other side [29]. This can be expressed with the help of dimensionless quantities [19]

$$f_{\mathrm{wr}} \cdot f_A = f_{\mathrm{sr}} \tag{11.6}$$

where $f_{\mathrm{wr}} \equiv F_{\mathrm{wr}}/F_{\mathrm{wr},0}$ is the weathering rate normalized by the present value, $f_A \equiv A_c/A_{c,0}$ is the continental area normalized by the present value, and $f_{\mathrm{sr}} \equiv S/S_0$ is the spreading rate normalized by the present value. Equation (11.6) can be rearranged by introducing the geophysical forcing ratio, GFR [25]:

$$f_{\mathrm{wr}}(T_{\mathrm{surf}}, P_{\mathrm{atm}}) = \frac{f_{\mathrm{sr}}}{f_A} =: GFR(t) \tag{11.7}$$

With the help of Eq. (11.1) it is possible to replace P_{atm} by T_{surf} in Eq. (11.7) and this results in:

$$f_{\mathrm{wr}}(T_{\mathrm{surf}}, L, R) = GFR(t) \tag{11.8}$$

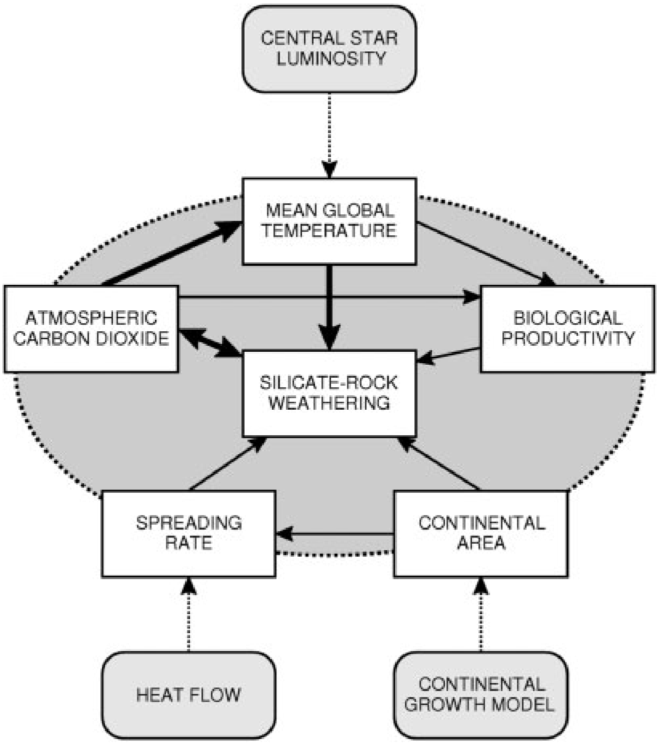


Fig. 11.2 Earth system box model. The arrows indicate the different forcing and feedback mechanisms.

The evolution of the surface temperature can be derived by solving Eq. (11.8) for known luminosity, distance to the central star and geophysical forcing ratio. Eq. (11.1) yields the corresponding evolution of the atmospheric CO_2 partial pressure.

Therefore, knowing continental area and spreading rate, one can calculate the weathering rate for both the geological history and the planetary future. This is called the geodynamic model that constitutes a significant improvement over the previous approach. Figure 11.3 depicts the continental growth models fed into the calculations.

The spreading rate can be determined according to boundary layer theory of convection [13], which is a main result of the so-called parameterized convection models for the earth's thermal evolution [31]. It is a function of the mantle temperature T_m, surface temperature T_{surf}, oceanic area A_o and heat flow from the mantle q_m [32]:

$$F_{\text{sr}} = \frac{q_m^2 \pi \kappa A_o(t)}{4k^2 (T_m - T_{\text{surf}})^2} \tag{11.9}$$

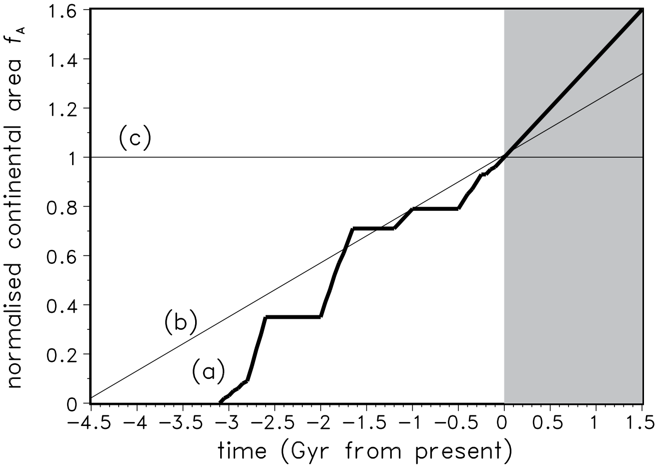


Fig. 11.3 Normalized continental area as a consequence of the following continental growth models: (a) based on Condie [30], (b) linear growth, (c) fixed area.

where k is the thermal conductivity and κ is the thermal diffusivity. Assuming conservation of energy we can write the following equation for the average mantle temperature, T_m:

$$\frac{4}{3}\pi\rho c(R_m^3 - R_c^3)\frac{dT_m}{dt} = -4\pi R_m^2 q_m + \frac{4}{3}\pi Q(t)(R_m^3 - R_c^3) \tag{11.10}$$

where ρ is the density, c is the specific heat at constant pressure, Q is the energy production rate by decay of radiogenic heat sources in the mantle, and R_m and R_c are the outer and inner radii of the mantle, respectively.

The mantle heat flow is parameterized in terms of the Rayleigh number, Ra:

$$q_m = \frac{k(T_m - T_{\text{surf}})}{(R_m - R_c)}\left(\frac{Ra}{Ra_{\text{crit}}}\right)^{\beta} \tag{11.11}$$

with

$$Ra = \frac{g\alpha(T_m - T_{\text{surf}})(R_m - R_c)^3}{\kappa\nu} \tag{11.12}$$

where Ra_{crit} is the critical value of Ra for the onset of convection, β is an empirical constant, g is the acceleration of gravity, α is the coefficient of thermal expansion, and ν is the water-dependent kinematic viscosity. This viscosity can be calculated with the help of a water fugacity dependent mantle creep rate. For details of this parameterization see Franck and Bounama [33].

11.3
The Habitable Zone in the Solar System

In general, the habitable zone (HZ) around the sun can be defined as the region within which an earth-like planet might enjoy moderate surface temperatures needed for advanced life forms. The more specific definition related to the existence of liquid water at the planet's surface was introduced by Huang [34, 35] and extended by Dole [36] and Shklovskii and Sagan [37]. Hart [38, 39] calculated the evolution of the terrestrial atmosphere over geologic time at varying distances. He found that the HZ between "runaway greenhouse" and "runaway glaciation" is surprisingly narrow for G2 stars like our sun: $R_{\text{inner}} = 0.958$ AU, $R_{\text{outer}} = 1.004$ AU. A main disadvantage of those calculations is the neglect of the negative feedback between atmospheric CO_2 content and mean global surface temperature. The full consideration of this feedback by Kasting [40] provided the interesting result of an almost constant inner boundary, yet a remarkable extension of the outer boundary. Later on, the calculations of the HZ were improved and extended to other main sequence stars [16, 41, 42]. The authors define the boundaries of the HZ via so-called critical solar fluxes. For the inner radius of the HZ, they give three different estimations. The first assumes loss of planetary water by a moist greenhouse [40]; the second assumes loss of planetary water by a runaway greenhouse; and the third is based on the observation that there was no liquid water on Venus' surface at least for the last one Gyr. The outer radius of the HZ is also estimated in three different ways. The first one is based on arguments that early Mars had a warm and wet climate [43, 44], the second one assumes a maximum possible CO_2 greenhouse heating; and the third one is related to first condensation limit of CO_2 clouds that increase the planetary albedo. The critical CO_2 partial pressure for the onset of this effect is about 5 to 6 bar. On the other hand, the effects of CO_2 clouds have been challenged by Forget and Pierrehumbert [45]. CO_2 clouds have the additional effect of reflecting the outgoing thermal radiation back to the surface.

Assuming the possibility of a "cold start", i.e., an originally ice-covered planet that was initially beyond the outer HZ boundary could be habitable in a later stage if the outer boundary of the HZ shifts outward to its orbit, Kasting et al. [42] found the following values for the present HZ in the solar system:

Most conservative case:	0.95 AU to 1.37 AU
Least conservative case:	0.75 AU to 1.90 AU
Intermediate case:	0.84 AU to 1.77 AU

The precise inner and outer limits of the climatic habitable zone are still unknown because of the limitations of climate model used until now. For the present sun it is probably smaller than the 0.7–2 AU region but it is still impossible to give a better constraint, especially for the outer boundary of the HZ.

Based on the earth system model described in the previous section, the HZ for the solar system can be determined in a different way. Here the HZ for an earth-like planet is the region around the sun within which the surface temperature of the planet stays between 0°C and 100°C and the atmospheric CO_2 partial pressure is higher than 10^{-5} bar, i.e., suitable for photosynthesis-based life (biological productivity $\Pi > 0$):

$$HZ := \{R \mid \Pi(P_{\text{atm}}(R,t), T_{\text{surf}}(R,t)) > 0\} \tag{11.13}$$

The upper limit of the CO_2 partial pressure is given by the maximum available amount of CO_2 in an earth-like planet's atmosphere, which is taken as 10 bar. The term "earth-like" explicitly implies the occurrence of plate tectonics as a necessary condition for the operation of the carbonate–silicate cycle as the mechanism to compensate the gradual brightening of the sun during its "life" on the main sequence. The geodynamical evolution of the considered earth-like planet provides an even stronger constraint. In the geological past, the volcanic input of CO_2 to the atmosphere was much higher and the continental area (available for weathering) was much smaller than today.

The results for the estimation of the HZ via the geodynamic model are summarized in Fig. 11.4, where we have plotted the width and position of the HZ for three different points in time (past, present, future). At this point we should emphasize that all calculations are done for a planet with earth mass and size, and an earth-like radioactive heating rate in its interior. In about 500 Myr the inner boundary, R_{inner}, reaches the earth distance from the sun ($R = 1$ AU) and the biosphere ceases to exist. The outer boundary, R_{outer}, decreases in a strongly nonlinear way. This result is in contrast to the results of Kasting et al. [42] and Kasting [41]. The calculations show that the inner boundary of the HZ is determined by the 10^{-5} bar limit and the outer boundary by the 0° C limit. From the view of geocybernetics [46], the inner and outer boundaries of the HZ can be considered as critical values of an ecological niche for life in general and humanity in particular. Of course, there may exist chemolithoautotrophic hyperthermophiles that might survive even in a future of higher temperatures, rather independently of atmospheric CO_2 pressures [see, e.g., 47]. But, under such conditions, all higher forms of life would certainly be eliminated. The biosphere model (Eq. (11.4)) is actually only relevant to photosynthesis-based life. Therefore, in the timespan under consideration, the upper temperature does not affect the results for the HZ. An earth-like planet at the Martian position would have been within the HZ from about 3.5 Gyr to about 0.5 Gyr ago [48]. The outer boundary of the HZ is mainly determined by the total amount of CO_2 that can be in the atmosphere. One should note that the required high atmospheric CO_2 content for a ter-

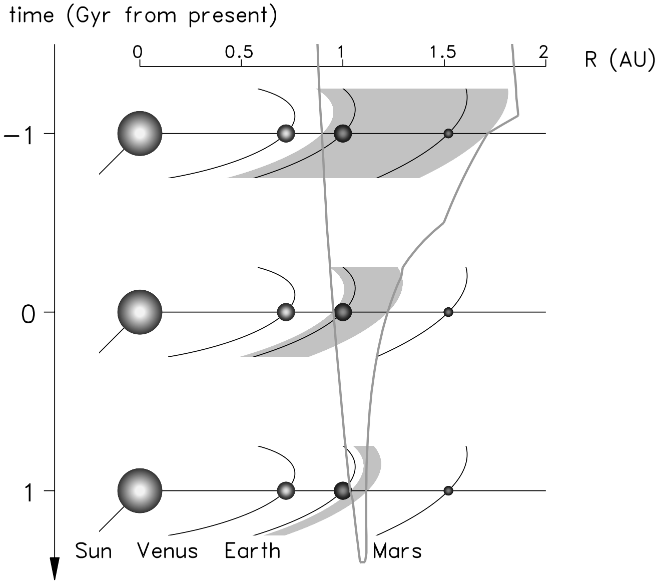


Fig. 11.4 Habitable zone (gray shading) for the solar system at three different time steps. The orbits of the three terrestrial planets Venus, Earth, and Mars are shown. The solid lines describe the evolution of the inner and outer boundary of the HZ.

restrial planet in the outer HZ would make it less likely for complex aerobic organisms to exist there. The possible extension of the HZ up to the Martian position seems to be realistic for the solar system past. This is a motivation to scrutinize planet Mars in some detail. We know that due to its smaller size all geological processes caused by the internal cooling of the planet should have faded away much faster than for the earth. Nevertheless, we can speculate that the results given above about the HZ are an upper bound for the time that Mars was habitable in the past. This is in good agreement with investigations concerning an early warmer and wetter Martian environment [49–51] and with observations that plate tectonics may have once operated on Mars [52]. On the other hand, there are theories that the Martian atmosphere was blown away by the solar wind, following the demise of its magnetic field four billion years ago. This happened so soon after Mars's formation that it is unlikely complex life would have had time to evolve [53]. In contrast to Mars, an earth-like planet at the Venusian position was never and will never be within the HZ.

11.4
Habitable Zone Around Main-sequence Stars

The same type of HZ calculations, on the basis of climatic constraints as well as on the basis of earth-system modeling, can be performed for stars with masses different from the solar mass. Kasting et al. [42] restricted themselves to stellar life times greater than 2 Gyr, which correspond to masses less than 1.5 $M_\odot$ (1 $M_\odot$ = one solar mass). Stellar luminosities and temperatures were taken directly from Iben [54, 55], climatic constraints corresponded to their so-called intermediate case (see above). As expected, stellar HZs for more massive stars are rather short because they have to be truncated at the end of the main sequence. HZs for low-mass stars remain essentially constant over time. In Fig. 11.5, we show the so-called zero-age main sequence HZ from Kasting et al. [42] as a function of stellar mass.

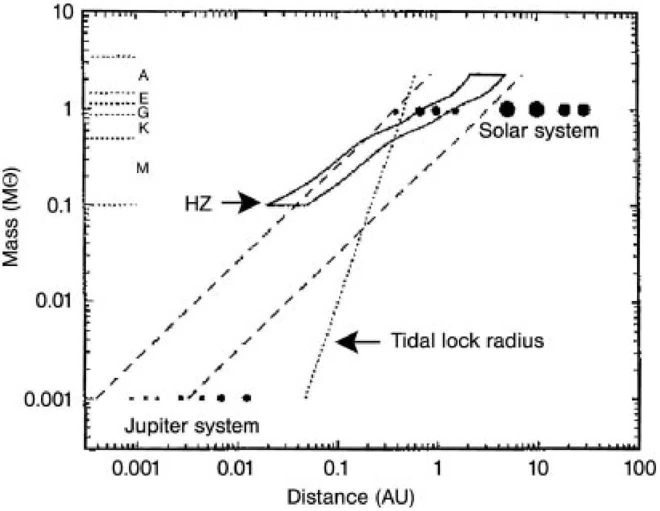


Fig. 11.5 The zero-age main sequence HZ as a function of central star mass (in solar masses $M_\odot$) for the intermediate case of climatic constraints. The long-dashed lines delineate the probable terrestrial planet accretion zone. The dotted line is the orbital distance for which an earth-like planet in a circular orbit would be locked into synchronous rotation (tidal locking) [42].

In Franck et al. [9], the HZ in extrasolar planetary systems is calculated using the luminosity evolution of central stars on the main sequence in the mass range between 0.8 and 2.5 $M_\odot$. The results have been obtained by polynomial fitting of detailed stellar evolution models by Schaller et al. [56]. The corresponding Hertzsprung–Russell diagram, i.e., a plot of luminosity versus effective radiating temperature, is shown in Fig. 11.6. The temperature toler-

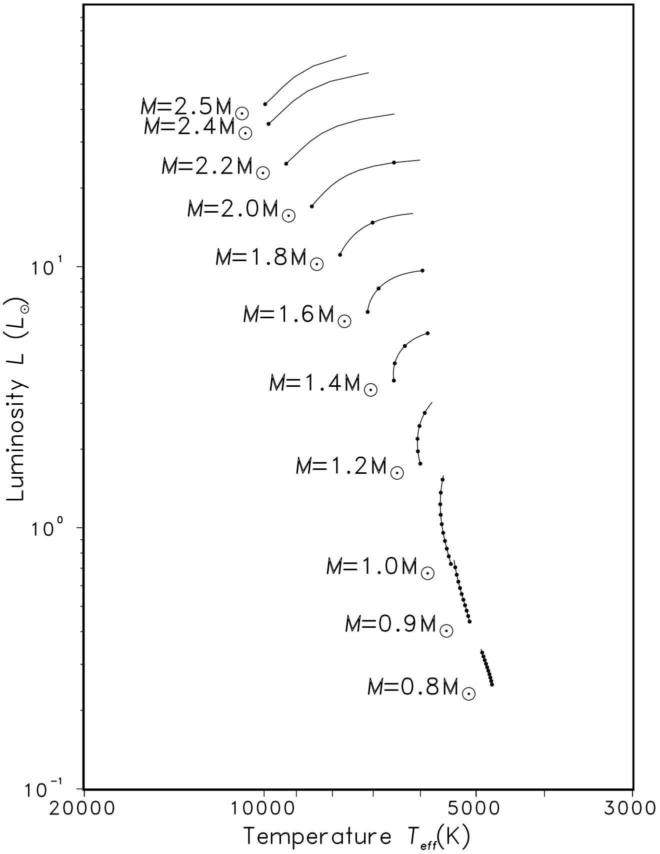


Fig. 11.6 Hertzsprung–Russell diagram for central stars in the mass range between 0.8 and 2.5 $M_\odot$. Only the main sequence evolution is considered. Successive dots on the mass-specific branches are separated in time by 1 Gyr.

ance window for the biological productivity was again in the range between 0°C and 100°C in order to incorporate thermophiles [57], but for this study a linear continental growth model (curve (b) in Fig. 11.3) was employed.

In principle, it is possible to calculate the HZ for any value of central-star mass shown in Fig. 11.6. As an illustration, we present the results for central-star masses of 0.8, 1.0, and 1.2 $M_\odot$, respectively, in Fig. 11.7.

An alternative way of presentation is to delineate the HZ for an earth-like extrasolar planet at a given (but arbitrary) distance *R* in the stellar mass-time plane. Here the following effects limit the HZ:

1. Stellar lifetime on the main sequence decreases strongly with mass. Using simple scaling laws [10], the central hydrogen burning period is estimated

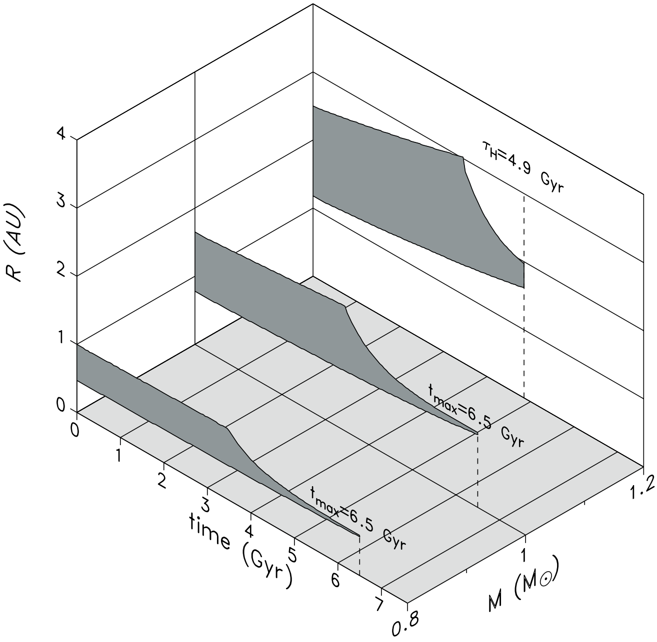


Fig. 11.7 Graphs of the width and position of the HZ derived from the geodynamic model for three different stellar masses M (0.8, 1.0, $1.2\,\mathrm{M}_\odot$). $t_{\max}$ is the maximum life span of the biosphere limited by geodynamic effects. τ_H indicates the hydrogen-burning time on the main sequence, limiting the life span of more massive stars.

to be $\tau_H < 0.8$ Gyr for $M > 2.2\,\mathrm{M}_\odot$. Therefore, there is no point in considering central stars with masses larger than $2.2\,\mathrm{M}_\odot$ because an earth-like planet may need 0.8 Gyr of habitable conditions for the development of life [38, 39]. Quite recently, smaller numbers for the timespan required for the emergence of life have been discussed, for instance 0.5 Gyr [58]. Performing calculations with $\tau_H < 0.5$ Gyr, one obtains qualitatively similar results, but the upper bound of central star masses is shifted to $2.6\,\mathrm{M}_\odot$.

2. When a star leaves the main sequence to turn into a red giant, there clearly remains no HZ for an earth-like planet. This limitation is relevant for stellar masses in the range between 1.1 and $2.2\,\mathrm{M}_\odot$.

3. In the stellar mass range between 0.6 and $1.1\,\mathrm{M}_\odot$, the maximum life span of the biosphere is determined exclusively by planetary geodynamics, which is independent (in a first approximation, but see the limiting effect 4.) of R. So one obtains the limitation $t < t_{\max}$, where $t_{\max} = 6.5$ Gyr.

4. There have been discussions about the habitability of tidally locked planets. This complication is taken into account by indicating the domain where an earth-like planet on a circular orbit experiences tidal locking. That domain consists of the set of (M, t)-couples, which generate an outer HZ boundary below the tidal-locking radius. This limitation is relevant for $M < 0.6\,\mathrm{M}_\odot$. Global climate models of tidally locked planets have indicated that earth-like planets should not necessarily be precluded from habitability [59].

As an example, we depict the HZ for $R = 2$ AU in Fig. 11.8. We have also shown the parameters of the 5.5 earth-mass planet OGLE-05-390Lb, which was assumed to orbit a $0.22\,\mathrm{M}_\odot$ star at a distance of $\approx$ 2.1 AU [60]. This system would be no candidate for extra-terrestrial life, because it is clearly outside the HZ.

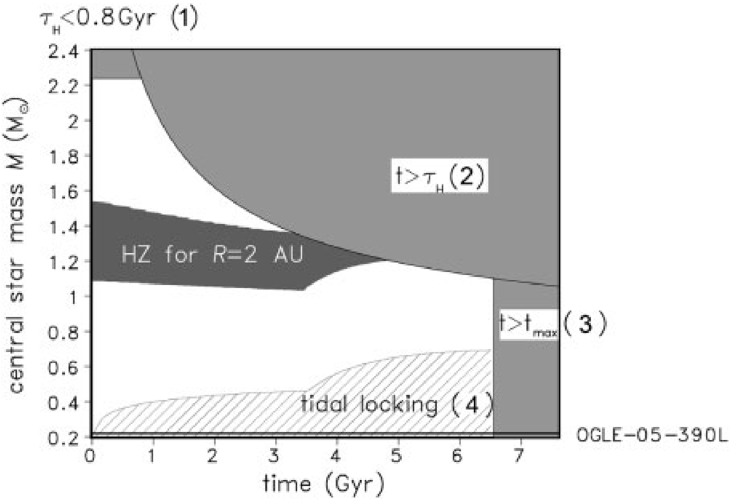


Fig. 11.8 Shape of the HZ (dark gray shading) in the mass–time plane for an earth-like planet at distance $R = 2$ AU from the central star. The potential overall domain for accommodating the HZ for planets at some arbitrary distance is limited by a number of factors that are independent of R: (1) minimum time for biosphere development, (2) central star lifetime on the main sequence, (3) geodynamics of the earth-like planet, and (4) tidal locking of the planet (nontrivial subdomain excluded). The excluded realms are marked by gray shading in the case of the first three factors and by gray hatching for the tidal-locking effect. The horizontal line represents the OGLE-05-390Lb planet, which is assumed to orbit a $0.22\,\mathrm{M}_\odot$ star.

11.5
Dynamic Habitability in Extrasolar Planetary Systems

The search for extrasolar earth-like planets is one of the main goals of present research. More than 200 extrasolar planets are known to orbit around main-sequence stars including several multiple-planet systems. Most of them are giant planets, with hydrogen and helium as the main constituents, and have atmospheres too turbulent to permit the emergence of life and also have no underlying solid surfaces or oceans that could support a terrestrial biosphere. Nevertheless, there is a possibility for habitable conditions at the surface of moons orbiting giant planets that are positioned within the habitable zone. Furthermore, earth-like Trojan planets in 1:1 mean-motion resonance on stable orbits with habitable conditions are possible [61]. The distribution of masses of all known exoplanets lets scientists suppose that there must be a multitude of planets with lower masses [62, 63]. A planet with a mass of $\approx$14 earth masses has been detected at a distance of 0.038 AU from the central star [64]. Whether this planet is a hot Neptune or a rocky "super earth" it is clear that the planet is inhabitable. As we have shown in Fig. 11.8 the recently discovered cool sub-Neptune-mass planet of 5.5 earth masses orbiting an M-dwarf star at a distance $\approx$ 2.1 AU [60] is not a candidate for a habitable world.

The existence of earth-type planets around stars other than the sun is strongly implied by various observational findings including (1) the steep rise of the mass distribution of planets with decreasing mass, which implies that more small planets form than giant ones; (2) the detection of protoplanetary disks (with masses between ten and one hundred times that of Jupiter) around many solar-type stars younger than $\approx$3 Myr; and (3) the discovery of "debris disks" around middle-aged stars, the presumed analogs of the Kuiper Belt and zodiacal dust ([63, 65] and references therein). Lineweaver and Grether [66] conclude that 25%–100% sun-like stars harbor planets.

Even if it seems today beyond technical feasibility to detect earth-mass planets, we can apply computer models to investigate known exoplanetary systems to determine whether they could host earth-like planets with surface conditions sufficient for the emergence and maintenance of life on a stable orbit. Such a configuration is described as dynamic habitable. Jones et al. [67] have investigated the dynamic habitability of several exoplanetary systems. They used the boundaries of the habitable zone (HZ) originating from Kasting et al. [42]. To test the intersection of stable orbits and the HZ, putative earth-mass planets were launched into various orbits in the HZ and a symplectic integrator was used to calculate the celestial evolution of the extrasolar planetary system.

11.5.1
Orbital Stability

Planetary habitability requires orbital stability of the earth-type planet over a biologically significant length of time in the HZ. The analysis of orbital stability of (hypothetical) terrestrial planets in extrasolar planetary systems has to take into account the effects of the giant planet(s) in those systems. In many cases the giant planets restrict the orbital stability of the terrestrial planet to a small or very small orbital domain, or prevent orbital stability completely. There exists a variety of papers discussing the orbital stability of (hypothetical) terrestrial planets in extrasolar planetary systems, which is strongly influenced by the masses, orbital positions and eccentricities of Jupiter-size planets in such systems (see, e.g., [68–71]). Jones et al. [67] analyzed the stability of orbits of terrestrial planets in several known extrasolar planetary systems. They used a mixed-variable symplectic integrator by Chambers [72] over a time scale of $\approx 10^9$ years, while Pilat-Lohinger and Dvorak [73] used a Lie-series method (e.g., [74]) for the calculation of the orbits of an earth-like planet.

In order to calculate the orbital stability we apply an approximation derived by Jones et al. [75]. They evaluated the dynamic habitability using nR_H derived from giant's orbital eccentricity without carrying out time-consuming orbital integrations, where R_H is the Hill radius of the giant planet and n is a multiplier that depends on the giant's orbital eccentricity.

The Hill radius R_H is defined as

$$R_H = \left(\frac{m}{3M}\right)^{1/3} a \tag{11.14}$$

where m is the mass of the giant planet, M the central star mass and a the semimajor axis. Then the inner and outer boundaries for unstable orbits around a giant planet are given by

$$R_{\text{int}} = a(1-e) - n_{\text{int}}(e)R_H \tag{11.15}$$

$$R_{\text{ext}} = a(1+e) + n_{\text{ext}}(e)R_H \tag{11.16}$$

where e is the eccentricity of the giant planet. The values of the functions $n_{\text{int}}(e)$ and $n_{\text{ext}}(e)$ are taken from Table 4 in Jones et al. [75]. They are in the range of $[2\ldots3]$ for $n_{\text{int}}(e)$ and $[3\ldots16]$ for $n_{\text{ext}}(e)$. Even if the orbit is outside the boundaries given in Eqs. (11.15) and (11.16) the eccentricity of the earth-like planet will generally increase until the orbit is possibly outside the HZ for a significant fraction of the orbital period. The critical limit for the eccentricity is between 0.5 and 0.7. Williams and Pollard [76] conclude that, even under such conditions, an earth-like planet might be habitable for a dense enough atmosphere. Jones et al. [75] found that the eccentricity is usually less than 0.3–0.4 outside the interval $[R_{\text{int}}, R_{\text{ext}}]$.

In our study we use results of Espresate [77] for the stellar parameters of 133 extrasolar planetary systems. Missing values for stellar ages were taken from Jones et al. [75]. For 86 of these 133 systems the stellar luminosities, effective temperatures and ages are given. It must be pointed out that the determination of the age of low-mass stars is error-prone and has to be done with care. Together with the corresponding orbital parameters and masses of the giant planets (data taken from Jean Schneider's extrasolar planets encyclopaedia http://www.obspm.fr/encycl/encycl.html) the habitable zones and the orbital stability domains can be calculated using Eqs. (11.14), (11.15) and (11.16).

11.5.2
Results for 86 Extrasolar Planetary Systems

The HZ for a fixed central star luminosity of $L \equiv 1\,\mathrm{L}_\odot$ and fixed effective temperature of $T_{\mathrm{eff}} = 5700$ K, is calculated, as a function of the age of the planetary system. For the investigation of an earth-like planet under external forcing we adopt a model planet with a fixed continental area (see Fig 11.3). The fraction of continental area to the total planetary surface is varied between 0.1 and 0.9. According to Franck et al. [8], Table 1, a constant continental area yields the maximum life span of the biosphere. Therefore, we have chosen this scenario in order to get the most optimistic estimation of habitability of an extrasolar planetary system. The HZ is calculated by solving Eq. (11.8) for different distances to the central star R. According to Eq. (11.13) the inner and outer limits of the HZ are given by a vanishing biological productivity (Eq. (11.4)).

In Fig. 11.9 the shaded areas indicate the HZs for different constant continental areas of the earth-like planet. While the inner limit of the HZ does not change significantly with age, the outer limit shows a nontrivial behavior. Up to a critical age the outer limit is constant and is determined by the maximum CO_2 atmospheric pressure (5 bars). Beyond this critical point the outer boundary moves inward due to geodynamic effects. At this point the source of carbon released into the atmosphere is too low to prevent the planet from freezing. The ultimate life span of the earth-like planet is determined by the coincidence of the outer and inner boundary. For a planet older than this ultimate life span, no habitability can be found. The critical age and the ultimate timespan is a decreasing function of the relative continental area of the earth-like planet. It is obvious that an almost completely ocean-covered planet "water world" has the highest likelihood of being habitable [78, 79]. For a central star with different luminosity and effective temperature, the limits of the HZ have to be rescaled to

$$R' = R\sqrt{\frac{S_{\mathrm{eff}}(T_{\mathrm{eff}})}{L/L_\odot}} \tag{11.17}$$

where $S_{\mathrm{eff}}(T_{\mathrm{eff}})$ describes the influence of the effective temperature T_{eff} of the centrals star on the inner and outer boundary [42].

In Fig. 11.10 the intervals for orbital instability of an earth-like planet is plotted for the 86 extrasolar planetary systems. For comparison, they are rescaled

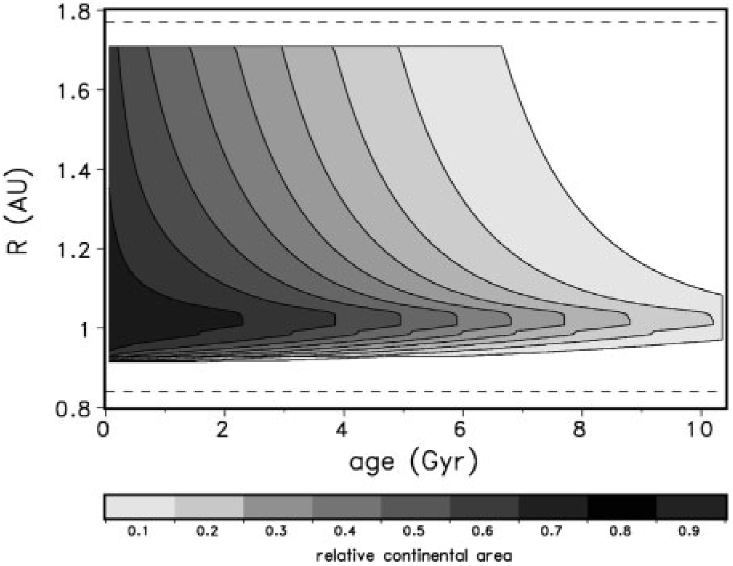


Fig. 11.9 The habitable zone around a central star for a luminosity of $L = 1\mathrm{L}_\odot$ and fixed effective temperature of the central star ($T_{\text{eff}} = 5700$ K) as a function of the age of the earth-like planet. The shaded areas indicate the extent of the HZ for different relative continental areas. The horizontal dashed lines indicate the HZ defined only by climatic constraints for the intermediate case [42].

to an equivalent solar distance R' and the HZ from Fig. 11.9 is plotted additionally. The range of dynamic habitability can be easily derived by excluding the horizontal bars from the HZ for a certain relative continental area.

We can distinguish four different classes for the 86 extrasolar planetary systems:

1. The HZ does not intersect with the intervals of orbital instability. An earth-like planet can be dynamic habitable in the entire HZ. 29 extrasolar planetary systems belong to this class.

2. The HZ does partially intersect with the intervals of orbital instability. 25 extrasolar planetary systems belong to this class.

3. The giant planet is located inside the HZ. A stable orbit of an earth-like planet does not exist around the central star. However, a moon around the giant planet can be habitable at its surface [76] or even a massive Trojan planet can be stable in 1:1 mean motion resonance. In particular, Schwarz et al. [80] have shown that in HD28185 (class 3) an earth-like planet can be stable on such an orbit. 20 extrasolar planetary systems belong to this class.

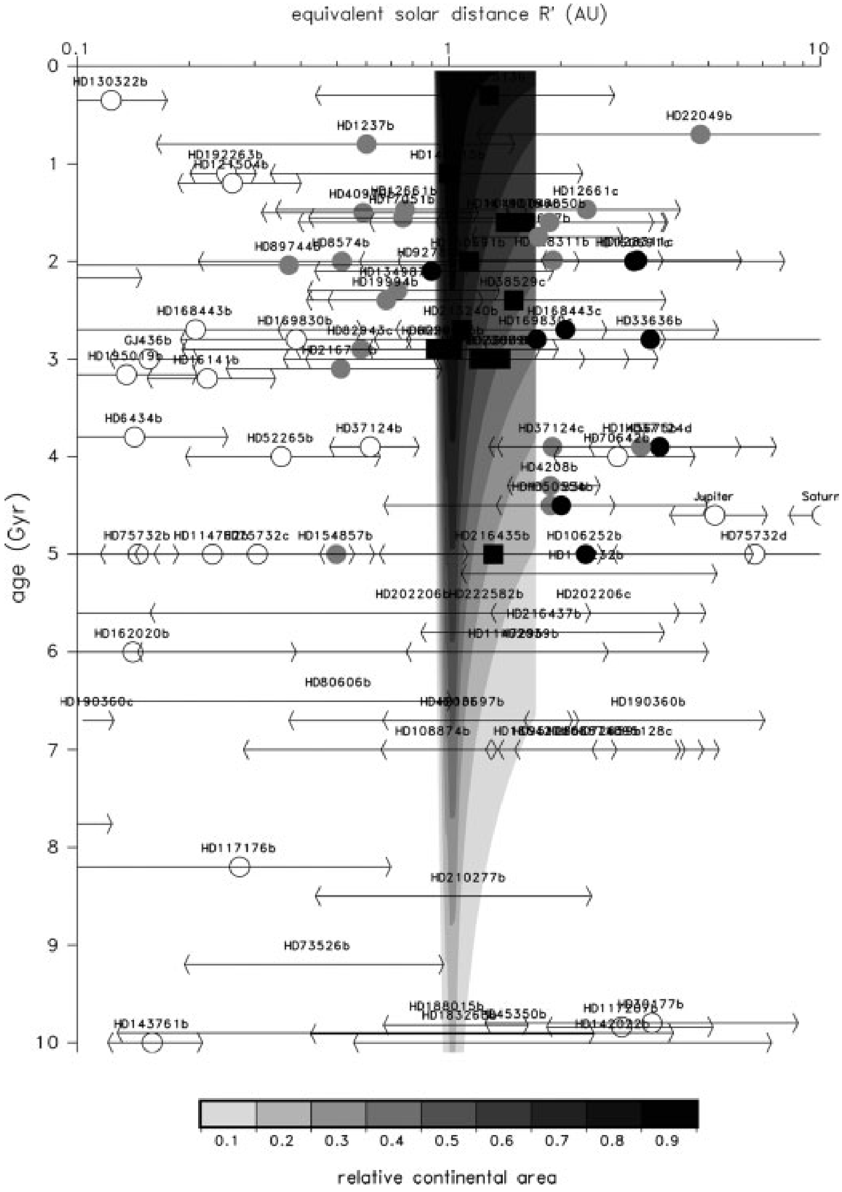


Fig. 11.10 The intervals of unstable orbits in the vicinity of giant planets of the solar system and of extrasolar planetary systems. The intervals are rescaled to an equivalent solar distance $R' = R\sqrt{S_{\text{eff}}L_{\odot}/L}$. The shaded areas indicate the extent of the HZ for different relative continental areas. Open circles denote giant planets not limiting dynamic habitability (class 1), gray shaded circles denote giant planets partially limiting dynamic habitability (class 2), black shaded circles denote planets excluding dynamic habitability (class 4), while shaded squares denote giant planets with habitable moons or Trojan planets (class 3).

4. The giant planet is outside the HZ and the interval of unstable orbits fully covers the HZ. No habitable earth-like planet is possible on a stable orbit. 12 extrasolar planetary systems belong to this class.

The likelihood that an earth-like planet is both on a stable orbit and also within the HZ can be quantitatively estimated from the width of the HZ excluding the interval of orbital instability:

$$\Delta R = \max(HZ \setminus \{R_{\text{int},i}, R_{\text{ext},i}\}_{i=1,n_p}) - \min(HZ \setminus \{R_{\text{int},i}, R_{\text{ext},i}\}_{i=1,n_p}) \quad (11.18)$$

where n_p is the number of detected planets in the extrasolar planetary system and $R_{\text{int},i}$ and $R_{\text{ext},i}$ are the inner and outer limits for unstable orbits, respectively. The widths ΔR of the dynamic habitable zones are shown in Fig. 11.11 ordered by the width of the HZ for the 55 extrasolar planetary systems with $\Delta R > 0$ for different relative continental areas. For comparison, the width of the dynamic HZ for the solar system is additionally plotted. We find that 18 systems have a larger width of the HZ than the solar system. These systems are characterized by a more luminous star, hosting a hot Jupiter, not perturbing the orbit of an earth-like planet. On the other hand, rather old systems like, e.g., 70 Vir, HD117207 and ρ CrB have much smaller HZs than the solar system. The habitable zone differs significantly from the HZ defined only by climatic constraints for the intermediate case (0.84AU, 1.77AU). In between are systems with a giant planet outside the HZ partially destroying stability inside the HZ. Two examples are the systems 47 UMa and ε Eri.

11.6
Summary

An earth-system model has been applied to study the habitability of an earth-like planet in the solar system and in extrasolar planetary systems with a different mass and age of the central star. Habitability is defined by a nonvanishing biological productivity of a photosynthesis-driven biosphere. The most prominent features observed in this approach are the age-dependent narrowing and the final disappearance of the HZ. Furthermore, we studied the principle possibility of orbitally stable earth-like habitable planets in 86 extrasolar systems. In particular, we considered earth-like planets with different ratios of land / ocean coverages and found that in 54 out of 86 systems (63%) potential dynamic habitable earth-like planets can exist (class 1 + 2). In 20 systems the giant is located within the HZ and could at least harbor a habitable moon or a Trojan planet (class3). Only in 12 systems (14%) is there no chance for the existence of dynamic habitable earth-like planets (class 4). According to our results the solar system is a relatively ordinary planetary system as shown in Fig. 11.11. There are at least 18 extrasolar planetary systems with better pre-

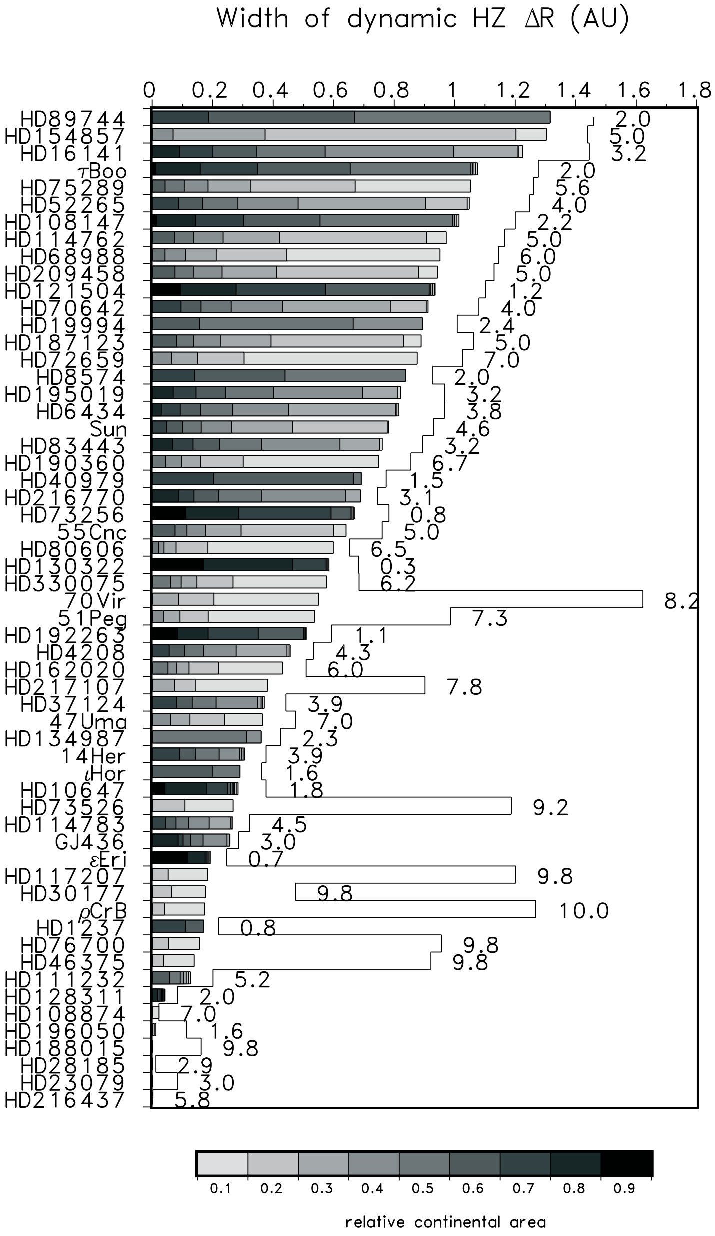


Fig. 11.11 Widths, ΔR, of the orbital range both warranting habitability and orbital stability for the solar system and extrasolar planetary systems (class 1 and 2) for different continental areas. The solid curve denotes the results for the HZ defined only by climatic constraints for the intermediate case [42]. The numbers at the horizontal bars are the corresponding ages of the planetary systems.

requisites to harbor dynamic habitable earth-like planets. This supports the so-called "Principle of Mediocrity." This principle proposes that our planetary system and life on earth are about average and that life will develop by the same rules wherever the proper conditions and the needed time are given [81].

Comparing our results based on the earth-system model with those of the HZ determination following Kasting et al. [42] there are, at least in some cases, remarkable differences. In particular, in our approach rather old extrasolar planetary systems are less probable for hosting dynamic habitable planets. On the other hand, the results of both approaches agree roughly if one considers water worlds with ages less than about 7 Gyr. Our results for land worlds, however, are less optimistic. It must be pointed out that two planets might follow different evolutions due to initial mass and composition. One planet at a given orbital distance might still be habitable at an age t while another planet with different initial characteristics might not. Even more important than different masses and compositions is the existence of plate tectonics.

In our study we neglected the influence of strong x-rays and shorter wavelengths and possible dense plasma fluxes. These effects have to be taken into account for central stars in the low-mass range of 0.5–0.7 $M_\odot$. Planets inside the habitable zone of M and K stars are tidally locked and may not develop plate tectonics. Due to the tidal locking, a weaker intrinsic magnetic field is expected and coronal mass ejections can compress the magnetospheres and erode the upper atmosphere. Furthermore, low-mass stars have strong XUV irradiations during longer time periods. Inside the habitable zone such radiations can erode the atmosphere of an earth-size planet and result in an additional limitation on the definition of habitability [82]. A detailed discussion about these effects can be found in Chapter 6 of this book. Only three stars out of the 86 stars with known age, however, have a mass lower than 0.7 $M_\odot$. Future studies focusing on the dynamic habitability planetary systems around M and K stars should include short-wave radiation effects.

Based on the facts that some parameters in the earth-system model are poorly constrained and based on processes that are not well understood, we get at least qualitative trends distinct from studies based on a climatic definition of habitability. The ultimate test of the presented methodology will be given by the forthcoming TPF/Darwin space missions of NASA/ESA, which try to detect biosignatures of earth-like planets in extrasolar planetary systems [83]. Direct detection of remote life might verify the model forecasts of habitability. On the other hand, the absence of any signs of life could be explained by the additional limitations of habitability caused by the geodynamic effects.

References

1 McKay, D. S., Gibson, E. K., Thomas-Keprta, K. L., Vali, H., Rommanek, C.S., Clemett, S.J., Chillier, D. F., Maechling, C. R., and Zare, R. H.: 1996, Search for life on Mars: possible relic biogenic activity in Martian meteorite ALH84001, *Science*, **273**, 924–930.

2 Lovelock, J. E.: 1965, A physical basis for life detection experiments, *Nature*, **207** , 568–570.

3 Lovelock, J. E.: 1975, Thermodynamics and the recognition of alien biospheres, *Proc. R. Soc. Lond. B.*, *189*, 167–181.

4 Hitchcock, D. R. and Lovelock, J. E.: 1967, Life detection by atmospheric analysis, *Icarus*, **7**, 140–150.

5 Lovelock, J. E. and Margulis, L.: 1972, Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis, *Tellus*, **26**, 2–14.

6 Lovelock, J. E.: 1987, *The Ages of Gaia: A Biography of Our Living Earth*, Oxford University Press, Oxford.

7 Schneider, S. H., Miller, J. R., Crist, E., and Boston, P. J. (eds.): 2003, *Scientists Debate Gaia*, MIT Press, Cambridge (MA).

8 Franck, S., Block, A., von Bloh, W., Bounama, C., Schellnhuber, H.-J., and Svirezhev, Y.: 2000a, Reduction of biosphere life span as a consequence of geodynamics, *Tellus*, **52B**, 94–107.

9 Franck, S., von Bloh, W., Bounama, C., Steffen, M., Schönberner, D., and Schellnhuber, H.-J.: 2000b, Determination of habitable zones in extrasolar planetary systems: Where are Gaia's sisters? *J. Geophys. Res.*, **105(E1)**, 1651–1658.

10 Kippenhahn, R. and Weigert, A.: 1990, *Stellar Structure and Evolution*, Springer, Berlin, Heidelberg.

11 Sackmann, I.-J., Boothroyd, A. I., and Kraemer, K. E.: 1993, Our Sun. III. Present and future, *ApJ*, **418**, 457–468.

12 Schellnhuber, H.-J.: 1999, "Earth system" analysis and the second Copernican revolution, *Nature*, **402**, Supp C19–C23.

13 Franck, S., Kossacki, K., and Bounama, C.: 1999, Modelling the global carbon cycle for the past and future evolution of the earth system, *Chem. Geol.*, **159**, 305–317.

14 Arrhenius, A.: 1896, On the influence of carbonic acid on the air upon the temperature of the ground, *Philos. Mag.*, **41 (5)**, 237–276.

15 Gough, D.O.: 1981, Solar interior structure and luminosity variations, *Sol. Phys.*, **74**, 21–34.

16 Williams, D. M.: 1998, *The Stability of Habitable Planetary Environments*, PhD Dissertation, Pennsylvania State University, University Park.

17 Walker, J. C. G., Hays, P. B., and Kasting, J. F.: 1981, A negative feedback mechanism for the long-term stabilization of Earth's surface temperature, *J. Geophys. Res.*, **86**, 9776–9782.

18 Berner, R. A.: 1993, Paleozoic atmospheric CO_2: importance of solar radiation and plant evolution, *Science*, **262**, 68–70.

19 Berner, R. A., Lasaga, A. C., and Garrels, R. M.: 1983). The carbonate–silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years, *Am. J. Sci.*, **283**, 641–683.

20 Lasaga, A. C., Berner R. A and Garrels, R. M.: 1985, An improved geochemical model of atmospheric CO_2 fluctuations over past 100 million years, in: Sundquist, E. T. and Broecker, W. S. (eds.) *The Carbon Cycle and Atmospheric CO_2: Natural Variations Archaean to Present*, AGU, Washington DC, pp 397–411.

21 von Bloh, W., Bounama, C., and Franck, S.: 2003a, Cambrian explosion triggered by geosphere–biosphere feedbacks. *Geophys Res. Lett.*, **30(18)**, 1963–1967.

22 Franck, S., Bounama, C. and von Bloh, W.: 2006, Causes and timing of future biosphere extinctions, *Biogeosciences*, **3**, 85–92.

23 Tajika, E. and Matsui, T.: 1992, Evolution of terrestrial proto-CO_2 atmosphere coupled with thermal history of the Earth. *Earth Planet. Sci. Lett.*, **113**, 251–266.

24 Stumm, W. and Morgan, J. J.: 1981, *Aquatic Chemistry*, Wiley, New York.

25 Volk, T.: 1987, Feedbacks between weathering and atmospheric CO_2 over the last 100 million years, *Am. J. Sci.*, **287**, 763–779.

26 Larcher, W.: 1995, *Physiological Plant Ecology: Ecophysiology of Functional Groups*, Springer-Verlag, New York.

27 Pearcy, R. W. and Ehleringer, J.: 1984, Comparative ecophysiology of C_3 and C_4 plants, *Plant Cell Environ.*, **7**, 1–13.

28 Caldeira, K. and Kasting, J. F.: 1992, The life span of the biosphere revisited, *Nature*, **360**, 721–723.

29 Kasting, J. F.: 1984, Comments on the BLAG model: the carbonate–silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years, *Am. J. Sci.*, **284**, 1175–1182.

30 Condie, K. C.: 1990, Growth and accretion of continental crust: inferences based on Laurentia. *Chem. Geol.*, **83**, 183–194.

31 Franck, S.: 1998, Evolution of the global heat flow over 4.6 Gyr, *Tectonophysics*, **291**, 9–18.

32 Turcotte, D. C. and Schubert, G.: 1982, *Geodynamics*, John Wiley, New York.

33 Franck, S. and Bounama, C.: 1995, Effects of water dependent creep rate on the volatile exchange between mantle and surface reservoirs, *Phys. Earth Planet. Inter.*, **92**, 57–65.

34 Huang, S. S.: 1959, Occurrence of life in the universe, *Sci. Am.*, **47**, 397–402.

35 Huang, S. S.: 1960, Life outside the solar system, *Sci. Am.*, **202 (4)**, 55–63.

36 Dole, S. H.: 1964, *Habitable Planets for Man*, Blaisdell, New York.

37 Shklovskii, I. S. and Sagan, C.: 1966, *Intelligent Life in the Universe*, Holden Day, San Francisco.

38 Hart, M. H.: 1978, The evolution of the atmosphere of the Earth, *Icarus*, **33**, 23–39.

39 Hart, M. H.: 1979, Habitable zones about main sequence stars, *Icarus*, **37**, 351–357.

40 Kasting J. F.: 1988, Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus, *Icarus*, **74**, 472–494.

41 Kasting, J. F.: 1997, Habitable zones around low mass stars and the search for extraterrestrial life, *Origins of Life*, **27**, 291–307.

42 Kasting, J. F., Whitmire, D. P., and Reynolds, R. T.: 1993, Habitable zones around main sequence stars, *Icarus*, **101**, 108–128.

43 Golombek, M. P.: 1999, A message from warmer times, *Science*, **283**, 1470–1471.

44 Malin, M. C. and Edgett, K. S.: 2000, Evidence for recent groundwater seepage and surface runoff on Mars, *Science*, **288**, 2330–2335.

45 Forget, F. and Pierrehumbert, R. T.: 1997, Warming early Mars with carbon dioxide clouds that scatter infrared radiation, *Science*, **278**, 1273–1276.

46 Schellnhuber, H.-J. and Kropp, J.: 1998, Geocybernetics: Controlling a complex dynamical system under uncertainty, *Naturwissenschaften*, **85**, 411–425.

47 Schwartzman, D. W.: 1999, *Life, Temperature and the Earth: The Self-organizing Biosphere*, Columbia University Press, New York.

48 Franck, S., Block, A., von Bloh, W., Bounama, C., Schellnhuber, H.-J., and Svirezhev, Y.: 2000c) Habitable zone for Earth-like planets in the solar system, *Planet. Space Sci.*, **48**, 1099–1105.

49 Head, J. W., Mustard, J. F., Kreslavsky, M. A., Milliken, R. E., and Marchant, D. R.: 2003, Recent ice ages on Mars, *Nature*, **436**, 797–802.

50 Jakosky, B. M., Haberle, R. M., and Arvidson, R. E.: 2005, The changing picture of volatiles and climate on Mars, *Science*, **310**, 1439–1440.

51 Leovy, C.: 2001, Weather and climate on Mars, *Nature*, **412**, 245–249.

52 Connerney, J. E. P., Acuna, M. H., Wasilewski, P. J., Ness, N. F., Reme, H., Mazelle, C., Vignes, D., Lin, R. P., Mitchell, D. L., and Cloutier, P. A.: 1999, Magnetic lineations in the ancient crust of Mars. *Science*, **284**, 794–798.

53 New Scientist: 2001, The day the dynamo died, *New Scientist*, **169 (No. 2277)**, 4.

54 Iben, I.: 1967a, Stellar evolution within and off the main sequence, *ARA&A.*, **5**, 571–626.

55 Iben, I.: 1967b, Stellar evolution. VI. Evolution from the main sequence to the red-giant branch for stars of mass 1 $M_\odot$, 1.25 $M_\odot$, and 1.5 $M_\odot$, *ApJ*, **147**, 624–649.

56 Schaller, G., Schaerer, D., Meynet, G., and Meader, A.: 1992, New grids of stellar models from 0.8 to 120 solar masses at Z=0.020 and Z=0.001, *A&A. Suppl. Ser.*, **96(2)**, 269–331.

57 Schwartzman, D. W., Mc Menamin M., and Volk, T.: 1993, Did surface temperatures constrain microbial evolution? *Bio. Sci.*, **43**, 390–393.

58 Jakosky, B. M.: 1998, *The Search for Life on Other Planets*, Cambridge University Press, Cambridge.

59 Joshi, M. M., Haberle, R. M., and Reynolds: 1997, Simulations of atmospheres of synchronously rotating terrestrial planets orbiting M dwarfs: Conditions for atmospheric collapse and the implications for habitability, *Icarus*, **129**, 450–465.

60 Beaulieu, J. P. et al.: 2006, Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing, *Nature*, **439**, 437–440.

61 Dvorak, R., Pilat-Lohinger, E., Funk, B., and Freistetter, F.: 2004, Extrasolar Trojan planets close to habitable zones, *A&A.*, **426**, L37–L40.

62 Marcy, G. W., Butler, R. P., Fischer, D. A., and Vogt, S. S.: 2003, Properties of extrasolar planets, in: Deming, D., Saeger, S.: (eds.), *Scientific Frontiers in Research on Extrasolar Planets*, ASP Confer. Ser. CS-294, pp. 1–16.

63 Marcy, G. W., Butler, R. P., Fischer, D., Vogt, S., Wright, J. T., Tinney, C. G., and Jones, H. R. A.: 2005, Observed properties of exoplanets: masses, orbits, and metallicities, *Progress Theoret. Phys. Suppl.*, **158**, 24–42.

64 McArthur, B., Endl, M., Cochran, W., Benedict, F., Fischer, D. A., Marcy, G. W., Butler, R. P., Naef, D., Mayor, M, Queloz, D., Udry, S., and Harrison, T. E.: 2004) Detection of a Neptune-Mass planet in the ρ^1 Cancri system using the Hobby-Eberly telescope, *ApJ*, **614**, L81–L84.

65 Santos, N. C., Benz, W., and Mayor, M.: 2005, Extrasolar planets: constraints for planet formation models, *Science*, **310**, 251–255.

66 Lineweaver, C. H. and Grether, D.: 2003, What fraction of Sun-like stars have planets? *ApJ*, **598**, 1350–1360.

67 Jones, B. W., Sleep, P. N., and Chambers, J. E.: 2001, The stability of the orbits of terrestrial planets in the habitable zones of known exoplanetary systems, *A&A.*, **366**, 254–262.

68 Asghari, N. et al.: 2004, Stability of terrestrial planets in the habitable zone of Gl 777 A, HD 72659, Gl 614, 47 Uma and HD 4208, *A&A*, **426**, 353–365.

69 Goździewski, K.: 2002, Stability of the 47 UMa planetary system, *A&A.*, **393**, 997–1013.

70 Menou, K. and Tabachnik, S.: 2003, Dynamical habitability of known extrasolar planetary systems, *ApJ*, **583**, 473–488.

71 Noble, M., Musielak, Z. E., and Cuntz, M.: 2002, Orbital stability of terrestrial planets inside the habitable zones of extra-solar planetary systems, *ApJ*, **572**, 1024–1030.

72 Chambers, J. E.: 1999, A hybrid symplectic integrator that permits close encounters between massive bodies, *MNRAS*, **305**, 793–799.

73 Pilat-Lohinger, E. and Dvorak, R.: 2002, Stability of S-type orbits in binaries, *CMDA*, **82**, 143–153.

74 Hanslmeier, A. and Dvorak, R.: 1984, Numerical integration with Lie series. *A&A.*, **132**, 203–207.

75 Jones, B. W., Underwood, D. R., and Sleep, P. N.: 2005, Prospects for habitable "Earth" in known exoplanetary systems, *ApJ*, **622**, 1091–1101.

76 Williams, D. M. and Pollard, D.: 2002, Earth-like worlds on eccentric orbits: excursions beyond the habitable zone, *Int. J. Astrobiol.*, **1(1)**, 61–69.

77 Espresate, J.: 2005, Catalog of 156 confirmed extrasolar planets and their 133 parent stars, *Astro-ph/0508317 v1 15 Aug 2005*, pp. 13.

78 Franck, S., Cuntz, M., von Bloh. W., and Bounama C.: 2003, The habitable zone of Earth-mass planets around 47 UMa: results for land and water worlds, *Int. J. Astrobiol.*, **2(1)**, 35–39.

79 von Bloh, W., Cuntz, M., Franck, S., and Bounama, C.: 2003b, On the existence of habitability around 55 Cancri, *Astrobiology*, **3(4)**, 681–688.

80 Schwarz, R., Pilat-Lohinger, E., Dvorak, R., Erdi, B., and Sandor, Z.: 2005, Trojans in habitable zones, *Astrobiology*, **5**, 579–586.

81 Darling, D. J.: 2001, *Life Everywhere: The Maverick Science of Astrobiology*, Basic Books, New York.

82 Grießmeier, J. M., Stadelmann, A., Motschmann, U., Belisheva, N.K., Lammer, H., and Biernat, H. K.: 2005, Cosmic ray impact on extrasolar Earth-like planets in close-in habitable zones. *Astrobiology*, **5**, 487–603.

83 Des Marais, D. J., Harwitt, M. O., Jucks, K. W., Kasting, J. F., Lin., D. N. C., Lunine, J.I., Schneider, J., Seager, S., Traub, W. A., and Woolf, N. J.: 2002, Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets. *Astrobiology*, **2**, 153–181.

Index

Extrasolar Planets: Formation, Detection and Dynamics. Edited by Rudolf Dvorak

ISBN: 978-3-527-40671-5

***Rudolf Dvorak** is a renowned expert in celestial body dynamics. He is now serving at the Institute of Astronomy in Vienna, has gained research experience in planet research at international institutions including the 'Observatoire de Paris', the University of Florida in Gainesville, and the University of Sao Paulo. His connections allowed him to assemble the team of leading authors to cover all aspects of the detection of distant worlds.*

The extrasolar planets are one of the research focuses of modern astronomy. Their detection in the close vicinity of their brilliant central stars requires the highest possible skills in observational accuracy, instrument performance and data evaluation. Special emphasis is given to the dynamical stability of these systems and the possibility of terrestrial planets in habitale zones. The investigation of those objects yield new insights into the physics and chemistry of planets in general, hence into the origin of our home world.

Detection methods employ the observation of various parameters of stellar light as well as of the evolving planets themselves.

This work covers the evolution and dynamics of planets as well as their detection now and in the future.

From the Contents:

- Planetary Masses and Orbital Parameters from Radial Velocity Measurements
- Terrestrial Planets in Extrasolar Planetary Systems
- Mission Requirements: How to Search for Extrasolar Planets
- Biomarkers Set in Context
- The Formation of Resonant Planetary Systems
- Impact of Stellar Activity on the Evolution of Planetary Atmospheres and Habitability
- Dynamics of the Extrasolar Planetary Systems
- Planets in Double Stars
- The Transit Method
- US Programs and Space Missions for Extrasolar Planet Research
- Habitable Zones in Extrasolar Planetary Systems

ISBN 978-3-527-40671-5

www.wiley-vch.de