Analysis of Exoplanetary Transit Light Curves

by

Joshua Adam Carter

B.S., Physics, University of North Carolina at Chapel Hill, 2004
B.S., Mathematics, University of North Carolina at Chapel Hill, 2004

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

Author ....
Department of Physics
August 31, 2009

Certified by ....
Joshua N. Winn
Assistant Professor of Physics
Class of 1942 Career Development Professor
Thesis Supervisor

Accepted by ....
Thomas J. Greytak
Lester Wolfe Professor of Physics
Associate Department Head for Education

Analysis of Exoplanetary Transit Light Curves

by

Joshua Adam Carter

Submitted to the Department of Physics
on August 31, 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This Thesis considers the scenario in which an extra-solar planet (*exo*planet) passes in front of its star relative to our observing perspective. In this event, the light curve measured for the host star features a systematic drop in flux occurring once every orbital period as the exoplanet covers a portion of the stellar disk. This exoplanetary transit light curve provides a wealth of information about both the planet and star. In this Thesis we consider the transit light curve as a tool for characterizing the exoplanet. The Thesis can divided into two parts.

In the first part, comprised of the second and third chapters, I assess what observables describing the exoplanet (and host) may be measured, how well they can be measured, and what effect systematics in the light curve can have on our estimation of these parameters. In particular, we utilize a simplified transit light curve model to produce simple, analytic estimates of parameter values and uncertainties. Later, we suggest a transit parameter estimation technique that properly treats temporally correlated stochastic noise when determining *a posteriori* parameter distributions.

In the second part, comprised of the fourth and fifth chapters, I direct my attention to real exoplanetary transit light curves, primarily for two exoplanets: HD 149026b and HD 189733b. We analyze four transits of the ultra-dense HD 149026b, as measured by an instrument on the *Hubble Space Telescope*, in an effort to properly constrain the stellar and exoplanetary radius. In addition, we assess a detection of strong, wavelength dependent absorption, possibly due to an unusual atmospheric composition. For HD 189733b, we utilize seven ultra-precise *Spitzer Space Telescope* transit light curves in an effort to make the first empirical measurement of asphericity in an exoplanet shape. In particular, we constrain the parameters describing an oblate spheriod shape for HD 189733b and, attributing oblateness to rigid-body rotation, we place lower bounds on the rotation period of the exoplanet.

Thesis Supervisor: Joshua N. Winn
Title: Assistant Professor of Physics
Class of 1942 Career Development Professor

Acknowledgements

Many people contributed to this work and to the completion of this degree.

I want to thank both my parents for instilling in me the virtues of perseverance, self-reliance, and humility which have time and again helped me succeed at any endeavor. I want to thank my Dad for being so proud of every good thing I have done, no matter how small. Dad, thanks for all the life lessons and for sharing all those baseball games together. I want to thank my Mom for shaping my mind from an early age to succeed in academics, science and life. Mom, thanks for the love and Legos. I also want to thank my siblings, Troy, Jason and Stephanie, for all their support and guidance.

I want to thank my surrogate parents, John and Janet Mustonen, for accepting me so completely into their family. Over the years they have given me tremendous support ranging from great advice to good company. I also want to thank Pete, my brother-in-law, for all his help and for bravely subjecting himself to endless good-natured competition despite my clear intellectual edge.

To my Boston family, Carly, Sarah, Stephanie and Eric, thanks for making my life away from MIT particularly awesome.

To all my current and past officemates of 37-602, thanks for making every day at work fun, even if no work was getting done. A special thanks goes to Ryan Lang whose friendship over the years left an indelible mark on my career and life as a graduate student at MIT.

Thanks to my advisor, Josh Winn, for making my last few years of graduate school exciting, stimulating and an excellent learning experience. I also want to thank Jean Papagianopoulos, Arlyn Hertz and the rest of the staff at MKI and the physics department for helping me so much over my time as a graduate student.

Most of all, I want to thank my perfect wife Erica for being so much more than just my wife; for being my drive, my motivation, and my purpose for reaching this goal. Without her, all of this work would be pointless.

Thank you Erica for being so patient through all my ups and downs during my

graduate career. Thank you for taking such good care of me. Thank you for being so beautiful even without trying. Thank you for making me feel so special and loved; I honestly feel so very lucky to have met and married you. I dedicate this work and, much more importantly, all the good I can offer in my entire life to you and our family to come. I love you so very much.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the burning questions of astronomy deals with frequency of planet-like bodies in the galaxy which belong to stars other than the Sun.
- Otto Struve (1952)

In looking over the long history of human science from time immemorial to our own times, it is impossible to overestimate the role played in it by the phenomena of eclipses of the celestial bodies both within our solar system as well in the stellar universe at large.
- Zdenek Kopal (1990)

1.1 Planets near and far

1.1.1 Our own Solar System

It was realized early in recorded history that, looking at the night sky, amongst the "fixed" stars in the "heavenly firmament" a group of wandering objects traced repeatable paths on the celestial sphere. These planets (literally "wanderers" in Greek), initially regarded as the physical manifestations of powerful mythological gods, were, in fact, worlds in many ways like the Earth, likely arriving from the same evolutionary process that gave birth to our common stellar host Sol. Upon closer inspection, famously first by Galileo's pioneering work identifying the moons of

Jupiter and the phases of Venus, each planet is found to be remarkably distinct from its siblings. In order of increasing semi-major orbital distance, the interior planets Mercury, Venus, Earth and Mars are small rocky worlds, while Jupiter and Saturn are gas giants lacking any substantial rocky core, and finally, Uranus and Neptune are "ice giants" having mean densities lying in between that of the terrestrial and Jovian worlds [see Carrol & Ostlie (2006) Part III for an excellent review of the Solar System][1]. Most planets are also accompanied by a collection of natural satellites (and now even artificial satellites) in the form of moons and diffuse rings. Each planetary system is, in its own right, a complicated and rich dynamical collection of gravitationally bound objects. Humanity has gone to extensive investigative lengths to classify, explain, or simply photograph these worlds with complex (and expensive) experiments including a series of manned (in the case of the Moon) and unmanned spacecraft missions right to the source. We work towards both an explanation of the planets in isolation from the remaining planets and as a Solar System as a whole. In particular, we ask the questions that relate most to our own planet's existence: how do these other planets (and moons) compare to Earth? Is Earth an atypical object in this small sample? Could other planets in our Solar System harbor their own form of life? Intelligent life? The final two questions are likely to induce the strongest inquisitive response from even the most uninformed, given that the answers to these questions will no doubt shed light on our very *relevance* in the universe.

1.1.2 Extrasolar planets: Planets outside our own Solar System

The most natural question following the above line of questioning is, given the preponderance of Sun-like stars in our own galaxy, how many planets are there that orbit stars other than our Sun? And assuming this answer is non-zero (yes, it is) are there multiple planets orbiting a single star other than our Sun? Going further, we may ask: How many of these extrasolar systems contain Jovian-type planets? Ice

---

[1] In August 2008, the International Astronomical Union (IAU) defined the term "planet." Unfortunately Pluto, previously the ninth planet in our Solar System, did not make the cut.

giant-like? Terrestrial planets? Earth-like? Do these planets have rings? Moons? Atmospheres? Life? We as a research community, at the time of writing, have at least some idea of the answers to many of these questions.

Extrasolar planets, or *exo*planets in the parlance of the field, number in the hundreds (353, as of July 2009). However, prior to 1995 [and the discovery of 51 Peg b by Mayor & Queloz (1995)], we only knew of the 9 Solar System planets (reduced now to 8, see footnote). The expectation for discovery was in place, as suggested in the short note by Struve (1952) in which the possibility of detection was first appreciated. The next section reviews Struve's recommended method of detection and other techniques, including the transit. Since 1995, however, the pace of discovery has steadily grown. In Figure (1-1) we show the number of exoplanets discovered by year, since 1995. In particular, the pace of discovery of planets that transit their stellar host (see § 1.2.2 below) has recently reached a doubling time that is less than one year [Charbonneau et al. (2009)].

As the number of exoplanets with precisely measured properties (see § 1.3) grows, we find ourselves on the frontier of a realm in which statistically meaningful generalizations may be drawn of planets as a whole. Homogeneous analyses of these precisely characterized systems (Torres et al. 2008, Southworth 2008) demonstrate trends in the parameter space [see Figure (1-2)] allowing us to reach somewhat-informed conclusions about the population of exoplanets yet to be discovered. However, in many ways, we are still very far from a complete understanding. But the prospects look good: initial estimates based upon our current sample of exoplanets imply that nearly 6% of stars harbor at least one giant planet within 4 AU. With this statistic as motivation, we need to (1) find more planets and (2), in the pursuit of a fundamental theory of planets, characterize these objects as accurately as possible.

While the work presented in this Thesis is geared more to the goal of point (2), we review the techniques relating to exoplanet discovery in the next section.

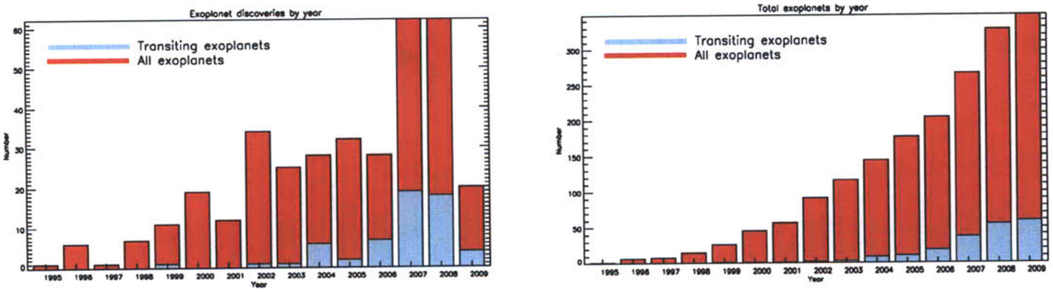

Figure 1-1 Exoplanet detections and totals by year. Data from `exoplanets.eu`, organized by J. Schneider.

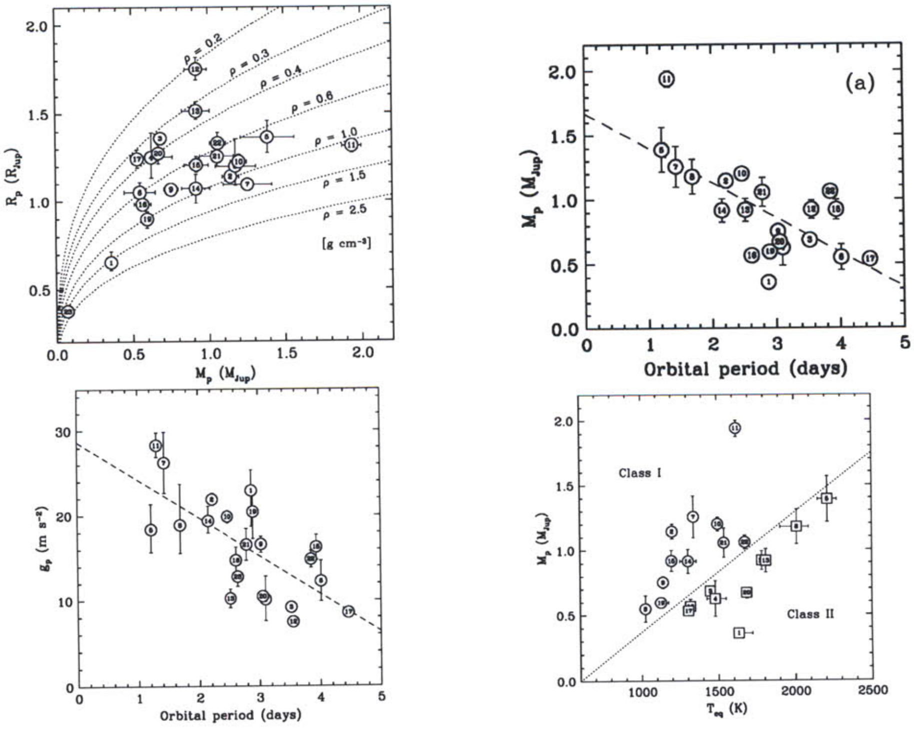

Figure 1-2 Exoplanet trends and correlations. Plotted are parameters determined in the homogeneous analysis by Torres et al. (2008) for a selection of exoplanets. In particular, correlations between exoplanet radius, R_p, mass M_p, orbital period P, surface gravity g_p and equilibrium temperature T_{eq} are shown. Figures by Torres et al. (2008); refer to their paper for details.

1.2 Detecting extrasolar planets

We must actually find exoplanets prior to attempting accurate parameter estimation. A number of techniques exist to detect exoplanets and may be organized into the following categories: (1) photometric (transits), (2) dynamical (primarily radial velocity, but also astrometry and timing), (3) microlensing, (4) direct imaging and (5) others (Perryman et al. 2005). Figure (1-3) organizes these detection techniques in a graphical manner indicating planetary mass detection limits. While this diagram is out-dated and statistics have changed [only 4 years old and missing hundreds of exoplanets that have been discovered since! See Fig. (1-1)], the top three techniques by total yield remain the same and in the following order: radial velocity (327 planets), transits (59 planets) and microlensing (7 planets). Each of these techniques has their respective advantages and disadvantages in terms of detection capability. We will show, in § 1.3, that while transit detection may be considered inferior to (or incomplete without) radial velocity detection, transit characterization of exoplanets is unrivaled. We will briefly describe the radial velocity technique before moving onto detection via transit. See Perryman et al. (2005) and the references therein for a discussion of alternate detection techniques (including microlensing).

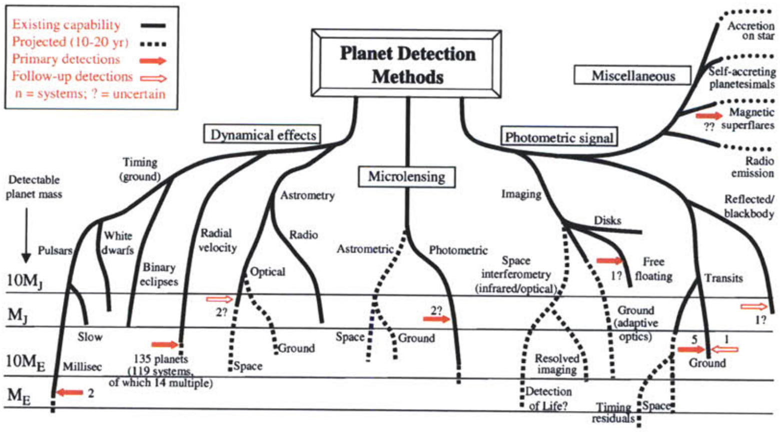

Figure 1-3 Exoplanet detection methods and yields as of 2005. Figure by Perryman et al. (2005).

1.2.1 Detection via radial velocity

The well-studied and understood two-body gravitational problem [see, e.g., Carrol & Ostlie (2006)] includes the basic prediction that each massive object moves in an elliptical orbit about the common center of mass. As a result, the velocity of the objects (in the center-of-mass frame, without loss of generality) oscillates at the orbital period, P, and each with a unique amplitude, $\hat{K}_i$, that depends on a formula involving the masses. In particular, if one of the objects is a star and the other is an unseen planet, we may infer the existence of the less massive component by monitoring the periodic signature on the stellar component's velocity [this was first appreciated by Struve (1952)]. We may measure the component of the velocity along the line of sight by using the principle of Doppler spectroscopy [see, for example, Butler et al. (1996)]. Here, information regarding the relative velocity along the radial direction is encoded in the spectral lines of the stellar spectrum as a result of Doppler frequency-shifting. One may obtain better than 3 m s^{-1} precision on the measurement of radial-velocity with a proper calibration of rest wavelengths of the spectral lines (and other instrumental calibrations, Butler et al. 1996). We may fit a model to the collection of radial velocity measurements to determine orbital parameters and masses. For a single planetary component, a simple Keplerian model will suffice. In particular, we may solve for the mass of the planetary object, M_p and orbital semi-major axis a,

$$
\begin{aligned}
M_p \sin i &= K\sqrt{1-e^2}\left[\frac{P\,(M_\star + M_p \sin i)^2}{2\pi G}\right]^{1/3} && (1.1)\\
\left(\frac{a}{1\ \mathrm{AU}}\right)^3 &= \left(\frac{M_\star + M_p \sin i}{M_\odot}\right)\left(\frac{P}{\mathrm{yr}}\right)^2, && (1.2)
\end{aligned}
$$

in terms of the stellar mass $M_\star$ and the inclination angle i of the orbital plane to the observational plane, where K is the amplitude of the radial velocity, e is the orbital eccentricity and P is the orbital period. The parameters K, e and P may be measured directly from the radial velocity data [e.g. Butler et al. 2006 and see Fig. (1-4)] . The stellar mass $M_\star$ may be precisely estimated via spectral identification, for example.

The inclination i is an unknown parameter degenerate with the planetary mass; we therefore can only estimate $M_p \sin i < M_p$.

The radial velocity exoplanet detection technique has several advantages. While the mass is degenerate with inclination i, only parallel orbital plane configurations ($i = 0°$) yield a non-detection. Current radial velocity exoplanet detection technology allows for the detection of exoplanet's with masses equal to a few times an Earth mass or less (the so-called "Super Earths"). Examples include the three orbiting HD 40307, with masses 4.2, 6.9, and 9.2 $M_\oplus$, found by Mayor et al. (2009) with the HARPS spectrograph (Pepe et al. 2002) at the La Silla Observatory in Chile. Currently, radial velocity is the detection method best suited to the detection of Earth-like analogs.

Exoplanet detection via radial velocity is unfortunately very costly in both dollars and time. Radial velocity measurements must uniformly sample the orbital phase in order to precisely measure orbital parameters of the unknown planetary component. For a Jupiter analog (5 AU from the Sun making one orbit every 12 years), it would take several years of observations to make detection possible. Radial velocity surveys are capable of tremendous yield for short-period giant exoplanets [i.e., 51 Peg b-like, Mayor & Queloz (1995)]. However, given the limited information that may be derived about the exoplanet and its orbit (namely $M_p \sin i$, e, P), radial velocity characterization of short-period giants has quickly diminishing returns.

1.2.2 Detection via transit

If the orbital plane of the exoplanetary system were to lie in the plane perpendicular to our observational plane ($i = 90°$), then the exoplanet will periodically pass in between its star and our telescopes. This fortunate configuration results in what is referred to as a transit[2]. The observational effect of transit is that the obscured star is perceived to experience a systematic decrease in total flux. For a planet in a stable

---

[2] The obscuration of one celestial body by another is referred to as an eclipse, in general, and is the subject of the general mathematical theory of Kopal et al. (1990) or as found in Mandel & Agol (2002). In practice, the word eclipse is reserved for the situation common to eclipsing stellar binaries where the two eclipsing components are of equal radial extent. If the object passing in front of the other from our perspective is significantly smaller (larger) than its companion then the eclipse is referred to as a transit (occultation).

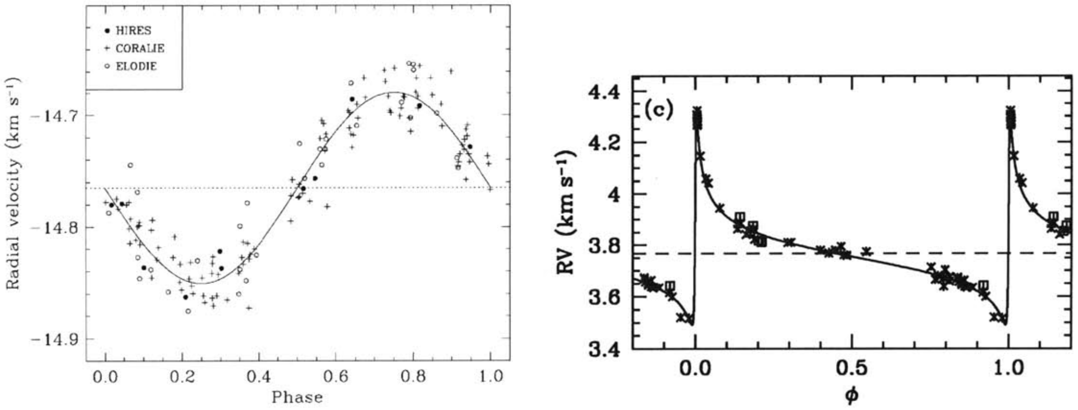

Figure 1-4 Exoplanet detection via radial velocity. The left panel shows the radial velocity signature of the exoplanet HD 209458b on HD 209458 [Figure by Mazeh et al. (2000)]. The right panel shows the radial velocity signature of HD 80606b on HD 80606 [Figure by Naef et al. (2001)]. The sharp features in the radial velocity curve for HD 80606 are as a result of the high orbital eccentricity of HD 80606b. See § 1.2.1 for details.

Keplerian orbit, the size and duration of the flux deficit are fixed. Additionally, the time between subsequent events is constant (equal to the orbital period, P). When the exoplanet passes behind the star from our perspective (so-called occultation), an analogous drop in total flux is observed, this time owing to the stellar obscuration of the planetary flux. The transit configuration is illustrated in Fig. (1-5) along with the "transit light curve," the dynamical curve describing the total flux measured for star and planet.

The depth δ of the transit is related to the fraction of area obscured by the exoplanet and, therefore, is related to the ratio of the radii of planet and star (see Chapter 5 for an alternate possibility). The spacing of subsequent transit events is directly related to the orbital period P. Thus, a photometric survey program can establish the existence of a transiting body with the observation of repeated, uniform drops in flux from a star. This was first realized in the same note as discussed above by Struve (1952) and considered further by Rosenblatt (1971) and Borucki & Summers (1984).

For a Jupiter sized planet transiting a Sun-like star, the expected deficit in flux, δ, should be about 1%. This precision may obtained for bright stars with modest

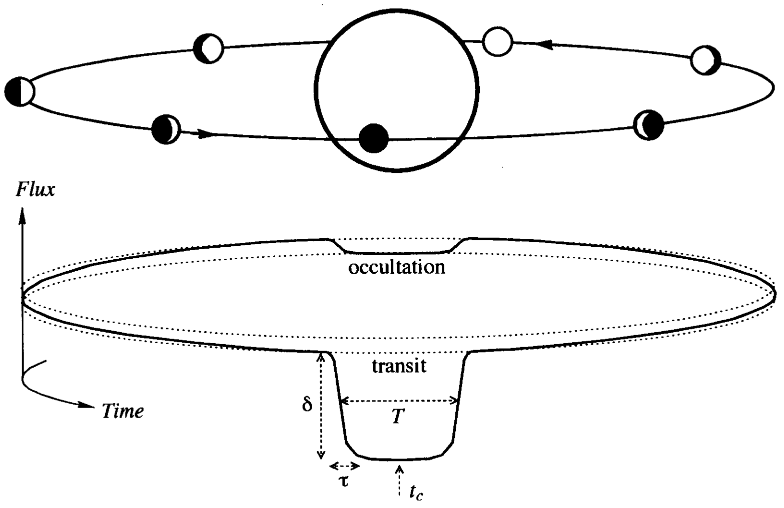

Figure 1-5 A transiting exoplanet configuration and transit light curve. Figure courtesy of J. Winn. The transit portion of the light curve can be minimally described by a depth, δ, transit duration, T and ingress or egress duration τ. See § 1.3 for details.

telescopes and imaging equipment. For this reason, a number of relatively cheap photometric surveys have sprung up for the purpose of exoplanet transit detection. See Perryman et al. (2005) for a comprehensive list of ground and space-based photometric surveys. The most "famous" of this collection are those with largest harvests. From the ground: OGLE [the first such survey, 7 planets, Udalski (2007)], TrES [4 planets, Alonso et al. (2004a)], XO [5 planets , McCullough et al. (2005)], HATNet [12 planets, Bakos et al. (2007)] and WASP [15 planets, Pollacco et al. (2006)]; from space: CoRoT [7 planets, Baglin et al. (2003), see Fig. (1-6)]. Recently, the space-based transit-survey mission *Kepler* (Borucki et al. 2009) was launched and has the potential for most transit-discoveries, likely out-pacing the radial velocity discovery rate.

One negative cost of transit discovery is based upon simple probability: given that planetary systems in the galaxy are likely to be randomly oriented, it is improbable to find a significant number of exoplanets whose chance alignment allows a transit from our perspective. We can quantify this probability via geometric analysis. Namely, the probability of observing a transit of a particular exoplanet is equal to the ratio of the solid angle in which the planet will be seen to transit and the total available

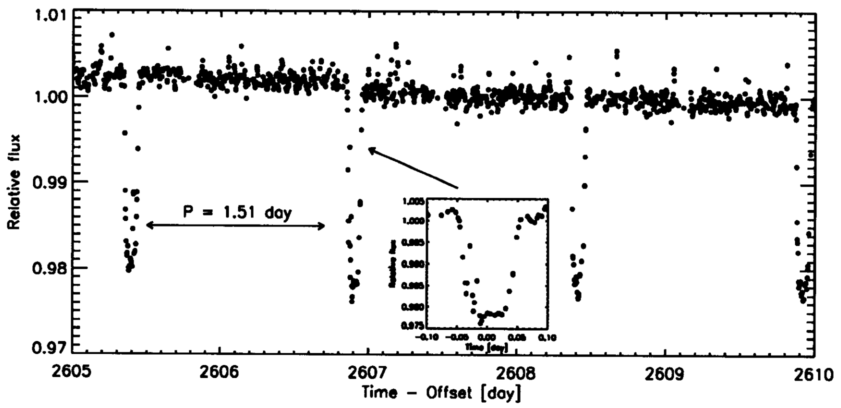

Figure 1-6 Detecting an exoplanet via transit. Plotted here is discovery data showing the transit of CoRoT-1b in successive transit epochs [Baglin et al. (2003)].

solid angle (4π). The probability of transit, $P_{\rm transit}$, is, therefore, mainly dependent on the ratio of the semi-major axis a and stellar radius $R_\star$ as

$$P_{\rm transit} = 0.0045 \left(\frac{1\ {\rm AU}}{a}\right) \left(\frac{R_\star + R_p}{R_\odot}\right) \left[\frac{1 + e\cos(\pi/2 - \omega)}{1 - e^2}\right] \tag{1.3}$$

where e and ω are the eccentricity and argument of periastron for the orbit and our line-of-sight, respectively (Charbonneau et al. 2007). For an Earth analog, the probability of transit is therefore $P_{\rm transit} \approx 0.45\%$. On the other hand, a Jupiter-sized planet orbiting a Sun-like star at 0.05 AU has a more palatable 10% transit probability. Before the identification of 51 Peg b-like "Hot Jupiters," or short-period gas giants, it was assumed, based upon our experience with our own Solar System, that transit surveys would be a low yield affair. Post 1995, as radial velocity survey yields implicated that fully 1% of nearby sun-like stars hosted these "Hot Jupiters," the interest in photometric surveys grew (Horne 2003).

The expectations from the research community for transit surveys were tremendous, with an anticipation of ~ 10 "Hot Jupiter" planets discovered per month for surveys with WASP-like characteristics (Horne 2003). That the current generation of surveys has not reached this rate is a testament to the additional difficulties associated with being able to identify transits in data and definitively declaring a flux decrement to be planetary in nature. The former problem is related to the discussion in Chapter 3 and is, in part, related to the fact that time-correlated noise can affect the exoplanet detection threshold, reducing the number of detected planets based on a uncorrelated assumption (see, Chapter 3 and Pont et al. 2006). The latter problem arises because, while radial velocity detection can make some statement about the mass of the secondary object, a transit light curve only can measure fractional obscuration. The signal interpreted as a planetary transit could be as a result of grazing eclipsing binaries, the transit of a brown dwarf across a giant star, the blending of light from a triple star system in which two components are transiting, and more (see, e.g., Alonso et al. 2004b, O'Donovan et al. 2007). Such confusion can be eliminated by a subsequent radial-velocity follow-up of the exoplanet candidate host thereby constraining the planetary mass M_p. Here, since the planet is seen to transit, $i \approx 90^\circ$ and the degeneracy between planetary mass and inclination is broken.

Even in the face of these difficulties, modern transit surveys survive by collecting light curves for a large number stars. Acquiring large numbers of stars is accomplished rather easily for a photometric survey covering a significant portion of the sky (wide) and/or capable of detecting very faint stars (deep).

1.3 Characterizing extrasolar planets that transit

While the detection of exoplanets via transit can be a profitable endeavor, the real power of transit light curve analysis lies in exoplanet characterization. The information encoded in the transit light curve is capable of uniquely determining a large number of exoplanet observables. In this section of the introduction, we quickly review the transit light curve as a tool for precise exoplanet analysis. Prior to diving into the details, it is useful to carefully study the model light curve in Fig. (1-5) and the "gallery" of real, space-based transit light curve data in Figure (1-7).

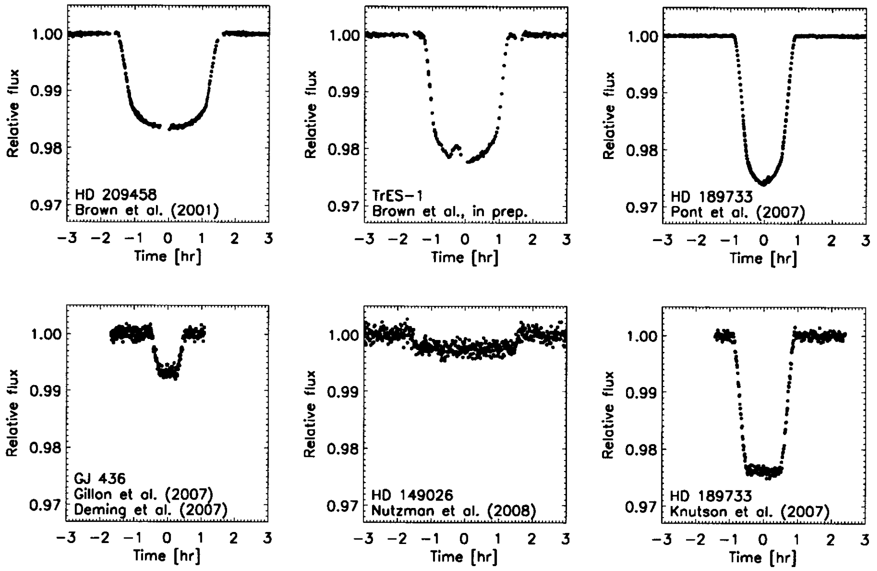

Figure 1-7 High precision exoplanetary transit light curves as measured from space. Figures by Winn (2009).

1.3.1 The exoplanetary transit light curve: From top to bottom

The effect of a transit of an exoplanet across the face of its star is most simply described by the following equation for F, the total flux measured for the combined exoplanet-host system

$$F = F_p(\text{planet}) + F_\star(\text{star}) - \begin{cases} F_\star(\text{planet} \cap \text{star}) & \text{if planet nearer} \\ F_p(\text{planet} \cap \text{star}) & \text{if star nearer} \end{cases} \qquad (1.4)$$

where $F_p(\Omega)$, $F_\star(\Omega)$, are the integrated flux of planet and star over the integration region Ω. We have used the shorthand "planet," "star," or "planet $\cap$ star" to indicate whether the integration region Ω is over the sky-projected planet surface, stellar surface or the intersection of the two. To first order, the sky-projected shape of exoplanet and star are disks with radius R_p and $R_\star$, respectively (see Chapter 5 for an alternative model).

Timescales and observables

The total flux F depends upon time, t. Most importantly, given that the exoplanet is in orbit around the star, planet $\cap$ star is a function of time. In particular, the duration of the transit, T, (when planet $\cap$ star $\neq \emptyset$) scales with the orbital period, P, as

$$T \approx \frac{R_\star}{a}\frac{P}{\pi} \tag{1.5}$$

while the duration of ingress or egress, τ, (for which planet $\cap$ star $\neq$ planet) scales as

$$\tau \approx \frac{R_p}{a}\frac{P}{\pi}. \tag{1.6}$$

If we use Kepler's third law [Eqn. (1.2)], we may write T in a more suggestive form

$$T \approx 13 \left(\frac{P}{1\ \mathrm{yr}}\right)^{1/3} \left(\frac{\rho_\star}{\rho_\odot}\right)^{-1/3} \mathrm{hr}. \tag{1.7}$$

It is therefore feasible to utilize the duration of the transit (or occultation) to make estimates of the mean stellar density, $\rho_\star$ [Perryman et al. 2005, Seager & Mallén-Ornelas 2003]. These precise density estimates may then be used in combination with stellar evolution models to constrain properties of star and planet (see Chapter 4). Again utilizing Kepler's third and also the radial velocity-determined mass M_p in Eqn. (1.1), we may write τ in the more suggestive form

$$\tau \approx 24 \left(\frac{P}{1\ \mathrm{yr}}\right)^{1/2} \left(\frac{g_p}{g_\oplus}\right)^{-1/2} \left(\frac{K}{1\ \mathrm{m/s}}\right)^{1/2} \mathrm{min}. \tag{1.8}$$

It is therefore also feasible to utilize the duration of transit ingress to make estimates of the planetary surface gravity, g_p [Southworth et al. (2007)].

Timing and additional, unseen planets

If we assume the exoplanet follows a Keplerian orbit around its star then the time between two successive transits Δt_c should be equal to the orbital period P. However, if the planet's orbit is perturbed by the gravitational tug of other unseen planets then $\Delta t_c = P + \delta P(t)$. The perturbation to the linear model, $\delta P(t)$, is a function of the mass of the unseen object (Holman & Murray 2005, Agol et al. 2005). It is therefore possible to detect additional planets and their masses from an analysis of a collection of midtransit times [see Fig. (1-8)]. It is important to note that midtransit times are acutely affected by time-correlated noise in the data; special care must therefore be taken to ensure these times are accurate for physical interpretation (see Chapter 3).

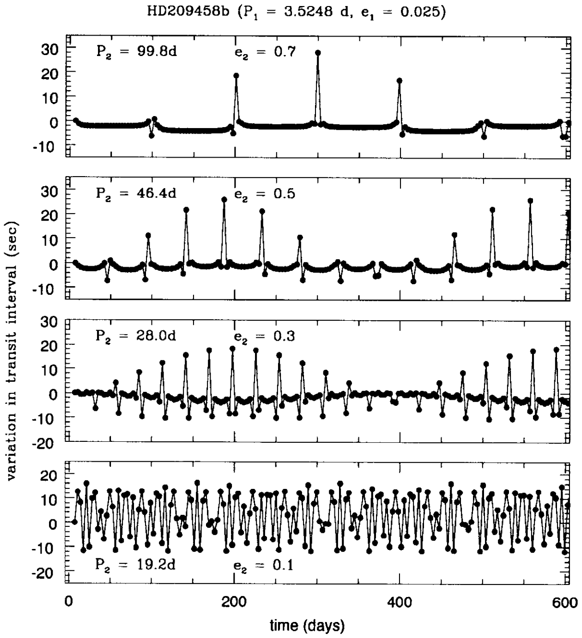

Figure 1-8 Changes in midtransit time as a result of a second planet. This figure, by Holman & Murray (2005), presents variation in the time of midtransit (between successive transits) of HD 209458b in response to gravitation perturbations from a second planet with orbital period P_2 and orbital eccentricity e_2.

Transit or occultation depth: Stellar and exoplanetary atmospheres

The relative transit depth, δ, is given by the normalized form of Eqn. (1.4) at maximum obscuration (planet $\cap$ star = planet),

$$\begin{aligned} \delta &= 1 - \frac{F_p(\text{planet}) + F_\star(\text{star}) - F_\star(\text{planet})}{F_p(\text{planet}) + F_\star(\text{star})} \\ &= \frac{F_\star(\text{planet})}{F_p(\text{planet}) + F_\star(\text{star})}. \end{aligned} \tag{1.9}$$

We have assumed that the stellar and planetary fluxes are constant, however, stellar variability can have a significant effect on the transit light curve [see, e.g., Czela et al. (2009), Silva (2003)]. If, for the moment, we assume the stellar brightness profile $I_\star(r,\theta)$ is constant then

$$\delta \;=\; \left(\frac{R_p}{R_\star}\right)^2 \left(1+\frac{F_p(\text{planet})}{F_\star(\text{star})}\right). \tag{1.10}$$

If we assume further the flux due to the planet, F_p, is negligible compared to that of the star then $\delta \approx (R_p/R_\star)^2$. Thus, the transit depth places a precise constraint on the exoplanetary radius.

Constant stellar brightness profiles are reasonable approximations at mid-infrared and longer wavelengths (see Chapter 5), however, in general, stellar limb-darkening suppresses flux at the stellar disk edges. The effect of the limb-darkening is strongly wavelength dependent and significantly affects the shape of the transit light curve at optical wavelengths [see Figs. (1-7,1-9)]. Limb-darkening tends to round the otherwise boxy transit light curve profile, working to confuse accurate estimation of transit parameters (see Chapter 2). On the other hand, exoplanetary transit light curves provide valuable information about limb-darkening profiles [such as those proposed by Claret (2000)] for stars other than our Sun (Knutson et al. 2007a).

So far, we have regarded the planetary radius, R_p, as independent of how we observe the transit. This is not generally true. In particular, the radial extent of the planet depends on the wavelength of observation, so that $R_p = R_p(\lambda)$. The reason for this dependence is simple: what we as the observer perceive as the radial extent of the exoplanet is determined by the height in the exoplanet atmosphere at which the optical depth for stellar light passing through the atmosphere on its way to us reaches unity, say. This height, z, is dependent on the structure of the atmosphere, the sources of opacity [rotation-vibrational molecular absorption, for example, see Fig. (1-10)], and the wavelength λ of our observation (Seager & Sasselov 2000, Brown 2001, Hubbard et al. 2001, Hui & Seager 2002 and Chapter 4). By observing transits and determining transit depths at multiple wavelengths, we may form an absorption spectrum of the planetary atmosphere. This technique, as executed by Swain et al. (2008) and illustrated in Fig. (1-10), is often referred to as transmission spectroscopy.

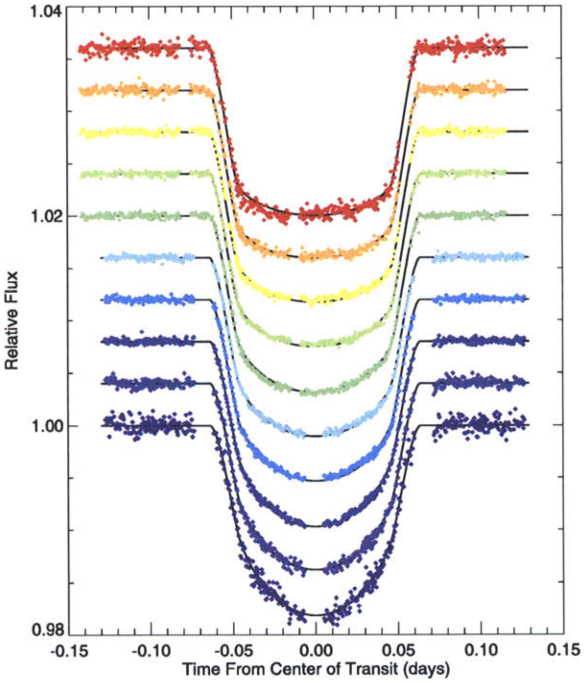

Figure 1-9 The effect of stellar limb-darkening on the transit light curve of HD 209458b. In this figure by Knutson et al. (2007), transit light curve data is shown for HD 209458b in wavelength bands spanning from 293 to 1019 nm. The curvature in each light curve is as a result of wavelength dependent stellar limb-darkening.

While we, in this Thesis, are concerned mainly with the transit portion of the total light curve [Eqn. (1.4)], an observation at occultation is extremely useful in constraining the atmosphere of the exoplanet. The occultation depth, δ_o, may be derived in an analogous fashion to Eqn. (1.9) as

$$\begin{aligned}\delta_o &= 1 - \frac{F_p(\text{planet}) + F_\star(\text{star}) - F_p(\text{planet})}{F_p(\text{planet}) + F_\star(\text{star})} \\ &= 1 - \frac{F_\star(\text{star})}{F_p(\text{planet}) + F_\star(\text{star})} \\ &\approx \frac{F_p(\text{planet})}{F_\star(\text{star})}. \qquad (1.11)\end{aligned}$$

If we again assume that the brightness profile of both planet and star are constant

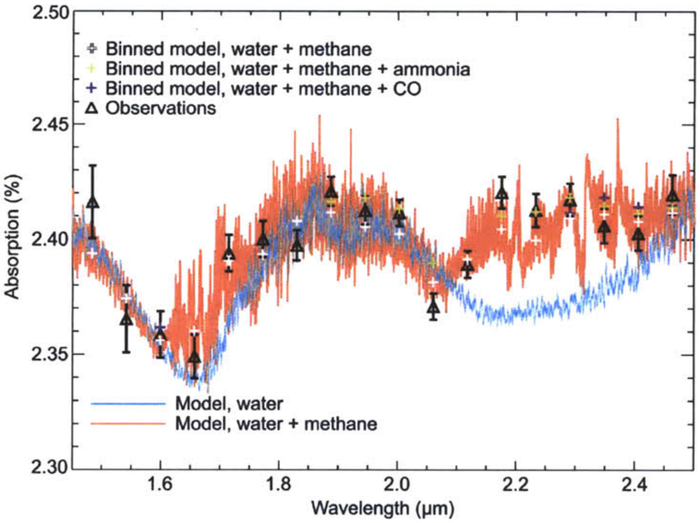

Figure 1-10 Transmission spectroscopy of HD 189733b. This figure by Swain et al. (2008) plots the wavelength-dependent transit depth for the exoplanet HD 189733b. The shape of the spectrum suggests the presence of methane in the exoplanet atmosphere.

then

$$\delta_o \approx \left(\frac{R_p}{R_\star}\right)^2 \frac{I_p}{I_\star} \tag{1.12}$$

where the ratio of the planetary and stellar intensities $I_p/I_\star$ is, to first order, related to the black-body temperature of planet and star. For an observation in the infrared, $I_p/I_\star \approx T_p/T_\star$. A measurement of occultation depth can, therefore, constrain the temperature of the photosphere of the exoplanet [see, for example, Harrington et al. (2007)].

The thermal emission from the exoplanetary photosphere may be non-uniform across its surface. Thus, as the planet rotates to show different faces while it moves through its orbit, we will measure a time-varying total flux, $F_\star(\text{star}) + F_p(\text{planet})$. By measuring this light curve (the so-called "phase function") we may learn how heat is transported through and redistributed throughout the exoplanet atmosphere. If the rotational period is known precisely, this phase function may be inverted into a temperature map of the planetary photosphere [as was done for HD 189733b, see Knutson et al. (2007b)].

At visible wavelengths, stellar light reflected off the planetary surface (by clouds, for example) is the dominant contribution to the occultation depth, whereby

$$\delta_o \approx \alpha \left(\frac{R_p}{a}\right)^2 \tag{1.13}$$

where α is the geometric albedo of the exoplanet (e.g., Seager 2008). By first measuring the quantity R_p/a during transit, a measurement of occultation depth in optical wavelengths would yield the albedo α.

Moons, Rings and Oblateness

While we have so far assumed the exoplanet is circular in projection, perturbations to the obscuring shape are possible, if not likely. If, for example, the exoplanet has close gravitationally bound companions, such as moons or rings, it is likely they will induce a non-trivial effect on the transit light curve. Even with no companions present, the exoplanet itself is likely to be non-spherical, as is the case with Solar System planets (Murray & Dermott 2000).

Moons present the most obvious perturbation, contributing to the total flux deficit as an additional transit on top of the transit of the exoplanet. Time varying effects in the photometry may help to constrain the exomoon mass and orbital period. Currently no transit data support the presence of an exomoon around any of the transiting planets [see, for example, Pont et al. (2007)]. Exomoons may also be detected by identifying the signature of their gravitational effect on their planetary host from anomalies in a collection of midtransit times or in the time variability of transit durations (Kipping 2000).

Rings present a subtler effect on the transit light curve, depending on the orientation of the rings in the sky plane and the level of extinction due to ring particles [Barnes & Fortney 2004, Ohta et al. 2009, see Figure (1-11)].

The shape of the exoplanet, most notably oblateness owing to rigid-body rotation of the bulk (see Chapter 5), is in principle measurable from the transit light curve. The effect is most evident during the phases of ingress and egress. With a measurement of oblateness, it is possible to constrain the planet's rotation period, its internal constitution and possible evolutionary histories (see Chapter 5).

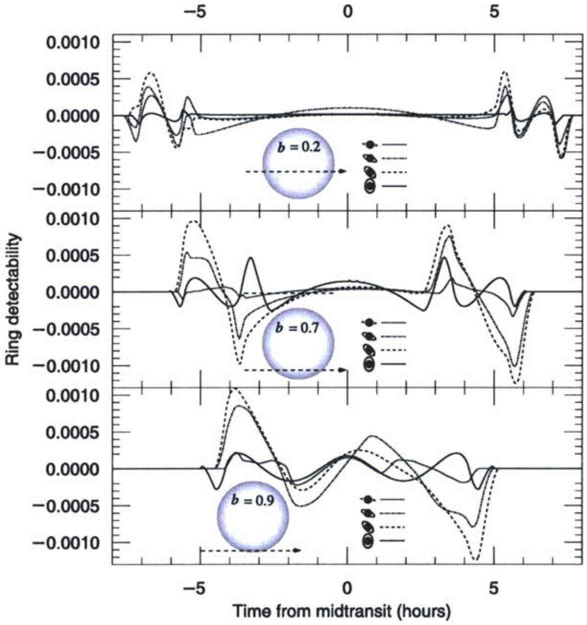

Figure 1-11 Signatures of exoplanetary rings in the transit light curve. Figure by Barnes & Fortney (2004).

Transits in radial velocity

The transit, when observed with Doppler spectroscopy (as used for detection in § 1.2.1), appears as an "anomalous" perturbation to the radial velocity of the stellar host [see Fig. (1-12)]. This so-called Rossiter-McLaughlin effect is as a result of the obscuring exoplanet covering a portion of the receding (approaching) half of the rotating stellar disk inducing an excess of "red" ("blue") Doppler-shifted photons. By measuring the radial velocity at transit, it is therefore possible to measure the sky projection of the angle between the spin axis of the star and that of the exoplanetary orbit (Gaudi & Winn 2007). This angle can be used to constrain possible dynamical scenarios involving additional planets in the stellar system [see, for example, Fabrycky & Tremaine (2007)].

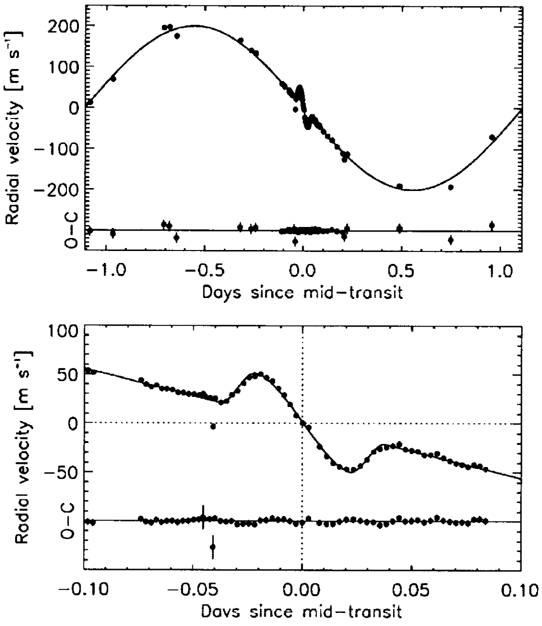

Figure 1-12 "Anomalous" velocity in radial velocity measured during transit. This figure showing the Rossiter-McLaughlin effect for HD 189733b is by Winn et al. (2006). See § 1.3.1 for details.

1.4 Thesis overview

As we have attempted to make clear, the light curve of an exoplanetary transit can be used to estimate the planetary radius and other parameters of interest. Because accurate parameter estimation is a non-analytic and computationally intensive problem, it is often useful to have analytic approximations for the parameters as well as their uncertainties and covariances. In Chapter 2, we give such formulas, for the case of an exoplanet transiting a star with a uniform brightness distribution. We also assess the advantages of some relatively uncorrelated parameter sets for fitting actual data. When limb darkening is significant, our parameter sets are still useful, although our analytic formulas underpredict the covariances and uncertainties.

We consider, in Chapter 3, the problem of fitting a parametric model to time-series data that are afflicted by correlated noise. The noise is represented by a sum of two stationary Gaussian processes: one that is uncorrelated in time, and another that has a power spectral density varying as $1/f^{\gamma}$. We present an accurate and fast $[O(N)]$ algorithm for parameter estimation based on computing the likelihood in a wavelet basis. The method is illustrated and tested using simulated time-series photometry of exoplanetary transits, with particular attention to estimating the midtransit time (see § 1.3.1). We compare our method to two other methods that have been used in the literature, the time-averaging method and the residual-permutation method. The algorithm presented in this chapter generally gives more accurate results for midtransit times and truer estimates of their uncertainties.

The transiting exoplanet HD 149026b is an important case for theories of planet formation and planetary structure, because the planet's relatively small size has been interpreted as evidence for a highly metal-enriched composition. We present, in Chapter 4, observations of 4 transits with the Near Infrared Camera and Multi-Object Spectrometer on the *Hubble Space Telescope* within a wavelength range of 1.1–2.0 μm. Analysis of the light curve gives the most precise estimate yet of the stellar mean density (see § 1.3.1), $\rho_\star = 0.497^{+0.042}_{-0.057}$ g cm^{-3}. By requiring agreement between the observed stellar properties (including $\rho_\star$) and stellar evolutionary models, we refine the estimate of the stellar radius: $R_\star = 1.541^{+0.046}_{-0.042}\ R_\odot$. We also find a deeper transit than has been measured at optical and mid-infrared wavelengths. Taken together, these findings imply a planetary radius of $R_p = 0.813^{+0.027}_{-0.025}\ R_{\rm Jup}$, which is larger than earlier estimates. Models of the planetary interior still require a metal-enriched composition, although the required degree of metal enrichment is reduced. It is also possible that the deeper NICMOS transit is caused by wavelength-dependent absorption by constituents in the planet's atmosphere (see § 1.3.1), although simple model atmospheres do not predict this effect to be strong enough to account for the discrepancy. We use the 4 newly-measured transit times to compute a refined transit ephemeris.

Finally, in Chapter 5, we place empirical constraints on the oblateness (see § 1.3.1) of the "Hot Jupiter" HD 189733b by completing a careful analysis of 7 transits observed with the InfraRed Array Camera (IRAC) onboard the *Spitzer Space Telescope.* We rule out, at 95% confidence, oblateness similar to that of Saturn at all or, for that of Jupiter, at most obliquities. By assuming the oblateness to be as a result of rigid-body rotation, we place constraints on the rotational period of the planet. In particular, we find that HD 189733b is rotating slower than once every 21 hours at 95% confidence. We also consider the detection of oblateness for the highly eccentric transiting exoplanet HD 80606b. The algorithm developed to quickly calculate the transit light curve of an oblate exoplanet is described in depth.

Bibliography

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567

Alonso, R., et al. 2004a, ApJ, 613, L153

Alonso, R., Deeg, H. J., Brown, T. M., & Belmonte, J. A. 2004b, Stellar Structure and Habitable Planet Finding, 538, 255

Baglin, A. 2003, Advances in Space Research, 31, 345

Bakos, G. Á., et al. 2007, ApJ, 656, 552

Barnes, J. W., & Fortney, J. J. 2004, ApJ, 616, 1193

Borucki, W., et al. 2009, IAU Symposium, 253, 289

Brown, T. M. 2001, ApJ, 553, 1006

Butler, R. P., et al. 2006, ApJ, 646, 505

Carroll, B. W., & Ostlie, D. A. 2006, Institute for Mathematics and Its Applications,

Charbonneau, D., Brown, T. M., Burrows, A., & Laughlin, G. 2007, Protostars and Planets V, 701

Claret, A. 2000, A&A, 363, 1081

Czesla, S., Huber, K. F., Wolter, U., Schröter, S., & Schmitt, J. H. M. M. 2009, arXiv:0906.3604

Fabrycky, D., & Tremaine, S. 2007, ApJ, 669, 1298

Gaudi, B. S., & Winn, J. N. 2007, ApJ, 655, 550

Harrington, J., Luszcz, S., Seager, S., Deming, D.,

Holman, M. J., & Murray, N. W. 2005, Science, 307, 1288

Horne, K. 2003, Scientific Frontiers in Research on Extrasolar Planets, 294, 361

Hubbard, W. B., Fortney, J. J., Lunine, J. I., Burrows, A., Sudarsky, D.,

Hui, L., & Seager, S. 2002, ApJ, 572, 540

Kipping, D. M. 2009, MNRAS, 392, 181

Knutson, H. A., Charbonneau, D., Noyes, R. W., Brown, T. M., & Gilliland, R. L. 2007a, ApJ, 655, 564

Knutson, H. A., et al. 2007b, Nature, 447, 183

Kopal, Z. 1990, Dordrecht, Netherlands, Kluwer Academic Publishers, 1990, 163 p.,

Mayor, M., et al. 2009, A&A, 493, 639

Mazeh, T., et al. 2000, ApJ, 532, L55

McCullough, P. R., Stys, J. E., Valenti, J. A., Fleming, S. W., Janes, K. A., & Heasley, J. N. 2005, PASP, 117, 783

Murray, C. D., & Dermott, S. F. 2000, Solar System Dynamics, by C.D. Murray and S.F. Dermott. Cambridge, UK: Cambridge University Press, 2000.,

Naef, D., et al. 2001, A&A, 375, L27

O'Donovan, F. T., & Charbonneau, D. 2007, Transiting Extrapolar Planets Workshop, 366, 58

Ohta, Y., Taruya, A., & Suto, Y. 2009, ApJ, 690, 1

Pepe, F., et al. 2002, The Messenger, 110, 9

Perryman, M., et al. 2005, arXiv:astro-ph/0506163

Pollacco, D., et al. 2006, Ap&SS, 304, 253

Pont, F., et al. 2007, A&A, 476, 1347

Pont, F., Zucker, S., & Queloz, D. 2006, MNRAS, 373, 231

Seager, S. 2008, Space Science Reviews, 135, 345

Seager, S., & Mallén-Ornelas, G. 2003, ApJ, 585, 1038

Seager, S., & Sasselov, D. D. 2000, ApJ, 537, 916

Silva, A. V. R. 2003, ApJ, 585, L147

Southworth, J., Wheatley, P. J., & Sams, G. 2007, MNRAS, 379, L11

Struve, O. 1952, The Observatory, 72, 199

Swain, M. R., Vasisht, G., & Tinetti, G. 2008, Nature, 452, 329

Torres, G., Winn, J. N., & Holman, M. J. 2008, ApJ, 677, 1324

Udalski, A. 2007, Transiting Extrapolar Planets Workshop, 366, 51

Udry, S., & Santos, N. C. 2007, ARA&A, 45, 397

Winn, J. N., et al. 2006, ApJ, 653, L69

Winn, J. N. 2009, IAU Symposium, 253, 99

Chapter 2

Analytic approximations for transit light-curve observables, uncertainties, and covariances

2.1 Introduction

In general, the parameters of a transiting system and their uncertainties must be estimated from the photometric data using numerical methods. For example, many investigators have used χ^2-minimization schemes such as AMOEBA or the Levenberg-Marquardt method, along with confidence levels determined by examining the appropriate surface of constant $\Delta\chi^2$ (see, e.g., Brown et al. 2001, Alonso et al. 2004) or by bootstrap methods (e.g., Sato et al. 2005, Winn et al. 2005). More recently it has become common to use Markov Chain Monte Carlo methods (e.g., Holman et al. 2006, Winn et al. 2007, Burke et al. 2007). However, even when numerical algorithms are required for precise answers, it is often useful to have analytic approximations for the parameters as well as their uncertainties and covariances.

Analytic approximations can be useful for planning observations. For example, one may obtain quick answers to questions such as, for which systems can I expect to obtain the most precise measurement of the orbital inclination? Or, how many transit light curves will I need to gather with a particular telescope before the statistical error in the planetary radius is smaller than the systematic error? Now that nearly 50 transiting planets are known, we enjoy a situation in which a given night frequently offers more than one observable transit event. Analytic calculations can help one decide which target is more fruitfully observed, and are much simpler and quicker than the alternative of full numerical simulations. Analytic approximations are also useful for understanding the parameter degeneracies inherent in the model, and for constructing relatively uncorrelated parameter sets that will speed the convergence of optimization algorithms. Finally, analytic approximations are useful in order-of-magnitude estimates of the observability of subtle transit effects, such as transit timing variations, precession-induced changes in the transit duration, or the asymmetry in the ingress and egress durations due to a nonzero orbital eccentricity.

Mandel & Agol (2005) and Giménez (2007) have previously given analytic formulas for the received flux as a function of the relative separation of the planet and the star, but their aim was to provide highly accurate formulas, which are too complex for useful analytic estimates of uncertainties and covariances. Protopapas et al. (2007) provided an analytic and differentiable approximation to the transit light curve, but they were concerned with speeding up the process of searching for transits in large databases, rather than parameter estimation. Seager & Mallén-Ornelas (2003) presented an approximate model of a transit light curve with the desired level of simplicity, but did not provide analytic estimates of uncertainties and covariances.

This chapter is organized as follows. In § 2.2 we present a simple analytic model for a transit light curve, using a convenient and intuitive parameterization similar to that of Seager & Mallen-Ornelas (2003). In § 2.3, we derive analytic approximations for the uncertainties and covariances of the basic parameters, and in § 2.4 we verify the accuracy of those approximations through numerical tests. Our model assumes that the flux measurements are made continuously throughout the transit, and that stellar limb-darkening is negligible; in § 2.4.1 and § 2.4.3 we check on the effects of relaxing these assumptions. In § 2.5 we derive some useful expressions for the uncertainties in some especially interesting or useful "derived" parameters, i.e., functions of the basic model parameters. In § 2.6 we present alternative parameter sets that are better suited to numerical algorithms for parameter estimation utilizing the analytic formalism given in § 2.3. We compare the correlations among parameters for various parameter sets that have been used in the transit literature. Finally, § 7 gives a summary of the key results.

2.2 Linear approximation to the transit light curve

Imagine a spherical star of radius $R_\star$ with a uniform brightness and an unocculted flux f_0. When a dark, opaque, spherical planet of radius R_p is in front of the star, at a center-to-center sky-projected distance of $zR_\star$, the received stellar flux is $F^e(r, z, f_0) = f_0(1 - \lambda^e(r, z))$, where

$$\lambda^e(r,z) = \begin{cases} 0 & 1+r < z \\ \frac{1}{\pi}\left(r^2\kappa_0 + \kappa_1 - \sqrt{\frac{4z^2-(1+z^2-r^2)^2}{4}}\right) & 1-r < z \leq 1+r \\ r^2 & z \leq 1-r \end{cases} , \tag{2.1}$$

with $\kappa_1 = \cos^{-1}[(1 - r^2 + z^2)/2z]$ and $\kappa_0 = \cos^{-1}[(r^2 + z^2 - 1)/2rz]$ (Mandel & Agol 2002). Geometrically, λ^e is the overlap area between two circles with radii 1 and r whose centers are z units apart. The approximation of uniform brightness (no limb darkening) is valid for mid-infrared bandpasses, which are increasingly being used for transit observations (see, e.g., Harrington et al. 2007, Knutson et al. 2007, Deming et al. 2007), and is a good approximation even for near-infrared and far-red bandpasses. We make this approximation throughout this chapter, except in § 2.4.3 where we consider the effect of limb darkening.

For a planet on a circular orbit, the relation between z and the time t is

$$z(t) = aR_\star^{-1}\sqrt{[\sin\ n(t-t_c)]^2 + [\cos\ i \cos\ n(t-t_c)]^2} \tag{2.2}$$

where a is the semimajor axis, i is the inclination angle, $n \equiv 2\pi/P$ is the angular frequency with period P, and t_c is the transit midpoint (when z is smallest).

The four "contact times" of the transit are the moments when the planetary disk and stellar disk are tangent. First contact ($t_{\rm I}$) occurs at the beginning of the transit, when the disks are externally tangent. Second contact ($t_{\rm II}$) occurs next, when the disks are internally tangent. Third and fourth contacts ($t_{\rm III}$ and $t_{\rm IV}$) are the moments of internal and external tangency, respectively, as the planetary disk leaves the stellar disk. The total transit duration is $t_{\rm IV} - t_{\rm I}$. The ingress phase is defined as the interval between $t_{\rm I}$ and $t_{\rm II}$, and likewise the egress phase is defined as the interval between $t_{\rm III}$ and $t_{\rm IV}$. We also find it useful to define the ingress midpoint $t_{\rm ing} \equiv (t_{\rm I} + t_{\rm II})/2$ and the egress midpoint $t_{\rm egr} \equiv (t_{\rm III} + t_{\rm IV})/2$.

Although Eqns. (2.1) and (2.2) give an exact solution, they are too complicated for an analytic error analysis. We make a few approximations to enable such an analysis. First, we assume the orbital period is large compared to transit duration, in which case Eqn. (2.2) is well-approximated by

$$z(t) \quad = \quad \sqrt{[(t-t_c)/\tau_0]^2 + b^2}, \tag{2.3}$$

where, for a circular orbit, $\tau_0 = R_\star P/2\pi a = R_\star/na$ and $b = a\cos i/R_\star$ is the normalized impact parameter. In this limit, the planet moves uniformly in a straight line across the stellar disk. Simple expressions may be derived for two characteristic timescales of the transit:

$$t_{\rm egr} - t_{\rm ing} \quad = \quad \tau_0\left(\sqrt{(1+r)^2 - b^2} + \sqrt{(1-r)^2 - b^2}\right) = 2\tau_0\sqrt{1-b^2} + O(r^2) \tag{2.4}$$

$$t_{\rm II} - t_{\rm I} \quad = \quad \tau_0\left(\sqrt{(1+r)^2 - b^2} - \sqrt{(1-r)^2 - b^2}\right) = 2\tau_0\frac{r}{\sqrt{1-b^2}} + O(r^3) \tag{2.5}$$

It is easy to enlarge the discussion to include eccentric orbits, by replacing a by the planet-star distance at midtransit, and n by the angular frequency at midtransit:

$$a \quad \rightarrow \quad \frac{a(1-e^2)}{1+e\sin\omega},$$

$$n \longrightarrow \frac{n(1+e\sin\omega)^2}{(1-e^2)^{\frac{3}{2}}},$$

where e is the eccentricity, and ω is the argument of pericenter. Here, too, we approximate the planet's actual motion by uniform motion across the stellar disk, with a velocity equal to the actual velocity at midtransit. Methods for computing these quantities at midtransit are discussed by Murray & Dermott (2000), as well as recent transit-specific studies by Barnes (2007), Burke (2008), Ford et al. (2008), and Gillon et al. (2007). We redefine the parameters τ_0 and b in this expanded scope as

$$b \equiv \frac{a\cos i}{R_\star}\left(\frac{1-e^2}{1+e\sin\omega}\right) \tag{2.6}$$

$$\tau_0 \equiv \frac{R_\star}{an}\left(\frac{\sqrt{1-e^2}}{1+e\sin\omega}\right). \tag{2.7}$$

We do not restrict our discussion to circular orbits ($e = 0$) unless otherwise stated.

Next, we replace the actual light curve with a model that is piecewise-linear in time, as illustrated in Figure 2-1. Specifically, we define the parameters

$$\delta \equiv f_0 r^2 = f_0(R_p/R_\star)^2 \tag{2.8}$$

$$T \equiv 2\tau_0\sqrt{1-b^2} \tag{2.9}$$

$$\tau \equiv 2\tau_0\frac{r}{\sqrt{1-b^2}} \tag{2.10}$$

and then we define our model light curve as

$$F^l(t) = \begin{cases} f_0 - \delta & |t-t_c| \le T/2 - \tau/2 \\ f_0 - \delta + \frac{\delta}{\tau}\left(|t-t_c| - T/2 + \tau/2\right) & T/2 - \tau/2 < |t-t_c| < T/2 + \tau/2 \\ f_0 & |t-t_c| \ge T/2 + \tau/2 \end{cases} \tag{2.11}$$

We use the symbol F^l to distinguish this piecewise-linear model (l for linear) from the exact uniform-source expression F^e given by Eqns. (2.1) and (2.2). The deviations between F^l and F^e occur near and during the ingress and egress phases. The approximation is most accurate in the limit of small r and b and is least accurate for grazing transits. As shown in Eqn. (2.5), when r is small, $\tau \approx t_{\rm II} - t_{\rm I}$ (the ingress or egress duration) and $T \approx t_{\rm egr} - t_{\rm ing}$ (the total transit duration). Neither this piecewise-linear model nor the choice of parameters is new. Seager & Mallén-Ornelas (2003) also used a piecewise-linear model, with different linear combinations of these parameters, and both Burke et al. (2007) and Bakos et al. (2007) have employed parameterizations that are closely related to the parameters given above. What is specifically new to this chapter is an analytic error and covariance analysis of this linear model, along with useful analytic expressions for errors in the physical parameters of the system. The "inverse" mapping from our parameterization to a more physical parameterization is

$$\begin{aligned}
r^2 = (R_p/R_\star)^2 &= \delta/f_0 && (2.12)\\
b^2 = \left(\frac{a\cos i}{R_\star}\right)^2\left(\frac{1-e^2}{1+e\sin\omega}\right)^2 &= 1 - r\frac{T}{\tau} && (2.13)\\
\tau_0^2 = \left(\frac{R_\star}{an}\right)^2\left(\frac{\sqrt{1-e^2}}{1+e\sin\omega}\right)^2 &= \frac{T\tau}{4r}. && (2.14)
\end{aligned}$$

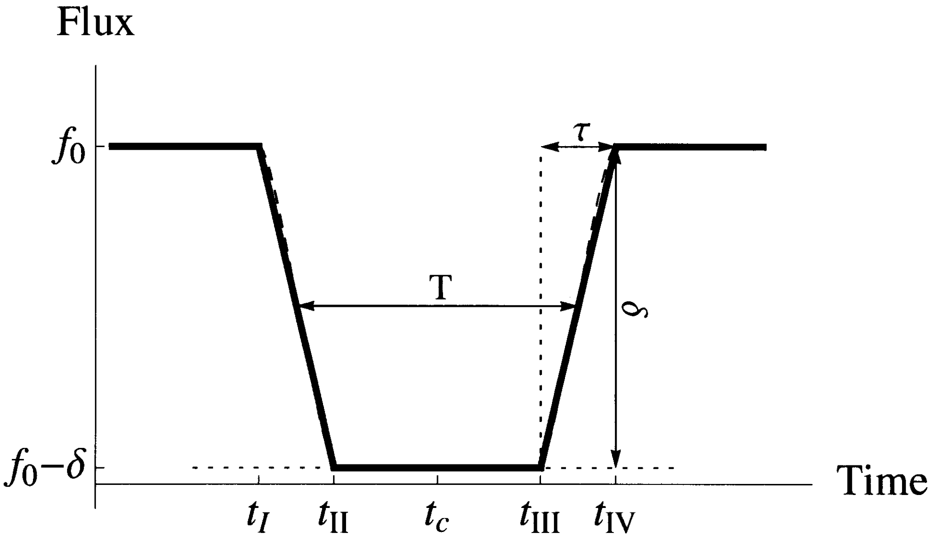

Figure 2-1 Comparison of the exact and piecewise-linear transit models, for the parameter choice $r = 0.2$, $b = 0.5$. The dashed line shows the exact uniform-source model F^e, given by Eqn. (2.1). The solid line shows the linear model F^l, given by Eqn. (2.11).

2.3 Fisher information analysis

Given a model $F(t; \{p_i\})$ with independent variable t and a set of parameters $\{p_i\}$, it is possible to estimate the covariance between parameters, $\mathrm{Cov}(p_i, p_j)$, that would be obtained by measuring $F(t)$ with some specified cadence and precision. (Gould 2003 gives a pedagogical introduction to this technique.) Suppose we have N data points taken at times t_k spanning the entire transit event. The error in each data point is assumed to be a Gaussian random variable, with zero mean and standard deviation σ_k. Then the covariance between parameters $\{p_i\}$ is

$$\mathrm{Cov}(p_i, p_j) = \left(B^{-1}\right)_{ij} \tag{2.15}$$

where B is the zero-mean Gaussian-noise Fisher information matrix, which is calculated as

$$B_{ij} = \sum_{k=1}^{N} \sum_{l=1}^{N} \left[\frac{\partial}{\partial p_i} F(t_k; \{p_m\})\right] \mathcal{B}_{kl} \left[\frac{\partial}{\partial p_j} F(t_l; \{p_m\})\right]. \tag{2.16}$$

Here, $\mathcal{B}_{kl}$ is the inverse covariance matrix of the flux measurements. We assume the measurement errors are uncorrelated (i.e., we neglect "red noise"), in which case $\mathcal{B}_{kl} = \delta_{kl}\sigma_k^{-2}$. We further assume that the measurement errors are uniform in time with $\sigma_k = \sigma$, giving $\mathcal{B}_{kl} = \delta_{kl}\sigma^{-2}$.

In Table (2.1), we compute the needed partial derivatives[1] of the piecewise-linear light curve F^l, which has five parameters $\{p_i\} = \{t_c, \tau, T, \delta, f_0\}$.

Fig. (2-2) shows the time dependence of the parameter derivatives, for a particular case. The time dependence of the parameter derivatives for the exact uniform-source model F^e is also shown, for comparison, as are the numerical derivatives for limb-darkenened light curves. This comparison shows that the linear model captures the essential features of more realistic models, and in particular the symmetries. The most

---

[1] In computing these derivatives we have ignored the dependence of the piecewise boundaries in Table. (2.1) on the parameter values. The derivatives associated with those boundary changes are finite, and have a domain of measure zero in the limit of continuous sampling. Thus they do not affect our covariance calculation.

	Totality	Ingress/Egress	Out of Transit
$\frac{\partial}{\partial t_c} F^l(t;\{p_m\})$	0	$-\frac{\delta}{\tau}\frac{t-t_c}{\lvert t-t_c\rvert}$	0
$\frac{\partial}{\partial \tau} F^l(t;\{p_m\})$	0	$-\frac{\delta}{\tau^2}\left(\lvert t-t_c\rvert-\frac{T}{2}\right)$	0
$\frac{\partial}{\partial T} F^l(t;\{p_m\})$	0	$-\frac{\delta}{2\tau}$	0
$\frac{\partial}{\partial \delta} F^l(t;\{p_m\})$	-1	$\frac{1}{\tau}\left(\lvert t-t_c\rvert-\frac{T}{2}\right)-\frac{1}{2}$	0
$\frac{\partial}{\partial f_0} F^l(t;\{p_m\})$	1	1	1

Table 2.1 Table of partial derivatives of the piecewise-linear light curve F^l, in the five parameters $\{p_i\} = \{t_c, \tau, T, \delta, f_0\}$. The intervals $|t - t_c| < T/2 - \tau/2$, $T/2 - \tau/2 < |t - t_c| < T/2 + \tau/2$, and $|t - t_c| > T/2 + \tau/2$ correspond to totality, ingress/egress, and out of transit respectively.

obvious problem with the linear model is that it gives a poor description of the τ-derivative and the δ-derivative for the case of appreciable limb darkening, as discussed further in § 2.4.3. From Fig. (2-2) and Table (2.1) we see that for the parameters T, τ, and δ, the derivatives are symmetric about $t = t_c$, while the derivative for the parameter t_c is antisymmetric about t_c. This implies that t_c is uncorrelated with the other parameters. (This is also the case for the exact model, with or without limb darkening.)

We suppose that the data points are sampled uniformly in time at a rate $\Gamma = N/T_{\rm tot}$, where the observations range from $t = t_0$ to $t = t_0 + T_{\rm tot}$ and encompass the entire transit event. In the limit of large $\Gamma\tau$ we may approximate the sums of Eqn. (2.16) with time integrals,

$$B_{ij} \;=\; \frac{\Gamma}{\sigma^2}\int_{t_0}^{t_0+T_{\rm tot}} \left[\frac{\partial}{\partial p_i}F^l(t;\{p_m\})\right]\left[\frac{\partial}{\partial p_j}F^l(t;\{p_m\})\right] dt. \qquad (2.17)$$

Using the derivatives from Table (2.1) we find

$$B \;=\; \frac{\Gamma}{\sigma^2}\begin{pmatrix} \frac{2\delta^2}{\tau} & 0 & 0 & 0 & 0 \\ 0 & \frac{\delta^2}{6\tau} & 0 & -\frac{\delta}{6} & 0 \\ 0 & 0 & \frac{\delta^2}{2\tau} & \frac{\delta}{2} & -\delta \\ 0 & -\frac{\delta}{6} & \frac{\delta}{2} & T-\frac{\tau}{3} & -T \\ 0 & 0 & -\delta & -T & T_{\rm tot} \end{pmatrix}. \qquad (2.18)$$

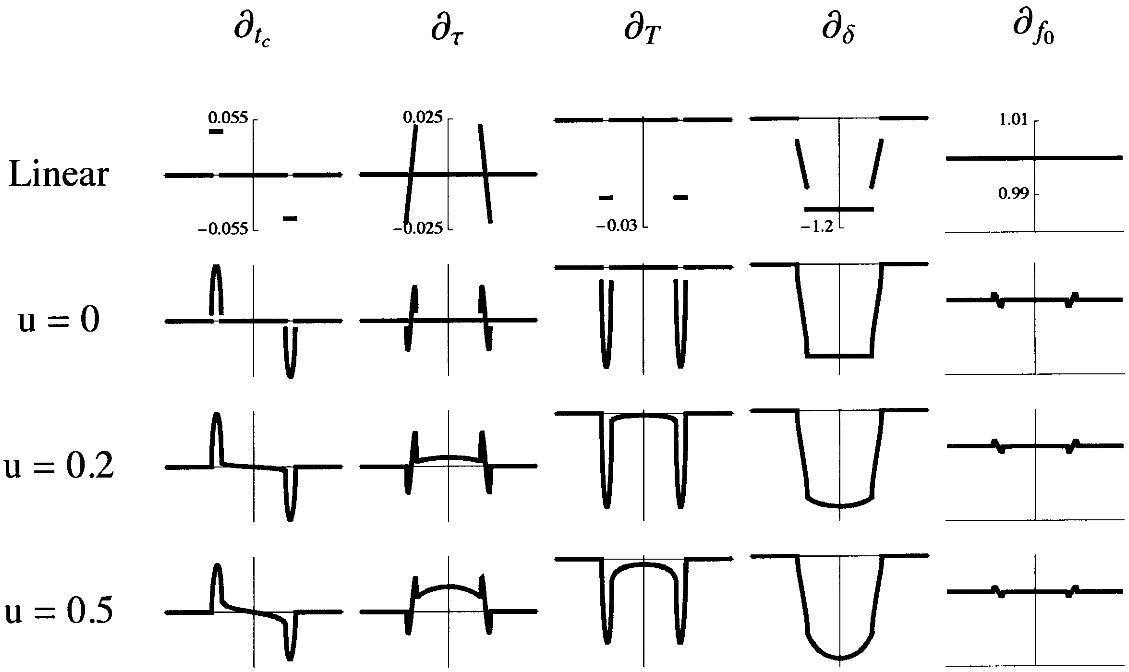

Figure 2-2 Parameter derivatives, as a function of time, for the piecewise-linear model light curve F^l (top row), the exact light curve for the case of zero limb darkening F^e (second row), and for numerical limb-darkened light curves with a linear limb-darkening coefficient $u = 0.2$ (third row) and $u = 0.5$ (bottom row). See § 2.4.3 for the definition of u. Typical scales are shown in the first row and are consistent in the following rows.

In what follows, it is useful to define some dimensionless variables:

$$\begin{aligned} Q &\equiv \sqrt{\Gamma T}\,\frac{\delta}{\sigma}, \\ \theta &\equiv \tau/T, \\ \eta &\equiv T/(T_{\rm tot} - T - \tau). \end{aligned} \tag{2.19}$$

The first of these variables, Q, is equal to the total signal-to-noise ratio of the transit in the limit $r \longrightarrow 0$. The second variable, θ, is approximately the ratio of ingress (or egress) duration to the total transit duration. The third variable, η, is approximately the ratio of the number of data points obtained during the transit to the number of data points obtained before or after the transit. Oftentimes, r and θ are much smaller than unity, which will later enable us to derive simple expressions for the variances and covariances, but for the moment we consider the general case.

Inverting B, we find the covariance matrix for the piecewise-linear model,

$$\mathrm{Cov}(\{t_c, \tau, T, \delta, f_0\}\ , \{t_c, \tau, T, \delta, f_0\}) =$$
$$\frac{1}{Q^2}\begin{pmatrix} \frac{\theta}{2}T^2 & 0 & 0 & 0 & 0 \\ 0 & \left[\eta\theta + \frac{6-5\theta}{1-\theta}\right]\theta T^2 & \left[\eta - \frac{1}{1-\theta}\right]\theta^2 T^2 & \left[\eta + \frac{1}{1-\theta}\right]\theta\delta T & \eta\theta\delta T \\ 0 & \left[\eta - \frac{1}{1-\theta}\right]\theta^2 T^2 & \left[\eta\theta + \frac{2-\theta}{1-\theta}\right]\theta T^2 & \left[\eta - \frac{1}{1-\theta}\right]\theta\delta T & \eta\theta\delta T \\ 0 & \left[\eta + \frac{1}{1-\theta}\right]\theta\delta T & \left[\eta - \frac{1}{1-\theta}\right]\theta\delta T & \left[\eta + \frac{1}{1-\theta}\right]\delta^2 & \eta\delta^2 \\ 0 & \eta\theta\delta T & \eta\theta\delta T & \eta\delta^2 & \eta\delta^2 \end{pmatrix}. \tag{2.20}$$

The elements along the diagonal of the covariance matrix are variances, or squares of standard errors, $\sigma_{p_i} = \sqrt{\mathrm{Cov}(p_i, p_i)}$.

This result can be simplified for the case when many out-of-transit observations are obtained and $\eta \longrightarrow 0$. In this limit, f_0 is known with negligible error, and we may assume $f_0 = 1$ without loss of generality. In this case, δ is the fractional transit depth, and the covariance matrix becomes

$$\mathrm{Cov}(\{t_c,\tau,T,\delta\}\quad,\quad\{t_c,\tau,T,\delta\}) = \frac{1}{Q^2}$$

$$\begin{pmatrix} \frac{\theta}{2}T^2 & 0 & 0 & 0 \\ 0 & \frac{\theta(6-5\theta)}{1-\theta}T^2 & -\frac{\theta^2}{1-\theta}T^2 & \frac{\theta}{1-\theta}\delta T \\ 0 & -\frac{\theta^2}{1-\theta}T^2 & \frac{\theta(2-\theta)}{1-\theta}T^2 & -\frac{\theta}{1-\theta}\delta T \\ 0 & \frac{\theta}{1-\theta}\delta T & -\frac{\theta}{1-\theta}\delta T & \frac{1}{1-\theta}\delta^2 \end{pmatrix}. \tag{2.21}$$

from which it is obvious that θ is the key controlling parameter that deserves special attention. Using Eqns. (2.9) and (2.10) we may write

$$\theta = \frac{r}{1-b^2}. \tag{2.22}$$

Unless the transit is grazing, we have $b \le 1-r$, and θ is restricted to the range $[r, \frac{1}{2-r}]$. Fig. (2-3) shows the dependence of θ on the impact parameter, for various choices of the transit depth. It is important to keep in mind that for $b \lesssim 0.5$, θ is nearly equal to r and depends weakly on b. This implies that θ is expected to be quite small for most transiting systems. For planetary orbits that are randomly oriented in space, the expected distribution of b is uniform, and hence we expect $\theta \lesssim 0.3$ for 90% of a random sample of transiting planets with $R_p \le R_{\mathrm{Jup}}$[2]. For this reason, in the following figures we use a logarithmic scale for θ, to emphasize the small values. Fig. (2-4) shows the (suitably normalized) elements of the covariance matrix as a function of θ.

In the limits $\eta \rightarrow 0$ (errorless knowledge of f_0) and $\theta \rightarrow r$ (small impact parameter), the expressions for the standard errors are especially simple:

$$\begin{aligned} \sigma_{t_c} &= Q^{-1}T\sqrt{\theta/2}, \\ \sigma_{\tau} &\approx Q^{-1}T\sqrt{6\theta}, \\ \sigma_T &\approx Q^{-1}T\sqrt{2\theta}, \end{aligned}$$

---

[2] In fact, the fraction of discovered systems with $\theta \lesssim 0.3$ may be even larger than 90%, because selection effects make it harder to detect grazing transits.

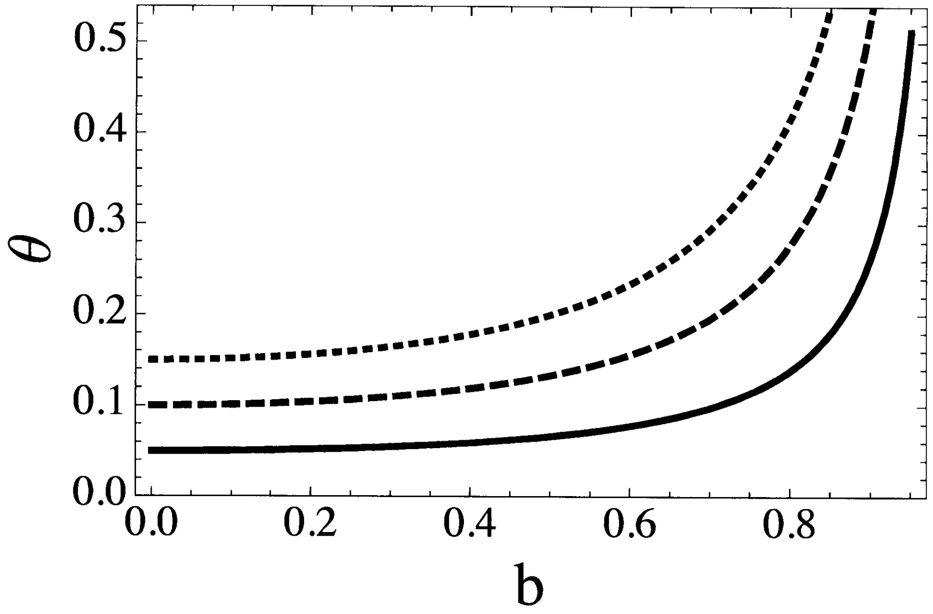

Figure 2-3 Dependence of $\theta = \frac{\tau}{T}$ on depth $\delta = r^2$ and normalized impact parameter b, for the cases $r = 0.05$ (solid line), $r = 0.1$ (dashed line), and $r = 0.15$ (dotted line).

$$\sigma_\delta \approx Q^{-1}\delta. \tag{2.23}$$

In this regime, we have a clear hierarchy in the precision with which the time parameters are known, with $\sigma_{t_c} < \sigma_T < \sigma_\tau$.

To further quantify the degree of correlation among the parameters, we compute the correlation matrix,

$$\mathrm{Corr}(\{t_c, \tau, T, \delta, f_0\}\ , \{t_c, \tau, T, \delta, f_0\}) = \left\{ \frac{\mathrm{Cov}(i,j)}{\sqrt{\mathrm{Cov}(i,i)\mathrm{Cov}(j,j)}} \right\} =$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{(\beta-1)\theta}{\sqrt{(6-\theta(5-\beta))(2-\theta(1-\beta))}} & \sqrt{\frac{(\beta+1)\theta}{6-\theta(5-\beta)}} & \sqrt{\frac{\beta\theta}{6-\theta(5-\beta)}} \\ 0 & \frac{(\beta-1)\theta}{\sqrt{(6-\theta(5-\beta))(2-\theta(1-\beta))}} & 1 & \frac{(\beta-1)\sqrt{\theta}}{\sqrt{(\beta+1)(2-\theta(1-\beta))}} & \sqrt{\frac{\beta\theta}{2-\theta(1-\beta)}} \\ 0 & \sqrt{\frac{(\beta+1)\theta}{6-\theta(5-\beta)}} & \frac{(\beta-1)\sqrt{\theta}}{\sqrt{(\beta+1)(2-\theta(1-\beta))}} & 1 & \sqrt{\frac{\beta}{\beta+1}} \\ 0 & \sqrt{\frac{\beta\theta}{6-\theta(5-\beta)}} & \sqrt{\frac{\beta\theta}{2-\theta(1-\beta)}} & \sqrt{\frac{\beta}{\beta+1}} & 1 \end{pmatrix} \tag{2.24}$$

where we have defined $\beta \equiv \eta(1-\theta)$ to simplify the resulting expression. For $\theta \to 0$, all correlations with f_0 vanish except for the correlation with δ. Due to the fact the

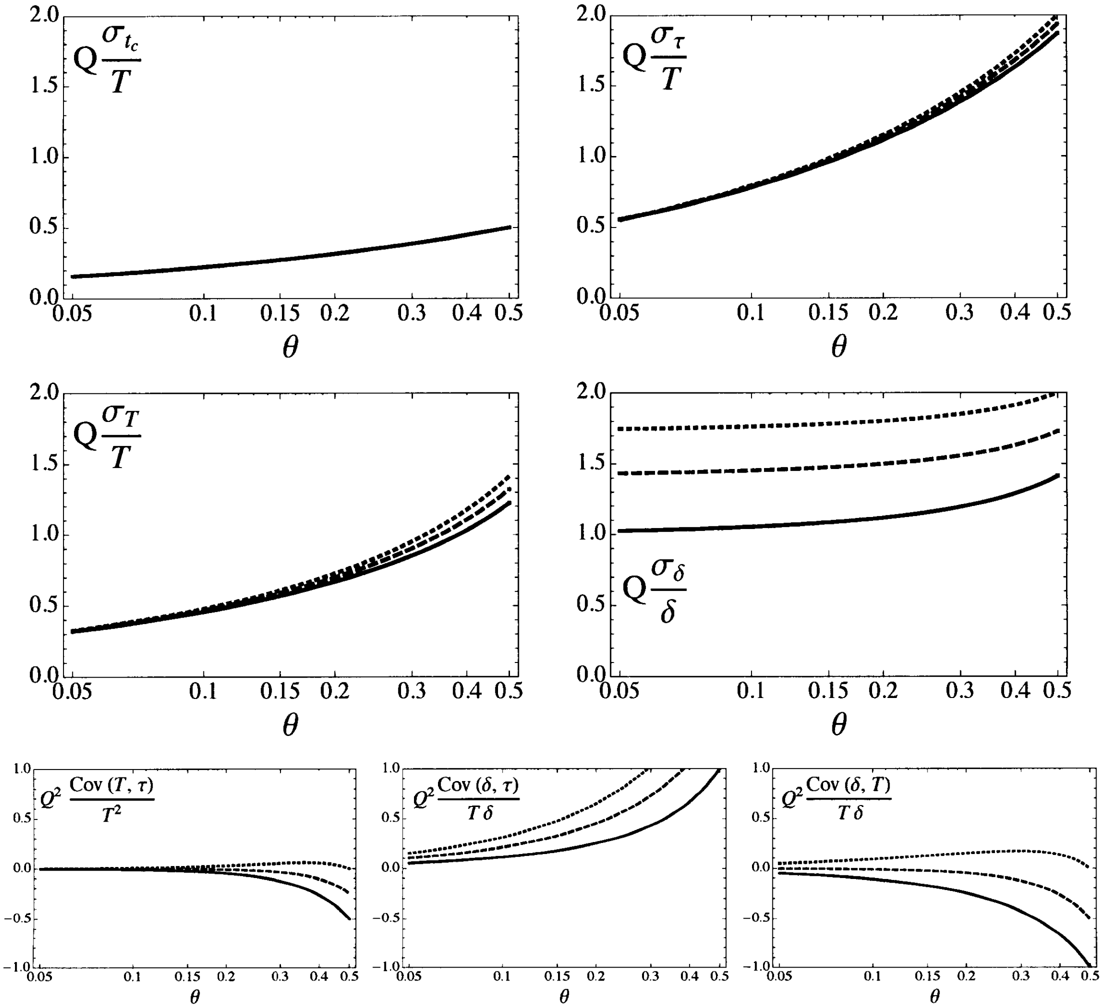

Figure 2-4 Standard errors and covariances, as a function of $\theta \equiv \tau/T$, for different choices of η. The analytic expressions are given in Eqn. (2.20). The definitions of η, θ, and Q are given in Eqn. (2.19). Solid line – $\eta = 0$; Dashed line – $\eta = 0.5$; Dotted line – $\eta = 1$.

correlation between δ and f_0 is $\propto \beta^{1/2}$, it remains large even for fairly small β. In the limit of $\eta \to 0$ ($\beta \to 0$), we remove all correlations with f_0 and have the remaining correlations depending only on the ratio θ:

$$\lim_{\eta\to 0} \mathrm{Corr}(\cdot\,,\cdot) \;=\; \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -\frac{\theta}{\sqrt{(6-5\theta)(2-\theta)}} & \sqrt{\frac{\theta}{6-5\theta}} & 0 \\ 0 & -\frac{\theta}{\sqrt{(6-5\theta)(2-\theta)}} & 1 & -\sqrt{\frac{\theta}{2-\theta}} & 0 \\ 0 & \sqrt{\frac{\theta}{6-5\theta}} & -\sqrt{\frac{\theta}{2-\theta}} & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (2.25)$$

Correlations with f_0 decline with η as $\sqrt{\eta}$.

In Fig. (2-5), we have plotted the nonzero correlations as a function of θ for a few choices of η. The special case of $\eta \to 0$ is plotted in Fig. (2-6). In the $\eta \to 0$ limit, all correlations are small ($\lesssim 0.3$) over a large region of the parameter space. Thus, our choice of parameters provides a weakly correlated set for all but grazing transits ($\theta \sim 1/2$), as noted during the numerical analysis of particular systems by Burke et al. (2007) and Bakos et al. (2007). One naturally wonders whether a different choice of parameters would give even smaller (or even zero) correlations. In § 2.6 we present parameter sets that are essentially uncorrelated and have other desirable properties for numerical parameter estimation algorithms.

The analytic formalism given in this section and more specifically the simple analytic covariance matrices in Eqns. (2.20, 2.21) provide a toolbox with which to evaluate the statistical merits of any parameter set that can be written in terms of our parameters. In § 2.5 this technique is defined and applied to produce analytic formulas for variances, covariances and uncertainties in several interesting parameters.

2.4 Accuracy of the covariance expressions

Before investigating other parameter sets, it is necessary to examine the validity of Eqns. (2.20, 2.21, 2.24, 2.25) when compared to similar quantities derived from more

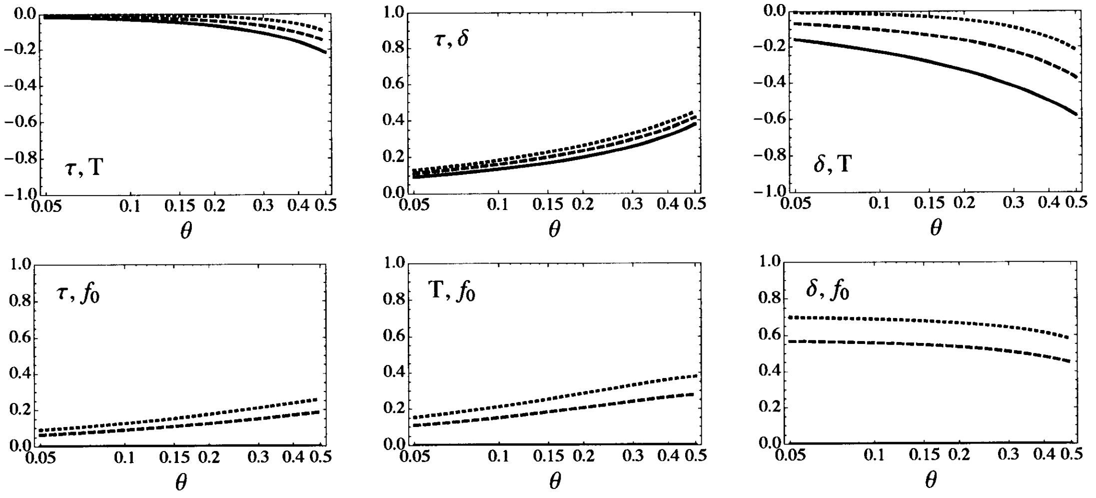

Figure 2-5 Correlations of the piecewise-linear model parameters, as a function of $\theta \equiv \tau/T$ for different choices of η. Solid line – $\eta = 0$; Dashed line – $\eta = 0.5$; Dotted line – $\eta = 1$.

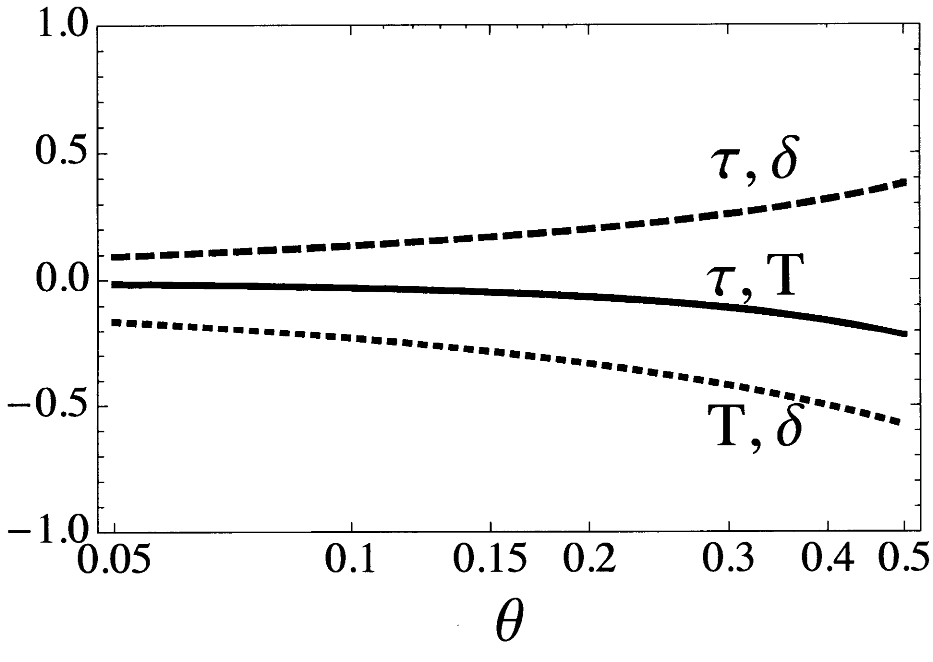

Figure 2-6 Correlations of the piecewise-linear model parameters, as a function of $\theta \equiv \tau/T$, for the case $\eta \to 0$ (errorless knowledge of the out-of-transit flux). Solid line – Corr(τ, T). Dashed line – Corr(τ, δ). Dotted line – Corr(T, δ).

realistic transit light curve models. The utility of the covariance matrix in Eqn. (2.20) depends on the accuracy of the integral approximation of Eqn. (2.17), and on the fidelity with which the parameter dependences of the piecewise-linear model mimic the dependences of the exact uniform-source model. In this section we investigate these two issues.

2.4.1 Finite cadence correction

The case of a finite observing cadence, rather than continuous sampling, can be analyzed by evaluating the exact sums of Eqn. (2.16). Generally, given a sampling rate Γ, we expect the integral approximation in Eqn. (2.17) to be valid to order $(\Gamma\tau)^{-1}$. In the $\eta \to 0$ limit we may evaluate the exact sums, under the assumption of a uniform sampling rate, with data points occurring exactly at the start and end of the ingress (and egress) phases as well as at some intermediate times. This directly summed covariance, $\mathrm{Cov}_{\mathrm{sum}}$, is related to the integral-approximation covariance Eqn. (2.21) as

$$\mathrm{Cov}_{\mathrm{sum}}(\cdot,\cdot) = \mathrm{Cov}(\cdot,\cdot) + 6\left(\frac{T}{Q}\right)^2 \frac{\theta}{1-\epsilon^2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \epsilon^2 & \epsilon & 0 \\ 0 & \epsilon & \epsilon^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.26}$$

where $\epsilon = (\Gamma\tau)^{-1}$.

The quantity $\Gamma\tau$ is approximately the number of data points obtained during ingress or egress. It is evident from Eqn. (2.26) that for this sampling scheme only the variances of T and τ along with their covariance are corrected. The corrections to the variances and covariance are $O(\epsilon^2)$ and $O(\epsilon)$ respectively.

2.4.2 Comparison with covariances of the exact uniform-source model

We tested the accuracy of the covariance matrix based on the piecewise-linear model by (1) performing a numerical Fisher analysis of the exact uniform-source model, and also (2) applying a Markov Chain Monte Carlo (MCMC) analysis of simulated data based on the exact uniform-source model. In both analyses, orbits are assumed to be circular. For the first task, we evaluated the analytic parameter derivatives of Eqn. (2.1), which are too cumbersome to be worth reproducing here, and numerically integrated Eqn. (2.17) to generate covariance matrices over a wide range of parameter choices. Fig. (2-2), in § 2, shows the parameter derivatives for the exact model, as well as the piecewise-linear model and some limb-darkened light curves. For the second task, idealized data was generated by adding Gaussian noise with standard deviation $\sigma/f_0 = 5 \times 10^{-4}$ to Eqn. (2.1) sampled at $\Gamma = 100$ (in units of the characteristic timescale τ_0, Eqn. (2.7)). With this sampling rate, approximately 50 samples occur during the ingress and egress phases. Approximately 10^4 links per parameter were generated with a Gibbs sampler and a Metropolis-Hasting jump-acceptance criterion. The jump-success fraction (the fraction of jumps in parameter space that are actually executed) was approximately 25% for all parameters. The effective length, defined as the ratio of the number of links to the correlation length (see the end of § 2.6 for the exact definition), was roughly $1000 - 2000$ for the piecewise-linear model parameter set. More details on the MCMC algorithm are given by Tegmark et al. (2004) and Ford (2005). Standard errors were determined by computing the standard deviation of the resulting distribution for each parameter. The Fisher-information analysis should mirror the MCMC results, as long as the log-likelihood function is well approximated as quadratic near the mean (Gould 2003).

The numerical Fisher analysis was performed for $\eta = 0$ and $0.05 \leq \theta \lesssim 1/2$. In practice this was done by choosing $r = 0.05$ and varying b across the full range of impact parameters. (The numerical analysis confirmed that the suitably-normalized covariances vary only as a function of $\theta \equiv \tau/T$, with the exception of slight δ-dependent positive offset in σ_δ that goes to zero as δ goes to zero.) The MCMC analysis for $\eta = 0$ was accomplished by fixing the out of transit flux, $f_0 = 1$, and varying the remaining parameters We chose $r = 0.1$ for the MCMC analysis. Fig. (2-7) shows all of the nonzero numerical correlation matrix elements, as a function of θ. The MCMC results, plotted as solid symbols, closely follow the curves resulting from the numerical Fisher analysis. Fig. (2-8) shows the nonzero numerical covariance matrix elements, also for the case $\eta = 0$.

The correlations of the piecewise-linear model match the correlations of the exact model reasonably well, with the most significant deviations occurring only in the grazing limit, $\theta \sim 1/2$. We have also confirmed that a similar level of agreement is obtained for nonzero η, although for brevity those results are not shown here. We concluded from these tests that the errors in the analytic estimates of the uncertainties are generally small enough for the analytic error estimates derived from the piecewise-linear model to be useful.

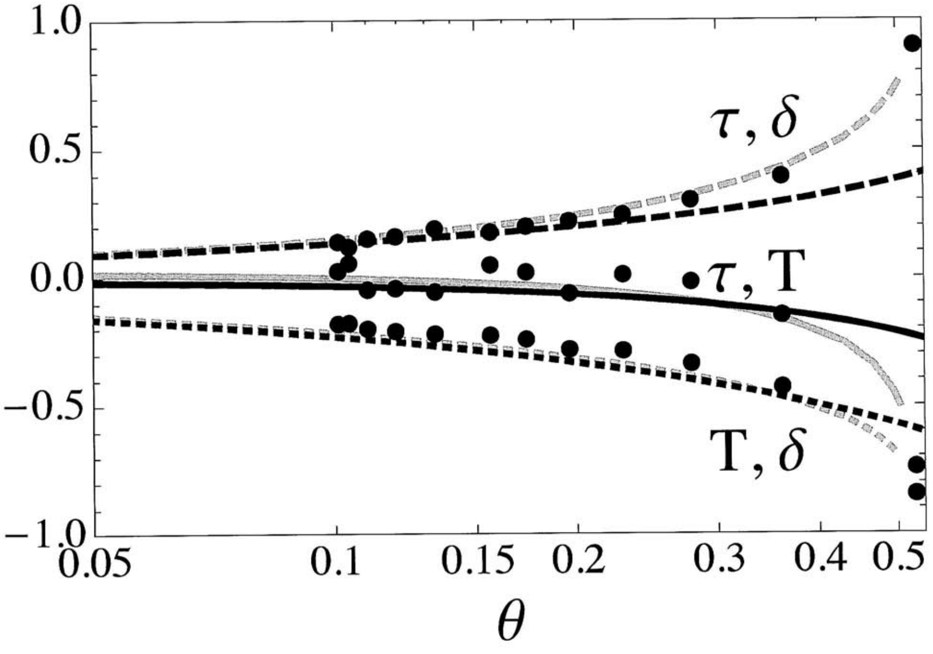

Figure 2-7 Comparison of the non-zero correlation matrix elements for the exact light-curve model and the piecewise-linear model, as a function of $\theta \equiv \tau/T$, for $\eta \to 0$. Black curves: correlations for the piecewise-linear model. Gray curves: correlations for the exact uniform-source model. Black dots: correlations based on an MCMC analysis of simulated data with Gaussian noise ($r = 0.1$).

2.4.3 The effects of limb darkening

The piecewise-linear function of Eqn. (2.11) was constructed as a model of a transit across a stellar disk of uniform brightness, with applications to far-red and infrared

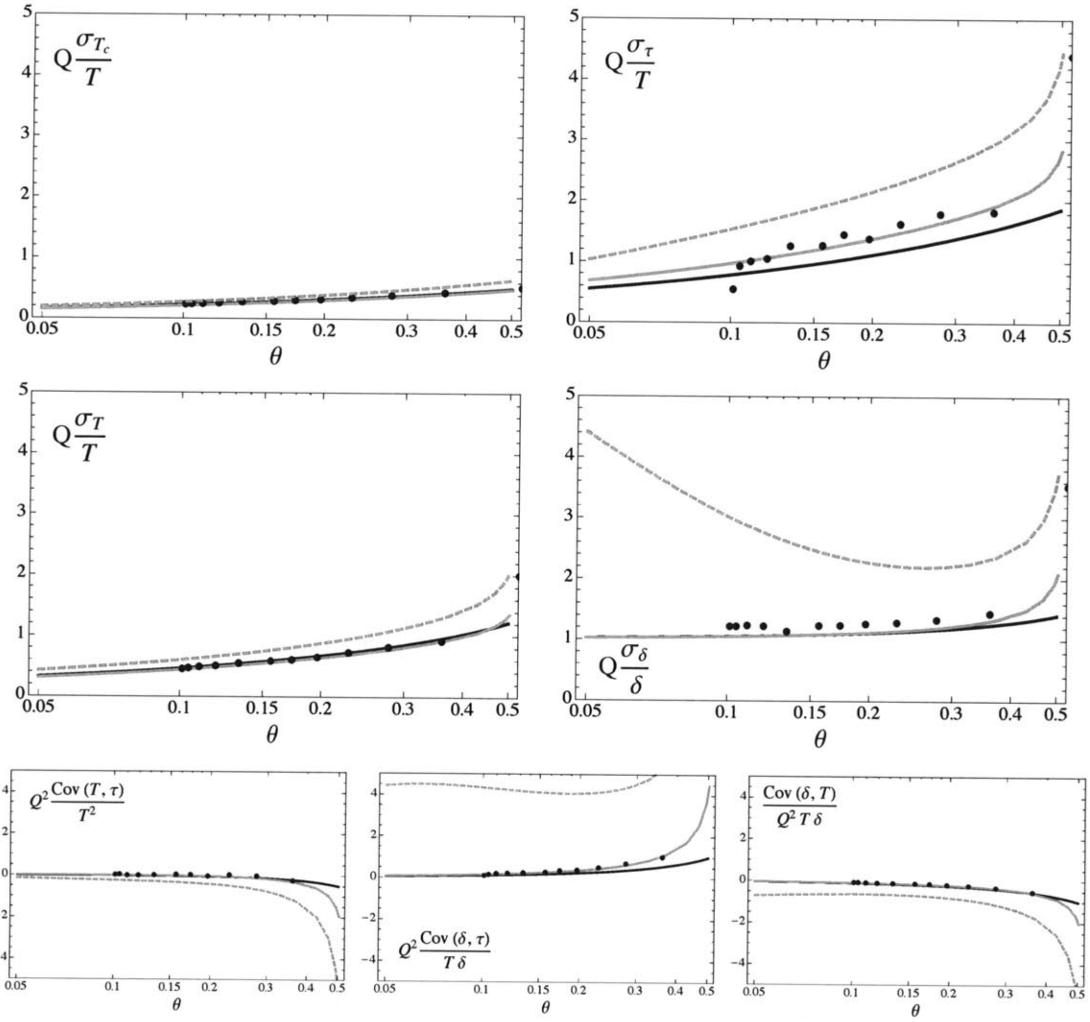

Figure 2-8 Comparison of the covariance matrix elements for the exact uniform-source model, linear limb-darkened model, and the piecewise-linear model, as a function of $\theta \equiv \tau/T$, for $\eta \to 0$. Black curves: covariances for the piecewise-linear model. Gray curves: covariances for the exact model with linear limb-darkening coefficient $u = 0$ (solid) and $u = 0.5$ (dashed). Black dots: covariances as determined by a MCMC analysis of simulated data with Gaussian noise ($u = 0$ and $r = 0.1$). The dimensionless number $Q \equiv \sqrt{\Gamma T}\delta/\sigma$ (see Eqn. 2.19) is approximately the signal-to-noise ratio of the transit.

photometry in mind. At shorter wavelengths, the limb darkening of the star is important. How useful are the previously derived results for this case, if at all? We used the limb-darkened light-curve models given by Mandel & Agol (2002) to answer this question.

To simplify the analysis we adopted a "linear" limb-darkening law, in which the surface brightness profile of the star is

$$\frac{I(z)}{I_0} = 1 - u\left(1 - \sqrt{1 - z^2}\right) \tag{2.27}$$

where u is the linear limb-darkening parameter. Claret (2000) finds values of u ranging from 0.5–1.2 in $UBVR$ for a range of main-sequence stars. Longer wavelength bands correspond to a smaller u for the same surface gravity and effective temperature. Solar values are $u \approx 0.5$ in the Johnson R band and 0.2 in the K band. Fig. (2-2) of § 2 shows the time-dependence of the parameter derivatives of a linear limb-darkened light curve, for the two cases $u = 0.2$ and $u = 0.5$, to allow for comparison with the corresponding dependences of the piecewise-linear model and the exact model with no limb darkening.

From the differences apparent in this plot, one would expect increased correlations (larger than our analytic formulas would predict) between the transit depth and the two timescales τ and T. This is borne out by our numerical calculations of the covariance matrix elements, which are plotted in Figs. (2-8,2-9). The analytic formulas underpredict the variances in δ and τ by a factor of a few, and they also severely underpredict the correlation between those parameters.

It is possible to improve the agreement with the analytic formulas by associating δ with the minimum of the transit light curve, rather than the square of the radius ratio. Specifically, one replaces the definition of Eqn. (2.12) with the new definition

$$\delta = f_0 r^2 \, \frac{9 - 8\left(\sqrt{1 - b^2} - 1\right) u}{9 - 8u}. \tag{2.28}$$

For the previously-derived formulas to be valid, we must adopt a value for u based on

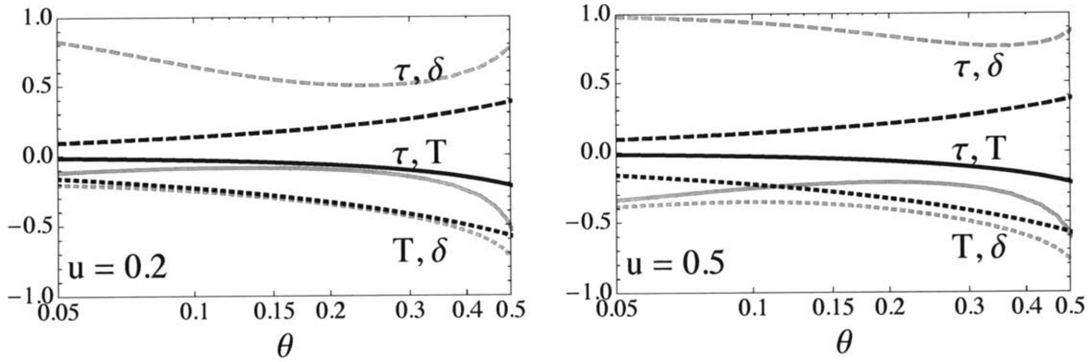

Figure 2-9 Comparison of the analytic correlations (black lines; Eqn. 2.24) numerically-calculated correlation matrix elements for a linear limb-darkened light curve (gray lines), as a function of $\theta \equiv \tau/T$, for $\eta \to 0$. Linestyles follow the conventions of Fig. (2-7).

other information about the parent star (its spectral energy distribution and spectral lines, luminosity, etc.) rather than determining u from the photometric data. Fig. (2-10) shows the correlations resulting from this new association, for the case $u = 0.5$. Fig. (2-11) shows the improvement with this new association for the variance in δ and the covariance between δ and τ, for the case $u = 0.5$. While this new association improves on the agreement with the analytic covariances (particularly at low normalized impact parameter), a disadvantage is that we no longer have a closed-form mapping from $\{\delta, T, \tau\}$ back to the more physical parameters $\{r, b, \tau_0\}$.

It should be noted that there is evidence that linear limb darkening may not adequately fit high-quality transit light curves relative to higher order models (Brown et al. (2001), Southworth (2008)). A more complete analysis with arbitrary source surface brightness would minimally include quadratic limb darkening but is outside the scope of this discussion. Pál (2008) completes a complementary analysis to this one of uncertainties in the quadratic limb darkening parameters themselves.

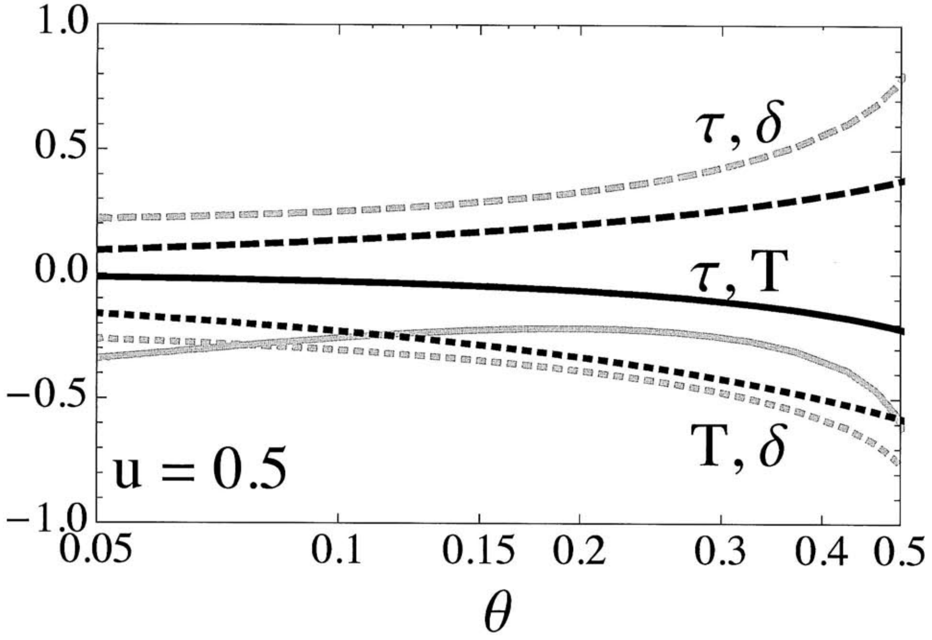

Figure 2-10 Comparison of correlation matrix elements for the piecewise-linear model (black curve) and a linear limb-darkened light curve ($u = 0.5$; gray curve), as a function of $\theta \equiv \tau/T$. Here, the δ parameter has been redefined as the minimum of the limb-darkened light curve, as approximated by Eqn. (2.28). Linestyles follow the conventions of Fig. (2-7).

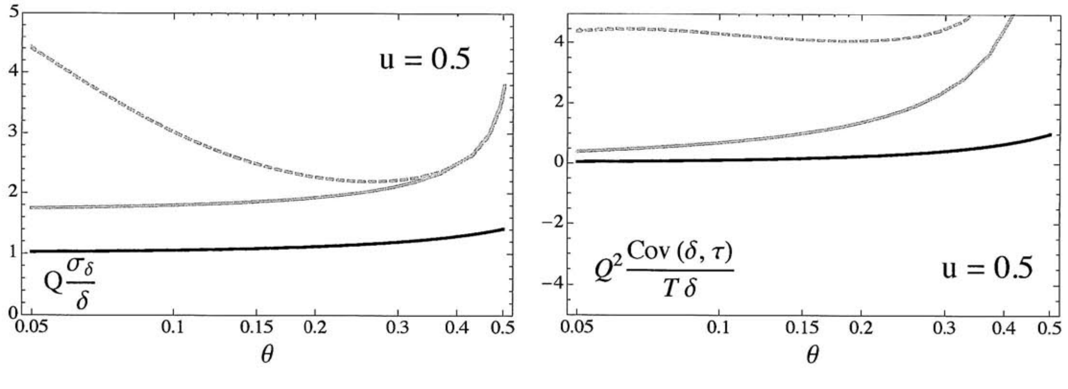

Figure 2-11 Comparison of select covariance matrix elements for the piecewise-linear model (black curve) and a linear limb-darkened light curve ($u = 0.5$; gray curves), as a function of $\theta \equiv \tau/T$. The δ parameter has been redefined as the minimum of the limb-darkened light curve, as approximated by Eqn. (2.28), in the solid gray curve. The dashed gray curve uses the initial δ association, as defined in Eqn. (2.8). Linestyles follow the conventions of Fig. (2-8).

2.5 Errors in derived quantities of interest in the absence of limb darkening

The parameters $\{t_c, \tau, T, \delta, f_0\}$ are preferred mainly because they lead to simple analytic formulas for their uncertainties and covariances. The values of these parameters are also occasionally of direct interest. In particular, when planning observations, it is useful to know the transit duration, depth, and the predicted midtransit time. Of more direct scientific interest are the values of the "physical" parameters, such as the planetary and stellar radii, the orbital inclination, and the mean density of the star. Those latter parameters also offer clearer *a priori* expectations, such as a uniform distribution in $\cos i$.

For affine parameter transformations $p \mapsto p'$, we may transform the covariance matrix C via the Jacobian $J = \frac{\partial p'}{\partial p}$ as

$$C' = J^{\mathrm{T}} C J \,. \tag{2.29}$$

Using Eqns. (2.12–2.14), we may calculate the Jacobian

$$\frac{\partial\{t_c, b^2, \tau_0^2, r, f_0\}}{\partial\{t_c, \tau, T, \delta, f_0\}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{rT}{\tau^2} & \frac{T}{4r} & 0 & 0 \\ 0 & -\frac{r}{\tau} & \frac{\tau}{4r} & 0 & 0 \\ 0 & -\frac{T}{2f_0 r\tau} & -\frac{T\tau}{8f_0 r^3} & \frac{1}{2f_0 r} & 0 \\ 0 & \frac{pT}{2f_0\tau} & \frac{T\tau}{8f_0 r} & -\frac{r}{2f_0} & 1 \end{pmatrix} \tag{2.30}$$

between the parameters of the piecewise-linear model and the more physical parameter set when limb darkening is negligible. Using this Jacobian, the transformed covariance matrix is

$$\mathrm{Cov}'(\{b^2, \tau_0^2, r, f_0\} \, , \{b^2, \tau_0^2, r, f_0\}) =$$

$$\frac{1}{Q^2}\begin{pmatrix} \frac{24-\theta(4(\theta-3)\theta+23)}{4(1-\theta)\theta^3}r^2 & \frac{24-\theta(23-4(\theta-2)\theta)}{16(1-\theta)\theta}T^2 & \frac{2\theta+1}{4\theta(1-\theta)}r^2 & 0 \\ \frac{24-\theta(23-4(\theta-2)\theta)}{16(1-\theta)\theta}T^2 & \frac{24-\theta(4(\theta-1)\theta+23)}{64r^2(1-\theta)}\theta T^4 & \frac{1-2\theta}{16(1-\theta)}\theta T^2 & 0 \\ \frac{2\theta+1}{4\theta(1-\theta)}r^2 & \frac{1-2\theta}{16(1-\theta)}\theta T^2 & \frac{1}{4(1-\theta)}r^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} +$$

$$\frac{\eta}{Q^2}\begin{pmatrix} \frac{(1-2\theta)^2}{4\theta^2}r^2 & \frac{1}{16}\left(1-4\theta^2\right)T^2 & \frac{(1-2\theta)}{4\theta}r^2 & \frac{r^3(1-2\theta)}{2\theta}f_0 \\ \frac{1}{16}\left(1-4\theta^2\right)T^2 & \frac{\theta^2(1+2\theta)^2}{64r^2}T^4 & \frac{1}{16}\theta\left(1+2\theta\right)T^2 & \frac{1}{8}r\theta\left(1+2\theta\right)f_0T^2 \\ \frac{(1-2\theta)}{4\theta}r^2 & \frac{1}{16}\theta\left(1+2\theta\right)T^2 & \frac{1}{4}r^2 & \frac{1}{2}r^3f_0 \\ \frac{r^3(1-2\theta)}{2\theta}f_0 & \frac{1}{8}r\theta\left(1+2\theta\right)f_0T^2 & \frac{1}{2}r^3f_0 & r^4f_0^2 \end{pmatrix} \tag{2.31}$$

where we have ignored the unmodified covariance elements involving t_c, and have kept only the leading-order terms in r in the η-dependent matrix.

The standard errors for other functions of the parameters, $f(\{p_i\})$, can be found via error propagation, just as in Eqn. (2.29),

$$\mathrm{Var}[f(\{p_i\})] = \sum_i\sum_j \mathrm{Cov}(p_i,p_j)\frac{\partial f}{\partial p_i}\frac{\partial f}{\partial p_j}. \tag{2.32}$$

The results for several interesting and useful functions, such as the mean densities of the star and planet, are given in Table (2.2). For brevity, the results are given in terms of the matrix elements of Eqn. (2.31). Simplified expressions for covariance matrix elements in the limit of $\eta \to 0$, θ small (plentiful out-of-transit data) and negligible limb darkening are given in Table (2.3).

2.6 Optimizing parameter sets for fitting data with small limb darkening

The parameter set $\{t_c, \tau, T, \delta, f_0\}$ has the virtues of simplicity and weak correlation over most of the physical parameter space. However, when performing numerical analyses of actual data, the virtue of simplicity may not be as important as the virtue of low correlation, which usually leads to faster and more robust convergence. To take one example, lower correlations among the parameters result in reduced correlation

Quantity	Variance (Standard Error Squared)	Notes
$R_p = rR_\star$	$R_p^2\,[\mathrm{Var}(r)/r^2 + (\log M_\star/M_\odot)^2\mathrm{Var}(x)]$	1
$R_\star/a = (\gamma_1/\gamma_2)2\pi\tau_0/P$	$\frac{1}{4}(R_\star/a)^2\mathrm{Var}(\tau_0^2)/\tau_0^4$	
$R_p/a = (\gamma_1/\gamma_2)2\pi\tau_0 r/P$	$(R_p/a)^2\left[\frac{1}{4}\mathrm{Var}(\tau_0^2)/\tau_0^4 + \mathrm{Var}(r)/r^2\right]$	
$\lvert b\rvert = (\gamma_2^2/\gamma_1)\lvert a\cos i/R_\star\rvert$	$\frac{1}{4}\mathrm{Var}(b^2)/b^2$	
$\lvert\cos i\rvert$ $= (\gamma_1^2/\gamma_2^3)2\pi\tau_0\,\lvert b\rvert\,/P$	$\frac{1}{4}\cos^2 i\,[\mathrm{Var}(b^2)/b^4 + \mathrm{Cov}(\tau_0^2, b^2)/\tau_0^2 b^2 + \mathrm{Var}(\tau_0^2)/\tau_0^4]$	
$\rho_\star$ $= (\gamma_2/\gamma_1)^3(3/8G\pi^2)P/\tau_0^3$	$\frac{9}{4}\rho_\star^2\mathrm{Var}(\tau_0^2)/\tau_0^4$	
ρ_p $= \gamma_2(K_\star\rho_\star/r^3\sin i)(P/2\pi GM_\star)^{1/3}$	$\rho_p^2\left[\frac{9}{4}\mathrm{Var}(\tau_0^2)/\tau_0^4 + 9\mathrm{Var}(r)/r^2 + \frac{9}{2}\mathrm{Cov}(r,\tau_0^2)/r\tau_0^2\right.$ $+\frac{1}{4}(\cos i/b)^4\mathrm{Var}(b^2) - \frac{3}{4}(\cos i/b)^2\mathrm{Cov}(b^2,\tau_0^2)/\tau_0^2$ $\left.-\frac{3}{2}(\cos i/b)^2\mathrm{Cov}(b^2,r)/r + \mathrm{Var}(K_\star)/K_\star^2\right]$	2
$g_\star$ $= (\gamma_2/\gamma_1)^3 R_\star P/(2\pi\tau_0^3)$	$g_\star^2\left[\frac{9}{4}\mathrm{Var}(\tau_0^2)/\tau_0^4 + (\log M_\star/M_\odot)^2\mathrm{Var}(x)\right]$	1
g_p $= (\gamma_2^3/\gamma_1^2)K_\star P/(2\pi r^2\tau_0^2\sin i)$	$g_p^2\,[\mathrm{Var}(\tau_0^2)/\tau_0^4 + 4\mathrm{Var}(r)/r^2 + 2\mathrm{Cov}(r,\tau_0^2)/r\tau_0^2$ $+\frac{1}{4}(\cos i/b)^4\mathrm{Var}(b^2) - \frac{1}{2}(\cos i/b)^2\mathrm{Cov}(b^2,\tau_0^2)/\tau_0^2$ $-(\cos i/b)^2\mathrm{Cov}(b^2,r)/r + \mathrm{Var}(K_\star)/K_\star^2]$	2

Table 2.2 Table of transit quantities and associated variances, in terms of the matrix elements given in Eqn. (2.31). We have assumed that both the orbital period, P, and stellar mass, $M_\star$, are known exactly. We have defined the noncircular-orbit parameters $\gamma_1 \equiv 1 + e\sin\omega$ and $\gamma_2 \equiv \sqrt{1-e^2}$ where e is the eccentricity and ω is the argument of pericenter (see § 2.2 for a discussion of eccentric orbits). Notes: (1) A mass-radius relation $R_\star \propto (M_\star/M_\odot)^x$ is assumed; (2) We have assumed $i \gtrsim 80°$ in simplifying the inclination dependence in the variance. Quantities in bold are not determined by the transit model and must be provided from additional information. $K_\star$ is the semi-amplitude of the source radial velocity. Terms have been arranged in order of relative importance with the largest in absolute magnitude coming first. Refer to Table (2.3) for matrix elements of Eqn. (2.31) for the case in which the planet is small, the out-of-transit flux is known precisely and limb darkening is negligible.

$Q^2\mathrm{Var}(r)/r^2$	$\approx 1/4$	$Q^2\mathrm{Cov}(b^2,\tau_0^2)/b^2\tau_0^2$	$\approx 6r/\theta^2 b^2$
$Q^2\mathrm{Var}(b^2)/b^4$	$\approx 6r^2/\theta^3 b^4$	$Q^2\mathrm{Cov}(b^2,r)/b^2 r$	$\approx r/4\theta b^2$
$Q^2\mathrm{Var}(\tau_0^2)/\tau_0^4$	$\approx 3/2\theta$	$Q^2\mathrm{Cov}(\tau_0^2,r)/\tau_0^2 r$	$\approx 1/16$

Table 2.3 Covariance matrix elements from Eqn. (2.31) in the limit $\eta \to 0$ and θ small for use in Table (2.2). These approximations are valid in the case in which the planet is small, the out-of-transit flux is known precisely and limb darkening is negligible.

lengths for Monte Carlo Markov Chains, and faster convergence to the desired *a posteriori* probability distributions, and can obviate the need for numerical Principal Component Analysis (Tegmark et al., 2004). In Fig. (2-12), we compare the degree of correlations for various parameter sets that have been used in the literature on transit photometry. Of note is the high degree of correlations among the "physical" parameter set $\{R_\star/a, R_p/a, b\}$, which is a poor choice from the point of view of computational speed.

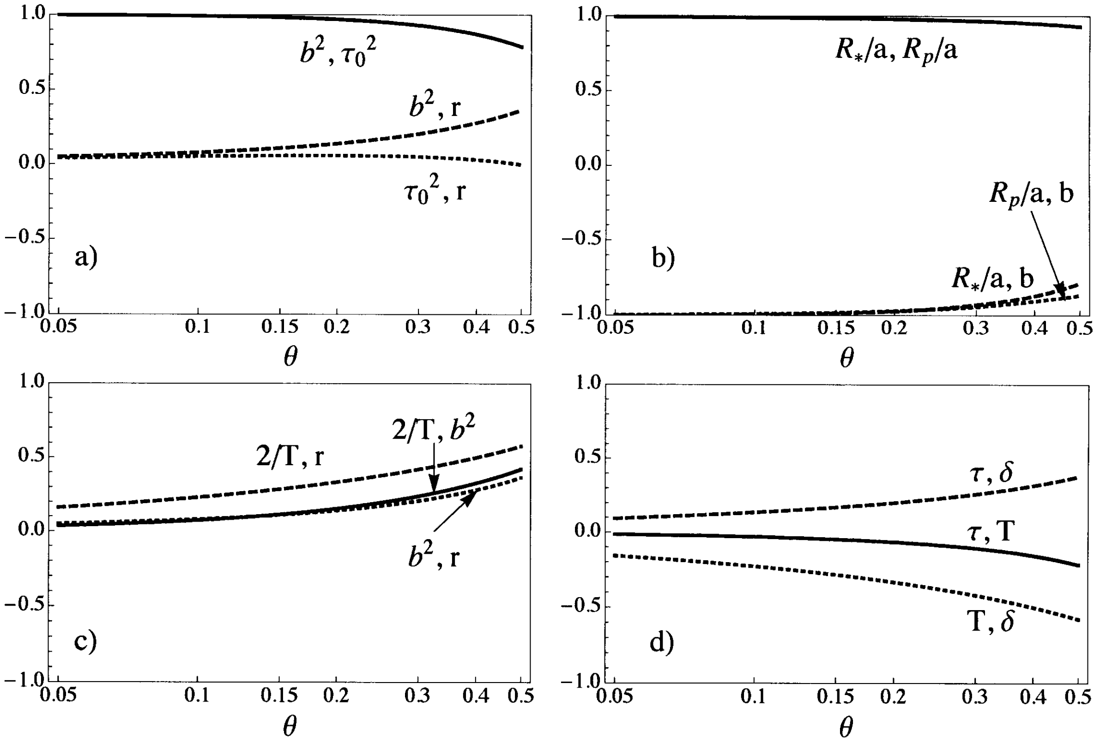

Figure 2-12 Comparison of correlations for various parameter sets that have been used in the literature. The correlations were derived from the piecewise-linear model (Eqn. 2.21) assuming $\eta = 0$. (a) Parameters $\{b^2, \tau_0^2, r\}$. (b) $\{R_\star/a = n\tau_0, R_p/a = n\tau_0 r, b^2\}$. (c) $\{2/T, b^2, r\}$ (e.g., Bakos et al. (2007)). (d) $\{T, \tau, \delta\}$, the set introduced in this chapter.

Nevertheless, one advantage of casting the model in terms of physical parameters is that the *a priori* expectations for those parameters are more easily expressed, such as a uniform distribution in b. The determinant of the Jacobian given by Eqn. (2.29), $|J|$, is also useful in translating *a priori* probability distributions from one parameter set to the other [see Burke et al. (2007) or Ford (2006) for an example of how this is done in practice]. For the case of the parameter set $\{t_c, \tau, T, \delta, f_0\}$, we may use the Jacobian, Eqn. (2.30), to convert *a priori* probability distributions via

$$\begin{aligned}
\mathrm{p}(t_c, \tau, T, \delta, f_0) dt_c\, d\tau\, dT\, d\delta\, df_0 &= \mathrm{p}(t_c, b^2, \tau_0^2, r, f_0) \frac{1}{4\, r\, \theta\, f_0} dt_c\, db^2\, d\tau_0^2\, dr\, df_0 \\
&= \mathrm{p}(t_c, b, \tau_0, r, f_0) \frac{1}{4\, r\, \theta\, f_0} \frac{1}{4\, b\, \tau_0} dt_c\, db\, d\tau_0\, dr\, df_0 \qquad (2.33) \\
&= \mathrm{p}(t_c, b, \tau_0, r, f_0) \left(\frac{1 - b^2}{16\, b\, r^2\, \tau_0\, f_0} \right) dt_c\, db\, d\tau_0\, dr\, df_0 .
\end{aligned}$$

where we have remeasured the phase space volume via the determinant,

$$\left|\left| \frac{\partial\{t_c, b^2, \tau_0^2, r, f_0\}}{\partial\{t_c, \tau, T, \delta, f_0\}} \right|\right| = \frac{1}{4\, r\, \theta\, f_0}. \tag{2.34}$$

One may use this expression to enforce a uniform prior in b, for example, by weighting the likelihood function as shown in Eqn. (2.34). However, there is a practical difficulty due to the singularity at $b = 0$. One way to understand the singularity is to note that uniform distributions in τ, T lead to a nearly uniform distribution in $\theta = \tau/T$, which highly disfavors $b = 0$; in order to enforce a uniform distribution in b, the prior must diverge at low b. Fig. (2-3) graphically captures the steep variation for small b with θ. Consider, instead, the parameter set $\{t_c, b, T, r \equiv \sqrt{\delta/f_0}, f_0\}$ where, from Eqn. (2.13), $b^2 = 1 - rT/\tau$. We may calculate the determinant of the Jacobian (not reproduced here)

$$\left|\left| \frac{\partial\{t_c, b, T, r, f_0\}}{\partial\{t_c, \tau, T, \delta, f_0\}} \right|\right| = \frac{(1 - b^2)^2}{4b\, r^2 f_0 T}. \tag{2.35}$$

Combining this result with Eqn. (2.34)

$$\begin{aligned}
\mathrm{p}(t_c, b, T, r, f_0) dt_c\, db\, dT\, dr\, df_0 &= \mathrm{p}(t_c, b, \tau_0, r, f_0) \frac{1}{1-b^2} \frac{T}{4\tau_0} dt_c\, db\, d\tau_0\, dr\, df_0 \\
&= \mathrm{p}(t_c, b, \tau_0, r, f_0) \frac{1}{2\sqrt{1-b^2}} dt_c\, db\, d\tau_0\, dr\, df_0 \quad (2.36)
\end{aligned}$$

The singularity at $b = 0$ has been removed with this parameter choice. There is a singularity at $b = 1$ instead, which is only relevant for near-grazing transits, and is not as strong of a singularity because of the square root. We confirm that this parameter set also enjoys weak correlations, as shown in Fig. (2-13), and therefore this set is a reasonable choice for numerical parameter-estimation algorithms. The merits of other parameter sets, from the standpoint of correlation and *a priori* likelihoods, may be weighed in a similar fashion, using the simple analytic covariance matrix of Eqn. (2.20), and the appropriate transformation Jacobian, in combination with Eqn. (2.29).

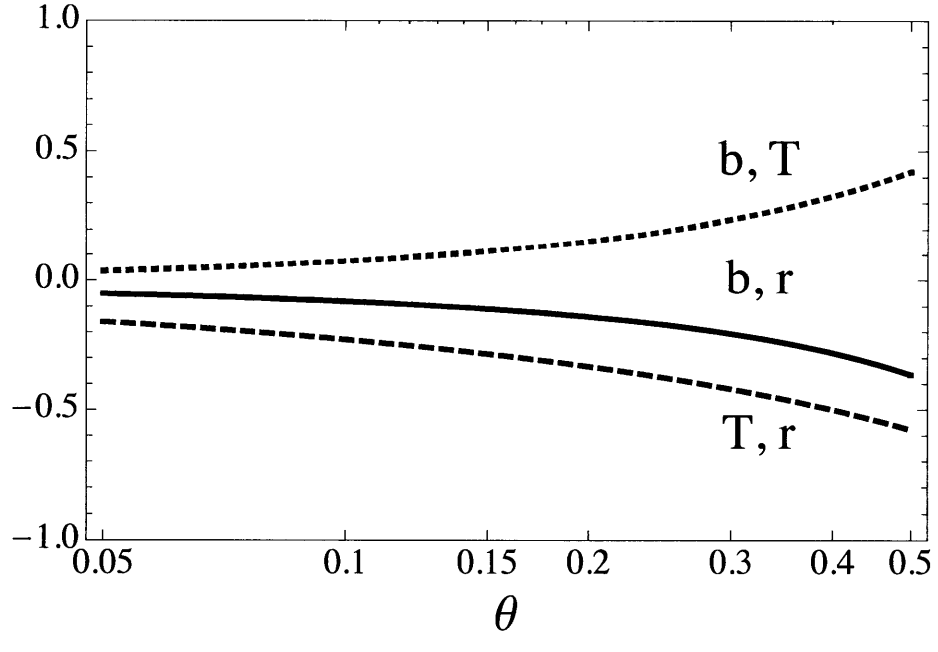

Figure 2-13 Correlations for the parameter set $\{b, T, r\}$. The correlations were derived from the piecewise-linear model (Eqn. 2.21) assuming $\eta = 0$.

If the issues associated with the transformation of priors are ignored (i.e. if the data are of such quality that the results will depend negligibly on the priors), we can give essentially uncorrelated parameter sets. Consider, for example, the parameter set $\{t_c, S_e \equiv \delta/\tau, T, A \equiv \delta T\}$. The new parameter S_e is the magnitude of the slope of the light curve during the ingress and egress phases, and the new parameter A is the area of the trapezoid defined by the transit portion of the light curve (i.e., the time integral of the flux decrement). For simplicity we assume $\eta = 0$ and fix $f_0 = 1$. The transformed correlation (Eqn. (2.25)) is found via the transformation Jacobian, Eqn. (2.29) as

$$\begin{aligned}\mathrm{Corr}(\{t_c, S_e, T, A\} \quad &, \quad \{t_c, S_e, T, A\}) = \\ &\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \sqrt{\frac{\theta(1-\theta)}{(2-\theta)(\theta+1)}} \\ 0 & 0 & \sqrt{\frac{\theta(1-\theta)}{(2-\theta)(\theta+1)}} & 1 \end{pmatrix}\end{aligned} \tag{2.37}$$

The determinant of the transformation Jacobian (for use with Eqn. (2.34)) is given as

$$\left\| \frac{\partial\{t_c, S_e, T, A\}}{\partial\{t_c, \tau, T, \delta\}} \right\| = \frac{(1-b^2)^2}{T}. \tag{2.38}$$

With this new parameter set, the only nonzero correlation is between T and A, and this correlation is $\lesssim 0.3$ even for grazing transits (see Fig. 2-14). We have found that these parameters provide a nearly optimal set for data fitting when little is known at the outset about the impact parameter of the transit.

It is possible to do even better when the impact parameter is known at least roughly. Consider the parameter set $\{t_c, S_e, \Pi = T\delta^{\tilde{\theta}}, \delta\}$ where S_e is the slope of ingress, and $\tilde{\theta}$ is a constant (whose chosen value will be discussed momentarily). The new parameter Π has no simple physical interpretation. We again assume $\eta = 0$ and $f_0 = 1$. The correlation matrix in this case is

$$\text{Corr}(\{t_c, S_e, \Pi, \delta\} \quad , \quad \{t_c, S_e, \Pi, \delta\}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{(\theta-\tilde{\theta})}{\sqrt{(\theta-\tilde{\theta})^2+2\theta(1-\theta)}} \\ 0 & 0 & \frac{(\theta-\tilde{\theta})}{\sqrt{(\theta-\tilde{\theta})^2+2\theta(1-\theta)}} & 1 \end{pmatrix} \tag{2.39}$$

The determinant of the transformation Jacobian is given as

$$\left|\left|\frac{\partial\{t_c, S_e, \Pi, \delta\}}{\partial\{t_c, \tau, T, \delta\}}\right|\right| = \frac{(1-b^2)^2 r^{2\tilde{\theta}}}{T^2}. \tag{2.40}$$

With this choice, the only nonzero correlation is between Π and δ. If the constant $\tilde{\theta}$ is chosen to be approximately equal to θ, then this sole correlation may be nullified. Thus, if θ is known even approximately at the outset of data fitting—from visual inspection of a light curve, or from the approximation $\theta \approx r$ valid for small planets on non-grazing trajectories—a parameter set with essentially zero correlation is immediately available. As an example, Fig. (2-14) shows the correlation between Π and δ as a function of θ, for the choice $\tilde{\theta} = 0.1$, which has a null at $\theta = 0.1$ as expected.

The utility of this parameter set is not lost if $\tilde{\theta}$ cannot be confidently specified when used with Markov chain Monte Carlo parameter estimation codes. At each chain step i, the next candidate state can be drawn from the candidate transition probability distribution function generated by the above parameter set with $\tilde{\theta} = \theta_{i-1}$. Thus, the Markov chain will explore the parameter space moving along principal axes at each chain step. Additionally, allowing the candidate transition function to vary as the Markov chain explores parameter space may prove useful for low S/N data sets.

As a concrete example of the effectiveness of uncorrelated parameters, we apply the MCMC algorithm to simulated data. For a given choice of the parameter set, we generate chains with a fixed jump-success fraction, and calculate the resulting autocorrelations of the Markov chain. For a particular parameter p (with value p_i at

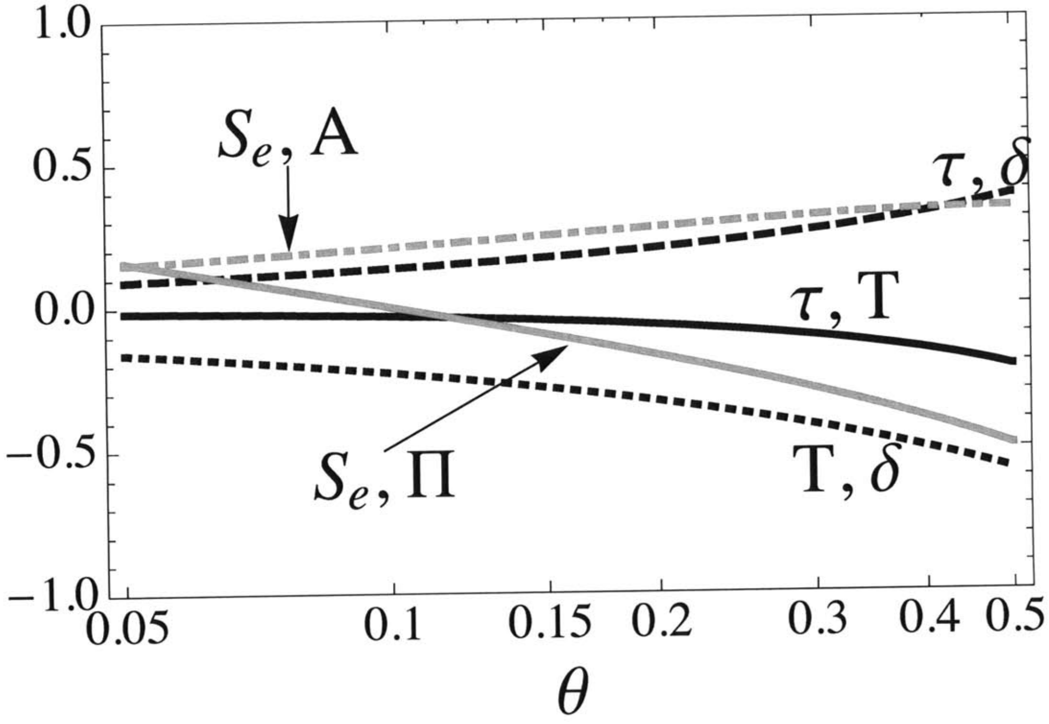

Figure 2-14 Comparison of the correlations among the parameters, for the set $\{\delta, T, \tau\}$ (black lines), the set $\{S_e, T, A = T\delta\}$ (dashed-dot gray line) and the set $\{S_e, \Pi \equiv T\delta^{\tilde{\theta}}, \delta\}$ (solid gray line) for the case $\tilde{\theta} = 0.1$. For the latter set, the only nonzero correlation is between Π and S_e, which vanishes at $\theta = 0.1$.

chain step i), the autocorrelation a at a given chain step j is defined as

$$a_j = \frac{\langle p_i p_{i+j} \rangle - \langle p_i \rangle^2}{\langle p_i^2 \rangle - \langle p_i \rangle^2} \tag{2.41}$$

where the averages refer to the averages over the whole chain (Tegmark et al., 2004). The correlation length of the chain is the number of steps N that are required before the autocorrelation drops below 0.5. The total chain length divided by the correlation length is referred to as the effective length of a chain. The effective chain length is approximately the number of independent samples, which quantifies the degree of convergence of the algorithm. A lower correlation length, for the same total chain length, gives a more accurate final distribution. This autocorrelation analysis was performed for both the "physical" parameter set $\{t_c, b^2, \tau_0^2, r^2\}$ as well as the parameter sets $\{t_c, \tau, T, \delta\}$ and $\{t_c, b, T, r\}$, with $\eta = 0$ in all cases (i.e., plentiful out-of-transit data). The MCMC was executed as detailed in § 2.4.2 with a fixed jump rate $\approx$50% for all parameter chains. (In practice this was achieved by adjusting the size of the Gaussian random perturbation that was added to each parameter at each trial step.) By choosing either the parameter set $\{t_c, \tau, T, \delta\}$ or $\{t_c, b, T, r\}$, the correlation lengths are reduced by a factor of approximately 150. By using the minimally-correlated parameter set $\{t_c, S_e, T, A\}$, the correlation lengths are reduced by an additional factor of $\sim$2.

To completely eliminate the correlations between parameters, one can diagonalize the symmetric covariance matrix, Eqn. (2.37), and find the linear combinations of parameters that eliminates correlations. This was done by Burke et al. (2007) for the particular case of the transiting planet XO-2b. Analytic expressions for the eigenvectors are available because there are only two entangled parameters. However, these eigenvectors are linear combinations of local parameter values; they do not constitute a global transformation rendering the covariance diagonal. Thus, this procedure is useful for numerical analysis of a particular system, although not for analytic insights.

2.7 Summary

We have presented formulas for uncertainties and covariances for a collection of parameters describing the light curve of an exoplanet transiting a star with uniform brightness. These covariances, given in Eqns. (2.20, 2.31), are derived using a Fisher information analysis of a linear representation of the transit light curve. The key inputs are the uncertainty in each measurement of the relative flux, and the sampling rate. We have verified the accuracy of the variance and covariance estimates derived from the piecewise-linear light curve with a numerical Fisher analysis of a more realistic (nonlinear) light-curve model, and with a Markov Chain Monte Carlo analysis of idealized data.

We focused on a particular parameterization of this piecewise-linear light curve that we believe to be most useful. The parameters are the midtransit time (t_c), the out-of-transit flux (f_0), the flux decrement during the full phase of the transit (δ), the duration of ingress or egress (τ), and the duration between the midpoint of ingress and the midpoint of egress (T). This set is observationally intuitive and gives simple analytic formulas for variances and covariances. The exact parameter definitions are provided in Eqns. (2.8, 2.9, 2.10) in terms of normalized impact parameter, stellar and planetary radii, semi-major axis and orbital period. Inverse mappings to more physical parameters are provided in Eqns. (2.12, 2.13, 2.14). The analytic covariance matrix is given in Eqn. (2.20) and the analytic correlation matrix is given in Eqn. (2.24). Some quick-and-dirty (but still rather accurate) expressions for the parameter uncertainties, for the case in which the planet is small, the out-of-transit flux is known precisely and limb darkening is negligible, are given as

$$\begin{aligned}
\sigma_{t_c} &= Q^{-1}T\sqrt{\theta/2}, \\
\sigma_{\tau} &\approx Q^{-1}T\sqrt{6\theta}, \\
\sigma_{T} &\approx Q^{-1}T\sqrt{2\theta}, \\
\sigma_{\delta} &\approx Q^{-1}\delta
\end{aligned}$$

where $\theta \equiv \tau/T$ is the ratio of the ingress or egress duration to the total duration, and $Q \equiv \sqrt{\Gamma T}\frac{\delta}{\sigma}$ is the total signal-to-noise ratio of the transit in the small-planet limit (see Eqn. 2.19).

We investigated the applicability of these results to a limb darkened brightness profile, in which the true light curve is not as well-described by a piecewise-linear function. We found that the analytic formulas underestimate some of the variances and covariances by a factor of a few, for a typical degree of limb darkening at optical wavelengths. Significant improvements to covariance estimates in the limb darkened case may be made by redefining the depth parameter as a function of darkening coefficient and impact parameter as in Eqn. (2.28). Unfortunately, no closed-form mapping to more physical parameters exists with this choice, and therefore most of the appeal of the analytic treatment is lost.

Quantities that are derived in part or in whole from the transit light curve (such as stellar mean density or exoplanet surface gravity) are provided in terms of the suggested parameter set. In Table (2.2), uncertainties propagated from the covariance estimates for these quantities are provided with simple analytic formulas. In Table 2.3, covariance elements relevant to the uncertainties in Table 2.2 are given for the case in which the planet is small and the out-of-transit flux is known precisely. This allows the uncertainty in a given physical parameter to be predicted in advance of any data, bypassing the need for time-consuming simulations. For transit surveys, these formulas may also be useful in giving closed-form expressions for the expected distributions for some of the key properties of a sample of transiting planets.

In § 2.6, with the tools provided, we approach the question of what parameter sets are best suited to numerical parameter estimation codes. This question depends both on the level of parameter correlation and the behavior of any *a priori* likelihood functions. We advocated a parameter set that has the virtue of both weak correlation and essentially uniform *a priori* expectations: specifically, the parameters are the mid-transit time, the out-of-transit flux, the ratio of planetary to stellar radii ($R_p/R_\star$), the normalized impact parameter, and the duration between the midpoint of ingress and the midpoint of egress. Fig. (2-13) graphically describes the parameter correlations while Eqn. (2.36) gives the *a priori* probability distribution. Finally, two parameter choices are given that are less intuitive than the suggested set but that provide smaller correlations, depending on information that may be inferred or guessed prior to analysis. Correlations may be tuned to zero with the second parameter choice for a non-grazing transit and an estimate of $R_p/R_\star$. The resulting correlation matrices for both parameter choices are given in Eqns. (2.37, 2.39). Lower correlations relate directly to more efficient data fitting, as demonstrated by reduced correlation lengths with a Markov Chain Monte Carlo method.

We thank Philip Nutzman for helpful comments on an early version of this draft, and in particular for pointing out the consequences of the singularity in Eqn. (2.34). We also thank the referee for helpful comments, and for suggesting the Markov chain technique for use with the parameter choices in Eqn. (2.39). Sara Seager and Paul Joss also provided helpful comments. We are grateful for support from the William S. Edgerly Innovation Fund and from NASA grant HST-GO-11165 from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

Bibliography

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567

Alonso, R., et al. 2004, ApJ, 613, L153

Bakos, G. Á., et al. 2007, ApJ, 671, L173

Barnes, J. W. 2007, PASP, 119, 986

Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W., & Burrows, A. 2001, ApJ, 552, 699

Burke, C. J., et al. 2007, ApJ, 671, 2115

Burke, C. J. 2008, ArXiv e-prints, 801, arXiv:0801.2579

Claret, A. 2000, A&A, 363, 1081

Deming, D., et al. 2007, ApJ, 667, L199

Diaz-Cordoves, J., Claret, A., & Gimenez, A. 1995, A&AS, 110, 329

Ford, E. B. 2005, AJ, 129, 1706

Ford, E. B. 2006, ApJ, 642, 505

Ford, E. B., Quinn, S. N., & Veras, D. 2008, ArXiv e-prints, 801, arXiv:0801.2591

Ford, E. B., & Holman, M. J. 2007, ApJ, 664, L51

Gillon, M., et al. 2007, A&A, 471, L51

Giménez, A. 2007, A&A, 474, 1049

Gould, A. 2003, ArXiv Astrophysics e-prints, arXiv:astro-ph/0310577

Heyl, J. S., & Gladman, B. J. 2007, MNRAS, 377, 1511

Holman, M. J., & Murray, N. W. 2005, Science, 307, 1288

Holman, M. J., et al. 2006, ApJ, 652, 1715

Knutson, H. A., et al. 2007, Nature, 447, 183

Mandel, K., & Agol, E. 2002, ApJ, 580, L171

Murray, C. D., & Dermott, S. F. 2000, Solar System Dynamics (Cambridge Univ. Press)

Pál, A. 2008, ArXiv e-prints, 805, arXiv:0805.2157

Richardson, L. J., Harrington, J., Seager, S., & Deming, D. 2006, ApJ, 649, 1043

Sato, B., et al. 2005, ApJ, 633, 465

Seager, S., & Mallén-Ornelas, G. 2003, ApJ, 585, 1038

Southworth, J., Wheatley, P. J., & Sams, G. 2007, MNRAS, 379, L11

Southworth, J. 2008, MNRAS, 386, 1644

Tegmark, M., et al. 2004, Phys. Rev. D, 69, 103501

Winn, J. N., et al. 2005, ApJ, 631, 1215

Winn, J. N., Holman, M. J., & Fuentes, C. I. 2007, AJ, 133, 11

Chapter 3

Parameter Estimation from Time-Series Data with Correlated Errors: A Wavelet-Based Method and its Application to Transit Light Curves

3.1 Introduction

Frequently one wishes to fit a parametric model to time-series data and determine accurate values of the parameters and reliable estimates for the uncertainties in those parameters. It is important to gain a thorough understanding of the noise and develop appropriate methods for parameter estimation, especially at the research frontier, where the most interesting effects are often on the edge of detectability. Underestimating the errors leads to unjustified confidence in new results, or confusion over apparent contradictions between different data sets. Overestimating the errors inhibits potentially important discoveries.

When the errors in the data are well understood and uncorrelated, the problem of parameter estimation is relatively straightforward (see, e.g., Bevington & Robinson 2003, Gould 2003, Press et al. 2007). However, when the noise is not well-understood—and particularly when the noise exhibits correlations in time—the problem is more challenging (see, e.g., Koen & Lombard 1993, Beran 1994). Traditional methods that ignore correlations often give parameter estimates that are inaccurate and parameter errors that are underestimated. Straightforward generalization of the traditional methods is computationally intensive, with time-complexity $O(N^2)$ in the worst cases (where N is the number of data points). This makes certain analyses impractical.

Our specific concern in this chapter is the analysis of time-series photometry of exoplanetary transits. During a transit, a planet passes in front of the disk of its parent star, which is evident from the slight diminution in the light received from the star. A model of a transit light curve may have many parameters, but we focus mainly on a single parameter, the midtransit time t_c, for three reasons. The first reason is the simplicity of a single-parameter model. The second reason is that t_c is a unique piece of information regarding each transit event, and as such, the accuracy cannot be improved by combining results from multiple transit observations. Instead one must make the most of single-event observations even if they are afflicted by correlated noise. The third reason is that transit timing offers a means of discovering additional planets or satellites by seeking anomalies in a sequence of transit times due to gravitational perturbations [Holman & Murray (2005), Agol et al. (2005)].[1]

Beginning with the work of Pont, Zucker, & Queloz (2006), it has been widely recognized that time-correlated noise ("red noise") is a limiting factor in the analysis of transit light curves. Many practitioners have attempted to account for correlated errors in their parameter estimation algorithms (see, e.g., Bakos et al. 2006, Gillon et al. 2006; Winn et al. 2007, 2009; Southworth 2008). Among these schemes are

---

[1] The transit duration is also expected to vary in the presence of additional gravitating bodies; see, e.g., Kipping (2009).

the "time-averaging" method, in which the effects of correlations are assessed by computing the scatter in a time-binned version of the data (Pont et al. 2006) and the "residual-permutation" method, a variant of bootstrap analysis that preserves the time ordering of the residuals (Jenkins et al. 2002).

In this chapter we present an alternative method for parameter estimation in the presence of time-correlated noise, and compare it to those two previously advocated methods. The method advocated here is applicable to situations in which the noise is well described as the superposition of two stationary (time-invariant) Gaussian noise processes: one which is uncorrelated, and the other of which has a power spectral density varying as $1/f^\gamma$.

A more traditional approach to time-correlated noise is the framework of autoregressive moving average (ARMA) processes (see, e.g., Box & Jenkins 1976). The ARMA noise models can be understood as complementary to our $1/f^\gamma$ model, in that ARMA models are specified in the time domain as opposed to the frequency domain, and they are most naturally suited for modeling short-range correlations ("short-memory" processes) as opposed to long-range correlations ("long-memory" processes). Parameter estimation with ARMA models in an astronomical context has been discussed by Koen & Lombard (1993), Konig & Timmer (1997), and Timmer et al. (2000). As we will explain, our method accelerates the parameter estimation problem by taking advantage of the discrete wavelet transform. It is based on the fact that a the covariance matrix of a $1/f^\gamma$ noise process is nearly diagonal in a wavelet basis. As long as the actual noise is reasonably well described by such a power law, our method is attractive for its simplicity, computational speed, and ease of implementation, in addition to its grounding in the recent literature on signal processing.

The use of the wavelets in signal processing is widespread, especially for the restoration, compression, and denoising of images (see, e.g., Mallat 1999). Parameter estimation using wavelets has been considered but usually for the purpose of estimating *noise* parameters (Wornell 1996). An application of wavelets to the problem of linear regression with correlated noise was given by Fadili & Bullmore (2002). What is new in this work is the extension to an arbitary nonlinear model, and the application to transit light curves.

This paper is organized as follows. In § 3.2, we review the problem of estimating model parameters from data corrupted by noise, and we review some relevant noise models. In § 3.3 we present the wavelet method and those aspects of wavelet theory that are needed to understand the method. In § 3.4, we test the method using simulated transit light curves, and compare the results to those obtained using the methods mentioned previously. In § 3.5 we summarize the method and the results of our tests, and suggest some possible applications and extensions of this work.

3.2 Parameter estimation with "colorful" noise

Consider an experiment in which samples of an observable y_i are recorded at a sequence of times $\{t_i : i = 1, \ldots, N\}$. In the context of a transit light curve, y_i is the relative brightness of the host star. We assume that the times t_i are known with negligible error. We further assume that in the absence of noise, the samples y_i would be given by a deterministic function,

$$y(t_i) = f(t_i; p_1, \ldots, p_K) = f(t_i; \vec{p}), \quad \text{(no noise)} \tag{3.1}$$

where $\vec{p} = \{p_1, \ldots, p_K\}$ is a set of K parameters that specify the function f. For an idealized transit light curve, those parameters may be the fractional loss of light δ, the total duration T, and ingress or egress duration τ, and the midtransit time t_c, in the notation of Carter et al. (2008). More realistic functions have been given by Mandel & Agol (2002) and Giménez (2007).

We further suppose that a stochastic noise process $\epsilon(t)$ has been added to the data, giving

$$y(t_i) = f(t_i; \vec{p}) + \epsilon(t_i). \quad \text{(with noise)} \tag{3.2}$$

As a stochastic function, $\vec{\epsilon} = \{\epsilon(t_1), \ldots \epsilon(t_N)\}$ is characterized by its joint distribution function $\mathcal{D}(\vec{\epsilon};\vec{q})$, which in turn depends on some parameters $\vec{q}$ and possibly also the times of observation. The goal of parameter estimation is to use the data $y(t_i)$ to calculate credible intervals for the parameters $\vec{p}$, often reported as best estimates $\hat{p}_k$ and error bars $\hat{\sigma}_{p_k}$ with some quantified degree of confidence. The estimate of $\vec{p}$ and the associated errors depend crucially on how one models the noise and how well one can estimate the relevant noise parameters $\vec{q}$.

In some cases one expects and observes the noise to be uncorrelated. For example, the dominant noise source may be shot noise, in which case the noise process is an uncorrelated Poisson process that in the limit of large numbers of counts is well-approximated by an uncorrelated Gaussian process,

$$\mathcal{D}(\vec{\epsilon};\vec{q}) = \mathcal{N}(\epsilon;\sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_i^2}{2\sigma^2}\right), \tag{3.3}$$

in which case there is only one error parameter, σ, specifying the width of the distribution.

If the noise is correlated then it is characterized by a joint probability distribution that is generally a function of all the times of observation. We assume that the function is a multivariate Gaussian function, in which case the noise process is entirely characterized by the covariance matrix

$$\Sigma(t_i, t_j) = \langle \epsilon(t_i)\epsilon(t_j) \rangle. \tag{3.4}$$

Here, the quantity $\langle \epsilon \rangle$ is the mean of the stochastic function ϵ over an infinite number of independent realizations. We further assume that the covariance depends only on the difference in time between two samples, and not on the absolute time of either sample. In this case, the noise source is said to be stationary and is described entirely by its autocovariance $R(\tau)$ (Bracewell 1965):

$$R(\tau) \equiv \langle \epsilon(t)\epsilon(t+\tau) \rangle. \tag{3.5}$$

The parameter estimation problem is often cast in terms of finding the set of parameters $\hat{p}_k$ that maximize a likelihood function. For the case of Gaussian *uncorrelated* noise the likelihood function is

$$\mathcal{L} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{r_i^2}{2\hat{\sigma}^2}\right), \tag{3.6}$$

where r_i is the *residual* defined as $y_i - f(t_i;\vec{p})$, and $\hat{\sigma}$ is an estimate of the single noise parameter σ. Maximizing the likelihood $\mathcal{L}$ is equivalent to minimizing the χ^2 statistic

$$\chi^2 = \sum_{i}^{N} \left(\frac{r_i}{\hat{\sigma}}\right)^2. \tag{3.7}$$

In transit photometry, the estimator $\hat{\sigma}$ of the noise parameter σ is usually *not* taken to be the calculated noise based on expected sources such as shot noise. This is because the actual amplitude of the noise is often greater than the calculated value due to noise sources that are unknown or at least ill-quantified. Instead, $\hat{\sigma}$ is often taken to be the standard deviation of the data obtained when the transit was not occurring, or the value for which $\chi^2 = N_{\rm dof}$ for the best-fitting (minimum-χ^2) model. These estimates work well when the noise process is Gaussian, stationary, and uncorrelated. For the case of correlated noise, Eqn. (3.7) is replaced by (Gould 2003)

$$\chi^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} r_i(\hat{\Sigma}^{-1})_{ij} r_j. \tag{3.8}$$

The case of uncorrelated noise corresponds to $\hat{\Sigma}_{ij} = \hat{\sigma}^2\delta_{ij}$.

It is at this point where various methods for modeling correlated noise begin to diverge. One approach is to estimate $\hat{\Sigma}$ from the sample autocovariance $\hat{R}(\tau)$ of the time series, just as $\hat{\sigma}$ can be estimated from the standard deviation of the residuals in the case of uncorrelated noise. However, the calculation of χ^2 has a worst-case time-complexity of $O(N^2)$ and iterative parameter estimation techniques can be prohibitively slow. One might ameliorate the problem by truncating the covariance matrix at some maximum lag, i.e., by considering the truncated χ^2 statistic

$$\chi^2(L) = \sum_{i=1}^{N} \sum_{\substack{l=-L \\ 1<i+l<N}}^{L} r_i(\hat{\Sigma}^{-1})_{i(i+l)} r_{i+l}, \tag{3.9}$$

but in the presence of long-range correlations one needs to retain many lags to obtain accurate parameter estimates. (In § 3.4.3, we will give an example where 50–75 lags were needed.) Alternatively, one may model the autocorrelation function and therefore the covariance matrix using an autoregressive moving-average (ARMA) model with enough terms to give a good fit to the data (see, e.g., Koen & Lombard 1993). Again, though, in the presence of long-range correlations the model covariance matrix will be non-sparse and computationally burdensome.

Pont et al. (2006) presented a useful simplification in the context of a transit search, when data are obtained on many different nights. In such cases it is reasonable to approximate the covariance matrix as block-diagonal, with different blocks corresponding to different nights. Pont et al. (2006) also gave a useful approximation for the covariance structure within each block, based on the variance in boxcar-averaged versions of the signal. Ultimately their procedure results in an equation resembling Eqn. (3.7) for each block, but where $\hat{\sigma}$ is the quadrature sum of σ_w (the "white noise") and σ_r (the "red noise," estimated from the boxcar-averaged variance). In this paper, all our examples involve a single time series with stationary noise properties, and the net effect of the Pont et al. (2006) method is to enlarge the parameter errors by a factor

$$\beta = \sqrt{1 + \left(\frac{\sigma_r}{\sigma_w}\right)^2} \tag{3.10}$$

relative to the case of purely white noise ($\sigma_r = 0$). We will refer to this method as the "time-averaging" method.

Another approach is to use Eqn. (3.7) without any modifications, but to perform the parameter optimization on a large collection of simulated data sets that are intended to have the same covariance structure as the actual data set. This is the basis of the "residual permutation" method that is also discussed further in § 3.4.4. As mentioned above, this method is a variant of a bootstrap analysis that takes into account time-correlated noise. More details on both the time-averaging and residual-permutation methods are given in § 3.4.4.

Our approach in this paper was motivated by the desire to allow for the possibility of long-range correlations, and yet to avoid the slowness of any method based on Eqn. (3.9) or other time-domain methods. Rather than characterizing the noise in the time domain, we characterize it by its Power Spectral Density (PSD) $\mathcal{S}(f)$ at frequency f, defined as the square of the Fourier transform of $\epsilon(t)$, or equivalently, the Fourier transform of the autocovariance $R(\tau)$. We restrict our discussion to noise sources with a PSD

$$\mathcal{S}(f) = \frac{A}{f^\gamma} \tag{3.11}$$

for some $A > 0$ and spectral index γ. For the special case of uncorrelated noise, $\gamma = 0$ and $\mathcal{S}(f)$ is independent of f. This type of noise has equal power density at all frequencies, which is why it is called "white noise," in an analogy with visible light. As γ is increased, there is an increasing preponderance of low-frequency power over high-frequency power, leading to longer-range correlations in time.

Noise with a power spectrum $1/f^\gamma$ is ubiquitous in nature and in experimental science, including astrophysics (see, e.g., Press 1978). Some examples of $1/f^\gamma$ noise are shown in Fig. 3-1 for a selection of spectral indices. In an extension of the color analogy, $\gamma = 1$ noise is sometimes referred to as "pink noise" and $\gamma = 2$ noise as "red noise." The latter is also known as a Brownian process, although not because of the color brown but instead because of the Scottish botanist Robert Brown. However, as we have already noted, the term "red noise" is often used to refer to any type of low-frequency correlated noise.

Here we do not attempt to explain how $1/f^\gamma$ noise arises in a given situation. Instead we assume that the experimenter has done his or her best to understand and to reduce all sources of noise as far as possible, but despite these efforts there remains a component of $1/f^\gamma$ noise. In transit photometry these correlations often take the

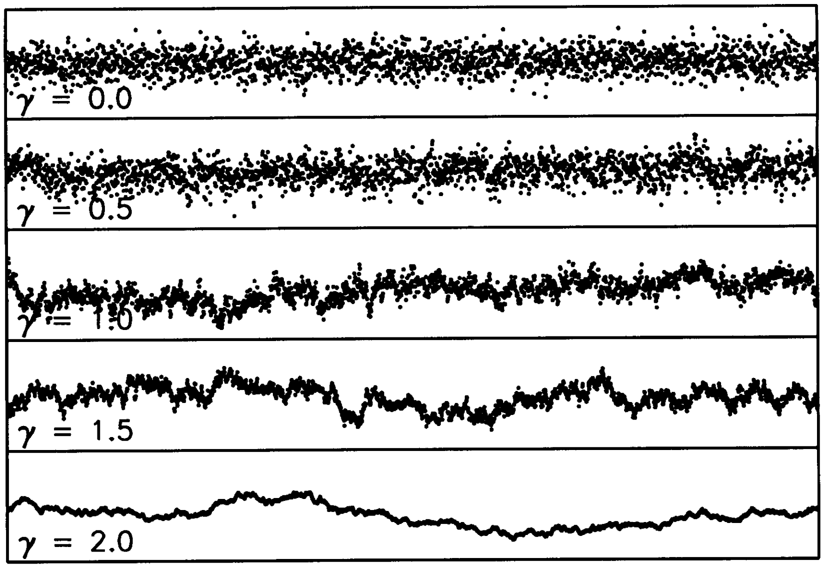

Figure 3-1 Examples of $1/f^{\gamma}$ noise. Uncorrelated (white) noise corresponds to $\gamma = 0$. "Pink" noise corresponds to $\gamma = 1$. "Red" noise or Brownian motion corresponds to $\gamma = 2$. These time series were generated using the wavelet-based method described in § 3.4.

form of "bumps," "wiggles," and "ramps" in a light curve and are often attributed to differential atmospheric extinction, instrumental artifacts such as imperfect flat-fielding, and stellar granulation or other astrophysical effects. The method presented in this paper is essentially a model of the likelihood function that retains the essential information in the covariance matrix without being prohibitively expensive to compute and store. It is based on a wavelet-based description, the subject of the next section.

3.3 Wavelets and $1/f^{\gamma}$ noise

One may regard a time series with N points as a vector in an N-dimensional space that is spanned by N orthonormal unit vectors, one for each time index (the "time basis"). The computational difficulty with correlated noise is that the sample covariance matrix $\hat{\Sigma}$ is not diagonal in the time basis, nor is it necessarily close to being diagonal in realistic cases. This motivates a search for some alternative basis spanning the data space for which the covariance matrix is diagonal or nearly diagonal. For example, if the noise took the form of additive quasiperiodic signals, it would be logical to work in a Fourier basis instead of the time basis.

The mathematical result that underpins our analysis algorithm is that in the presence of $1/f^\gamma$ noise, the covariance matrix is nearly diagonal in a suitable *wavelet basis.* Before giving the details of the algorithm we will briefly review the wavelet transform. Our discussion is drawn primarily from Wornell (1996), Teolis (1998), Daubechies (1988), and Mallat (1999). Practical details and an sample implementation of the wavelet transform are given by Press et al. (2007).

A wavelet is a function that is analogous to the sine and cosine functions of the Fourier transform. Some properties that wavelets share with sines and cosines are that they are localized in frequency space, and they come in families that are related by translations and dilations. Wavelets are *unlike* sine and cosine functions in that wavelets are strongly localized in time. A wavelet basis is derived from a single "mother wavelet" $\psi(t)$, which may have a variety of functional forms and analytic properties. The individual basis functions are formed through translations and dilations of $\psi(t)$. The choice of mother wavelet depends on the specific application. We restrict our focus to dyadic orthogonal wavelet bases with basis functions

$$\psi_n^m(t) = \psi(2^m t - n) \tag{3.12}$$

for all integers m and n, and we further require $\psi(t)$ to have one or more vanishing moments.[2] In this case, the pair of equations analogous to the Fourier series and its inversion is

$$\begin{aligned} \epsilon(t) &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \epsilon_n^m \psi_n^m(t) && (3.13) \\ \epsilon_n^m &= \int_{-\infty}^{\infty} \epsilon(t)\psi_n^m(t)dt && (3.14) \end{aligned}$$

---

[2] In particular it is required that the mother wavelet $\psi(t)$ has zero mean. This is a necessary and sufficient condition to ensure the invertibility of the wavelet transform.

where ϵ_n^m is referred to as the wavelet coefficient of $\epsilon(t)$ at resolution m and translation n.

3.3.1 The wavelet transform as a multiresolution analysis

We will see shortly that some extra terms are required in Eqn. (3.14) for real signals with some minimum and maximum resolution. To explain those terms it is useful to describe the wavelet transform as a multiresolution analysis, in which we consider successively higher-resolution approximations of a signal. An approximation with a resolution of 2^m samples per unit time is a member of a *resolution space* V_m. Following Wornell (1996) we impose the following conditions:

1. if $f(t) \in V_m$ then for some integer n, $f(t - 2^{-m}n) \in V_m$
2. if $f(t) \in V_m$ then $f(2t) \in V_{m+1}$.

The first condition requires that V_m contain all translations (at the resolution scale) of any of its members, and the second condition ensures that the sequence of resolutions is nested: V_m is a subset of the next finer resolution V_{m+1}. In this way, if $\epsilon_m(t) \in V_m$ is an approximation to the signal $\epsilon(t)$, then the next finer approxmation $\epsilon_{m+1}(t) \in V_{m+1}$ contains all the information encoded in $\epsilon_m(t)$ plus some additional *detail* $d_m(t)$ defined as

$$d_m(t) \equiv \epsilon_{m+1}(t) - \epsilon_m(t). \tag{3.15}$$

We may therefore build an approximation at resolution M by starting from some coarser resolution k and adding successive detail functions:

$$\epsilon_M(t) \quad = \quad \epsilon_k(t) + \sum_{m=k}^{M} d_m(t) \tag{3.16}$$

The detail functions $d_m(t)$ belong to a function space $W_m(t)$, the orthogonal complement of the resolution V_m.

With these conditions and definitions, the orthogonal basis functions of W_m are the wavelet functions $\psi_n^m(t)$, obtained by translating and dilating some mother wavelet

$\psi(t)$. The orthogonal basis functions of V_m are denoted $\phi_n^m(t)$, obtained by translating and dilating a so-called "father" wavelet $\phi(t)$. Thus, the mother wavelet spawns the basis of the detail spaces, and the father wavelet spawns the basis of the resolution spaces. They have complementary characteristics, with the mother acting as a high-pass filter and the father acting as a low-pass filter.[3]

In Eqn. (3.16), the approximation $\epsilon_k(t)$ is a member of V_k, which is spanned by the functions $\phi_n^k(t)$, and $d_m(t)$ is a member of W_m, which is spanned by the functions $\psi_n^m(t)$. Thus we may rewrite Eqn. (3.16) as

$$\epsilon_M(t) = \sum_{n=-\infty}^{\infty} \bar{\epsilon}_n^k \phi_n^k(t) + \sum_{m=k}^{M} \sum_{n=-\infty}^{\infty} \epsilon_n^m \psi_n^m(t). \tag{3.17}$$

The *wavelet coefficients* ϵ_n^m and the *scaling coefficients* $\bar{\epsilon}_n^m$ are given by

$$\epsilon_n^m = \int_{-\infty}^{\infty} \epsilon(t)\psi_n^m(t)dt \tag{3.18}$$

$$\bar{\epsilon}_n^m = \int_{-\infty}^{\infty} \epsilon(t)\phi_n^m(t)dt \tag{3.19}$$

Eqn. (3.17) reduces to the wavelet-only equation (3.13) for the case of a continuously sampled signal $\epsilon(t)$, when we have access to all resolutions m from $-\infty$ to ∞.[4]

There are many suitable choices for ϕ and ψ, differing in the tradeoff that must be made between smoothness and localization. The simplest choice is due to Haar (1910):

$$\phi(t) = \begin{cases} 1 & \text{if } 0 < t \leq 1 \\ 0 & \text{otherwise} \end{cases} . \tag{3.20}$$

$$\psi(t) = \begin{cases} 1 & \text{if } -\frac{1}{2} < t \leq 0 \\ -1 & \text{if } 0 < t \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.21}$$

---

[3] More precisely, the wavelet and scaling functions considered here are "quadrature mirror filters" (Mallat 1999).

[4] The signal must also be bounded in order for the approximation to the signal at infinitely coarse resolution to vanish, i.e., $\lim_{k \to -\infty} \epsilon_k(t) = 0$.

The left panel of Fig. 3-2 shows several elements of the approximation and detail bases for a Haar multiresolution analysis. The left panels of Fig. 3-3 illustrate a Haar multiresolution analysis for an arbitrarily chosen signal $\epsilon(t)$, by plotting both the approximations $\epsilon_m(t)$ and details $d_m(t)$ at several resolutions m. The Haar analysis is shown for pedagogic purposes only. In practice we found it advantageous to use the more complicated fourth-order Daubechies wavelet basis, described in the next section, for which the elements and the multiresolution analysis are illustrated in the right panels of Fig. 3-2-3-3.

3.3.2 The Discrete Wavelet Transform

Real signals are limited in resolution, leading to finite M and k in Eqn. (3.17). They are also limited in time, allowing only a finite number of translations N_m at a given resolution m. Starting from Eqn. (3.17), we truncate the sum over n and reindex the resolution sum such that the coarsest resolution is $k = 1$, giving

$$\epsilon_M(t) \quad = \quad \sum_{n=1}^{N_1} \bar{\epsilon}_n^1 \phi_n^1(t) + \sum_{m=2}^{M} \sum_{n=1}^{N_m} \epsilon_n^m \psi_n^m(t) \tag{3.22}$$

where we have taken $t = 0$ to be the start of the signal. Since there is no information on timescales smaller than 2^{-M}, we need only consider $\epsilon_M(t_i)$ at a finite set of times t_i:

$$\epsilon(t_i) \quad = \quad \sum_{n=1}^{N_1} \bar{\epsilon}_n^1 \phi_n^1(t_i) + \sum_{m=2}^{M} \sum_{n=1}^{N_m} \epsilon_n^m \psi_n^m(t_i). \tag{3.23}$$

Eqn. (3.23) is the inverse of the Discrete Wavelet Transform (DWT). Unlike the continuous transform of Eqn. (3.13), the DWT must include the coarsest level approximation (the first term in the preceding equation) in order to preserve all the information in $\epsilon(t_i)$. For the Haar wavelet, the coarsest approximation is the mean value. For data sets with $N = n_0 2^M$ uniformly spaced samples in time, we will have access to a maximal scale M, as in Eqn. (3.23), with $N_m = n_0 2^{m-1}$.

A crucial point is the availability of the Fast Wavelet Transform (FWT) to perform the DWT (Mallat 1989). The FWT is a pyramidal algorithm operating on data sets of size $N = n_0 2^M$ returning $n_0(2^M - 1)$ wavelet coefficients and n_0 scaling coefficients for some $n_0 > 0$, $M > 0$. The FWT is a computationally efficient algorithm that is easily implemented (Press et al. 2007) and has $O(N)$ time-complexity (Teolis 1998).

Daubechies (1988) generalized the Haar wavelet into a larger family of wavelets, categorized according to the number of vanishing moments of the mother wavelet. The Haar wavelet has a single vanishing moment and is the first member of the family. In this work we used the most compact member (in time and frequency), $\psi =_4\mathcal{D}$ and $\phi =_4\mathcal{A}$, which is well suited to the analysis of $1/f^\gamma$ noise for $0 < \gamma < 4$ (Wornell 1996). We plot ${}_4\mathcal{D}^m_n$ and ${}_4\mathcal{A}^m_n$ in the time-domain for several n, m in Fig. 3-2, illustrating the rather unusual functional form of ${}_4\mathcal{D}$. The right panel of Fig. 3-3 demonstrates a multiresolution analysis using this basis. Press et al. (2007) provide code to implement the wavelet transform in this basis.

3.3.3 Wavelet transforms and $1/f^\gamma$ noise

As alluded in § 3.3, the wavelet transform acts as a nearly diagonalizing operator for the covariance matrix in the presence of $1/f^\gamma$ noise. The wavelet coefficients ϵ^m_n of such a noise process are zero-mean, nearly uncorrelated random variables. Specifically, the covariance between scales m, m' and translations n, n' is (Wornell 1996, p. 65)

$$\langle \epsilon^m_n \epsilon^{m'}_{n'} \rangle \approx \left(\sigma_r^2 2^{-\gamma m}\right) \delta_{m,m'} \delta_{n,n'}. \tag{3.24}$$

The wavelet basis is also convenient for the case in which the noise is modeled as the sum of an uncorrelated component and a correlated component,

$$\epsilon(t) = \epsilon_0(t) + \epsilon_\gamma(t), \tag{3.25}$$

where $\epsilon_0(t)$ is a Gaussian white noise process ($\gamma = 0$) with a single noise parameter σ_w, and $\epsilon_\gamma(t)$ has $\mathcal{S}(f) = A/f^\gamma$. In the context of transit photometry, white noise might arise from photon-counting statistics (and in cases where the detector is wellcalibrated, σ_w is a known constant), while the $\gamma \neq 0$ term represents the "rumble" on many time scales due to instrumental, atmospheric, or astrophysical sources. For the noise process of Eqn. (3.25) the covariance between wavelet coefficients is

$$\langle \epsilon_n^m \epsilon_{n'}^{m'} \rangle \approx \left(\sigma_r^2 2^{-\gamma m} + \sigma_w^2\right) \delta_{m,m'} \delta_{n,n'}. \tag{3.26}$$

and the covariance betwen the scaling coefficients $\bar{\epsilon}_n^m$ is

$$\langle \bar{\epsilon}_n^m \bar{\epsilon}_n^m \rangle \approx \sigma_r^2 2^{-\gamma m} g(\gamma) + \sigma_w^2 \tag{3.27}$$

where $g(\gamma)$ is a constant of order unity; for the purposes of this work $g(1) = (2 \ln 2)^{-1} \approx 0.72$ (Fadili & Bullmore 2002). Eqns. (3.26) and (3.27) are the key mathematical results that form the foundation of our algorithm. For proofs and further details, see Wornell (1996).

It should be noted that the correlations between the wavelet and scaling coefficients are small but not exactly zero. The decay rate of the correlations with the resolution index depends on the choice of wavelet basis and on the spectral index γ. By picking a wavelet basis with a higher number of vanishing moments, we hasten the decay of correlations. This is why we chose the Daubechies 4th-order basis instead of the Haar basis. In the numerical experiments decribed in § 4, we found the covariances to be negligible for the purposes of parameter estimation. In addition, the compactness of the Daubechies 4th-order basis reduces artifacts arising from the assumption of a periodic signal that is implicit in the FWT.

3.3.4 The whitening filter

Given an observation of noise $\epsilon(t)$ that is modeled as in Eqn. (3.25), we may estimate the $\gamma \neq 0$ component by rescaling the wavelet and scaling coefficients and filtering out the white component:

$$\epsilon_\gamma(t) = \sum_{n=1}^{N_1} \left(\frac{\sigma_r^2 2^{-\gamma} g(\gamma)}{\sigma_r^2 2^{-\gamma} g(\gamma) + \sigma_w^2} \right) \bar{\epsilon}_n^1 \phi_n^1(t) + \tag{3.28}$$

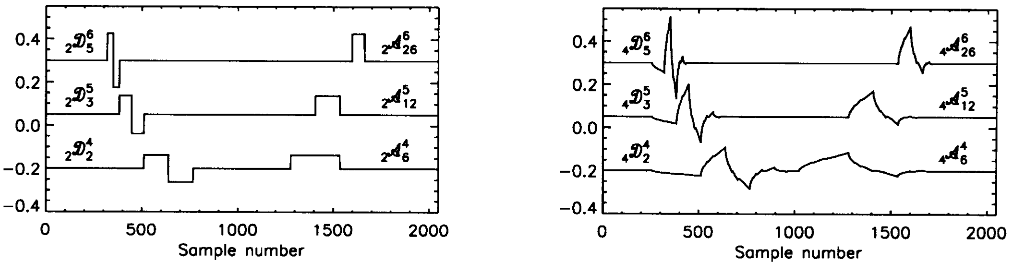

Figure 3-2 Examples of discrete wavelet and scaling functions, for $N = 2048$. *Left.*—Haar wavelets and the corresponding father wavelets, also known as 2nd-order Daubechies orthonormal wavelets or $_2\mathcal{D}^m_n$ and $_2\mathcal{A}^m_n$. *Right.*—4th-order Daubechies orthonormal wavelets, or $_4\mathcal{D}^m_n$ and $_4\mathcal{A}^m_n$.

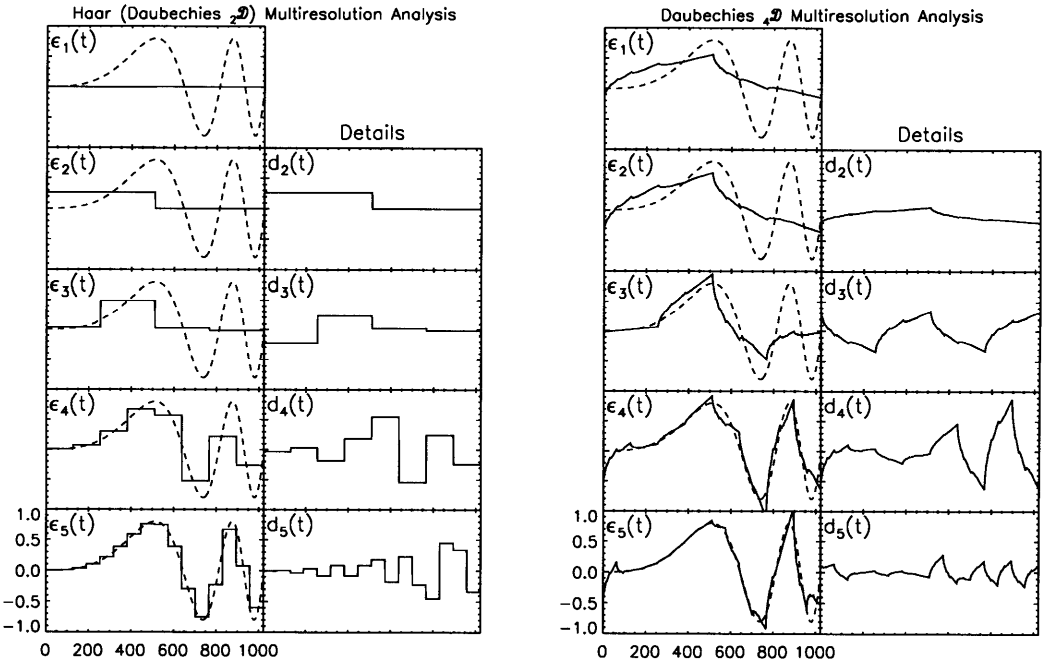

Figure 3-3 Illustration of a multiresolution analysis, for the function $\epsilon(t) = \sin[4\pi(t/1024)^3]$ (dashed line). Plotted are the approximations $\epsilon_m(t)$ to the function at successive resolutions, along with the detail functions $d_m(t)$. *Left.*—Using the Haar wavelet basis. *Right.*—Using the 4th-order Daubechies wavelet basis.

$$\sum_{m=2}^{M}\sum_{n=1}^{N_m}\left(\frac{\sigma_r^2 2^{-\gamma m}}{\sigma_r^2 2^{-\gamma m}+\sigma_w^2}\right)\epsilon_n^m \psi_n^m(t). \tag{3.29}$$

We may then proceed to subtract the estimate of the correlated component from the observed noise, $\epsilon_0(t) = \epsilon(t) - \epsilon_\gamma(t)$ (Wornell 1996, p. 76). In this way the FWT can be used to "whiten" the noise.

3.3.5 The wavelet-based likelihood

Armed with the preceding theory, we rewrite the likelihood function of Eqn. (3.6) in the wavelet domain. First we transform the residuals $r_i \equiv y_i - f(t_i;\vec{p})$, giving

$$\begin{aligned} r_n^m &= y_n^m - f_n^m(\vec{p}) = \epsilon_{\gamma,n}^m + \epsilon_{0,n}^m & (3.30)\\ \bar{r}_n^1 &= \bar{y}_n^1 - \bar{f}_n^1(\vec{p}) = \bar{\epsilon}_{\gamma,n}^1 + \bar{\epsilon}_{0,n}^1 & (3.31)\end{aligned}$$

where y_n^m and $f_n^m(\vec{p})$ are the discrete wavelet coefficients of the data and the model. Likewise, $\bar{y}_n^1$ and $\bar{f}_n^1(\vec{p})$ are the n_0 scaling coefficients of the data and the model. Given the diagonal covariance matrix shown in Eqns. (3.26) and (3.27), the likelihood $\mathcal{L}$ is a product of Gaussian functions at each scale m and translation n:

$$\begin{aligned}\mathcal{L} &= \left\{\prod_{m=2}^{M}\prod_{n=1}^{n_0 2^{m-1}}\frac{1}{\sqrt{2\pi\sigma_W^2}}\exp\left[-\frac{(r_n^m)^2}{2\sigma_W^2}\right]\right\}\\ &\times\left\{\prod_{n=1}^{n_0}\frac{1}{\sqrt{2\pi\sigma_S^2}}\exp\left[-\frac{(\bar{r}_n^1)^2}{2\sigma_S^2}\right]\right\} \end{aligned} \tag{3.32}$$

where

$$\begin{aligned}\sigma_W^2 &= \sigma_r^2 2^{-\gamma m} + \sigma_w^2 & (3.33)\\ \sigma_S^2 &= \sigma_r^2 2^{-\gamma} g(\gamma) + \sigma_w^2 & (3.34)\end{aligned}$$

are the variances of the wavelet and scaling coefficients respectively. For a data set with N points, calculating the likelihood function of Eqn. (3.32) requires multiplying N Gaussian functions. The additional step of computing the FWT of the residuals prior to computing $\mathcal{L}$ adds $O(N)$ operations. Thus, the entire calculation has a time-complexity $O(N)$.

For this calculation we must have estimators of the three noise parameters γ, σ_r and σ_w. These may be estimated separately from the model parameters $\vec{p}$, or simultaneously with the model parameters. For example, in transit photometry, the data obtained outside of the transit may be used to estimate the noise parameters, which are then used in Eqn. (3.32) to estimate the model parameters. Or, in a single step we could maximize Eqn. (3.32) with respect to all of γ, σ_r, σ_w and $\vec{p}$. Fitting for both noise and transit parameters simultaneously is potentially problematic, because some of the correlated noise may be "absorbed" into the choices of the transit parameters, i.e., the errors in the noise parameters and transit parameters are themselves correlated. This may cause the noise level and the parameter uncertainties to be underestimated. Unfortunately, there are many instances when one does not have enough out-of-transit data for the strict separation of transit and noise parameters to be feasible.

In practice the optimization can be accomplished with an iterative routine [such as AMOEBA, Powell's method, or a conjugate-gradient method; see Press et al. (2007)]. Confidence intervals can then be defined by the contours of constant likelihood. Alternatively one can use a Monte Carlo Markov Chain [MCMC; see, e.g., Gregory (2005)], in which case the jump-transition likelihood would be given by Eqn. (3.32). The advantages of the MCMC method have led to its adoption by many investigators (see, e.g., Holman et al. 2006, Burke et al. 2007, Collier Cameron et al. 2007). For that method, computational speed is often a limiting factor, as a typical MCMC analysis involves several million calculations of the likelihood function.

3.3.6 Some practical considerations

Some aspects of real data do not fit perfectly into the requirements of the DWT. The time sampling of the data should be approximately uniform, so that the resolution scales of the multiresolution analysis accurately reflect physical timescales. This is usually the case for time-series photometric data. Gaps in a time series can be fixed by applying the DWT to each uninterrupted data segment, or by filling in the missing elements of the residual series with zeros.

The FWT expects the number of data points to be an integral multiple of some integral power of two. When this is not the case, the time series may be truncated to the nearest such boundary; or it may be extended using a periodic boundary condition, mirror reflection, or zero-padding. In the numerical experiments described below, we found that zero-padding has negligible effects on the calculation of likelihood ratios and parameter estimation.

The FWT generally assumes a periodic boundary condition for simplicity of computation. A side effect of this simplication is that information at the beginning and end of a time series are artificially associated in the wavelet transform. This is one reason why we chose the 4th-order Daubechies-class wavelet basis, which is well localized in time, and does not significantly couple the beginning and the end of the time series except on the coarsest scales.

3.4 Numerical experiments with transit light curves

We performed many numerical experiments to illustrate and test the wavelet method. These experiments involved estimating the parameters of simulated transit light curves. We also compared the wavelet analysis to a "white" analysis, by which we mean a method that assumes the errors to be uncorrelated, and to two other analysis methods drawn from the literature. Because we used simulated transit light curves with known noise and transit parameters, the "truth" was known precisely, allowing both the absolute and relative merits of the methods to be evaluated.

3.4.1 Estimating the midtransit time: Known noise parameters

In this section we consider the case in which the noise parameters γ, σ_r, and σ_w are known with negligible error. We have in mind a situation in which a long series of out-of-transit data are available, with stationary noise properties.

We generated transit light curves with known transit parameters $\vec{p}$, contaminated by an additive combination of a white and a correlated $(1/f^{\gamma})$ noise source. Then we used an MCMC method to estimate the transit parameters and their 68.3% confidence limits. (The technique for generating noise and the MCMC method are described in detail below.) For each realization of a simulated light curve, we estimated transit parameters using the likelihood defined either by Eqn. (3.6) for the white analysis, or Eqn. (3.32) for the wavelet analysis.

For a given parameter p_k, the estimator $\hat{p}_k$ was taken to be the median of the values in the Markov chain and $\hat{\sigma}_{p_k}$ was taken to be the standard deviation of those values. To assess the results, we considered the "number-of-sigma" statistic

$$\mathcal{N} \equiv \left(\hat{p}_k - p_k\right)/\hat{\sigma}_{p_k}. \tag{3.35}$$

In words, $\mathcal{N}$ is the number of standard deviations separating the parameter estimate $\hat{p}_k$ from the true value p_k. If the error in p_k is Gaussian, then a perfect analysis method should yield results for $\mathcal{N}$ with an expectation value of 0 and variance of 1. If we find that the variance of $\mathcal{N}$ is greater than one, then we have underestimated the error in $\hat{p}_k$ and we may attribute too much significance to the result. On the other hand, if the variance of $\mathcal{N}$ is smaller than one, then we have overestimated σ_{p_k} and we may miss a significant discovery. If we find that the mean of $\mathcal{N}$ is nonzero then the method is biased.

For now, we consider only the single parameter t_c, the time of midtransit. The t_c parameter is convenient for this analysis as it is nearly decoupled from the other transit parameters (Carter et al. 2008). Furthermore, as mentioned in the introduction, the measurement of the midtransit time cannot be improved by observing other transit events, and variations in the transit interval are possible signs of additional gravitating bodies in a planetary system.

The noise was synthesized as follows. First, we generated a sequence of $N = 1024$ independent random variables obeying the variance conditions from Eqns. (3.26) and

(3.27) for 1023 wavelet coefficients over 9 scales and a single scaling coefficient at the coarsest resolution scale. We then performed the inverse FWT of this sequence to generate our noise signal. In this way, we could select exact values for γ, σ_r, and σ_w. We also needed to find the single parameter σ for the white-noise analysis; it is not simply related to the parameters γ, σ_r, and σ_w. In practice, we found σ by calculating the median sample variance among 10^4 unique realizations of a noise source with fixed parameters γ, σ_r, and σ_w.

For the transit model, we used the analytic formulas of Mandel & Agol (2002), with a planet-to-star ratio of $R_p/R_\star = 0.15$, a normalized orbital distance of $a/R_\star = 10$, and an orbital inclination of $i = 90°$, as appropriate for a gas giant planet in a close-in orbit around a K star. These correspond to a fractional loss of light $\delta = 0.0225$, duration $T = 1.68$ hr, and partial duration $\tau = 0.152$ hr. We did not include the effect of limb darkening, as it would increase the computation time and has little influence on the determination of t_c (Carter et al. 2009). Each simulated light curve spanned 3 hr centered on the midtransit time, with a time sampling of 11 s, giving 1024 uniformly spaced samples. A noise-free light curve is shown in Fig. 3-4.

For the noise model, we chose $\sigma_w = 1.35 \times 10^{-3}$ and $\gamma = 1$, and tried different choices for σ_r. We denote by α the ratio of the rms values of the correlated noise component and the white noise component.[5] The example in Fig. 3-4 has $\alpha = 1/3$. As α is increased from zero, the correlated component becomes more important, as is evident in the simulated data plotted in Fig. 3-5. Our choice of σ_w corresponds to a precision of 5.8×10^{-4} per minute-equivalent sample, and was inspired by the recent work by Johnson et al. (2009) and Winn et al. (2009), which achieved precisions of 5.4×10^{-4} and 4.0×10^{-4} per minute-equivalent sample, respectively. Based on our survey of the literature and our experience with the Transit Light Curve project (Holman et al. 2006, Winn et al. 2007), we submit that all of the examples shown in Fig. 3-5 are "realistic" in the sense that the bumps, wiggles, and ramps resemble features in actual light curves, depending on the instrument, observing site, weather

---

[5] We note that although σ_w is the rms of the white noise component, σ_r is generally not the rms of the correlated component. The notation is unfortunate, but follows that of Wornell (1996).

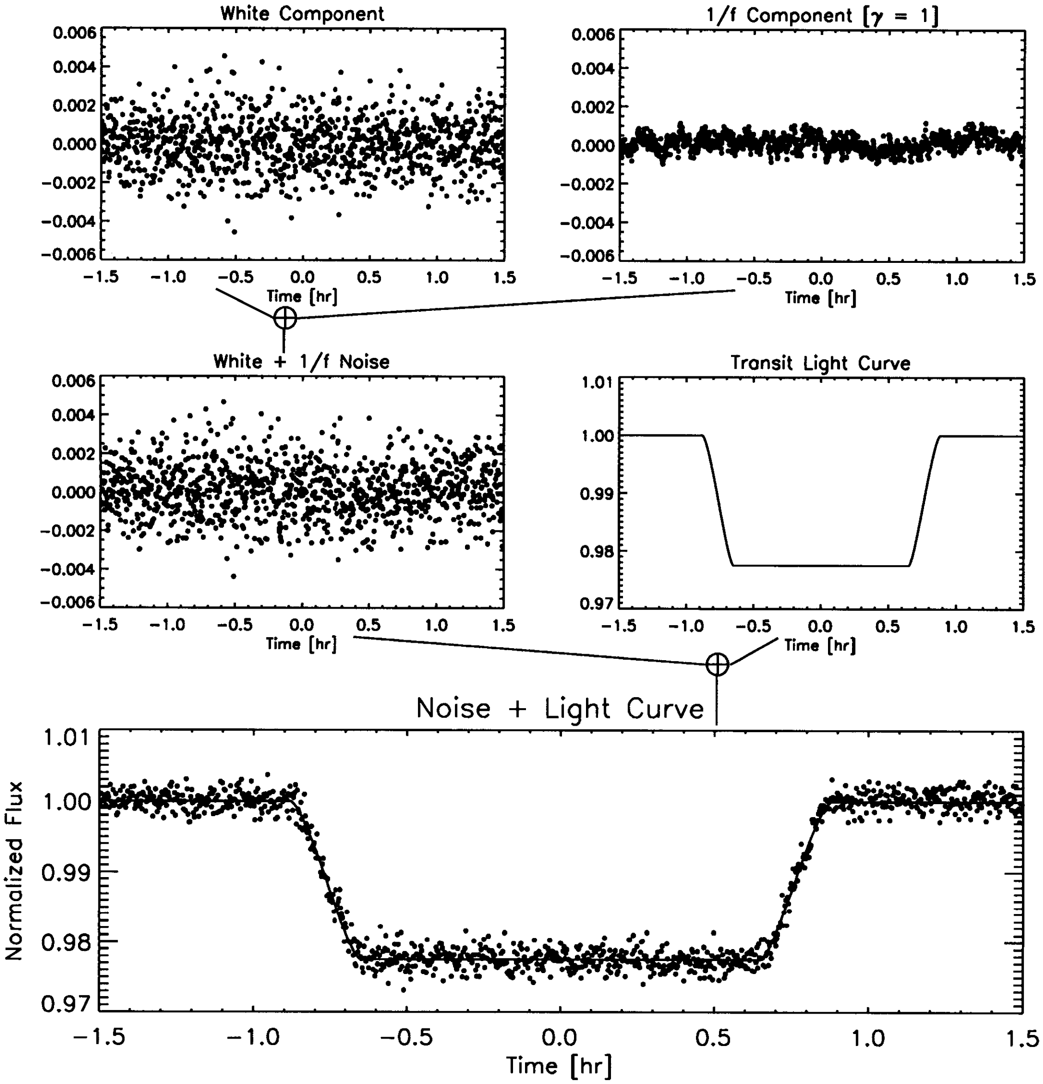

Figure 3-4 Constructing a simulated transit light curve with correlated noise. The total noise is the sum of uncorrelated Gaussian noise with standard deviation σ_w (upper left panel) and correlated noise with a power spectral density $S(f) \propto 1/f$ and an rms equal to $\sigma_w/3$ (upper right panel). The total noise (middle left panel) is added to an idealized transit model (middle right panel) to produce the simulated data (bottom panel).

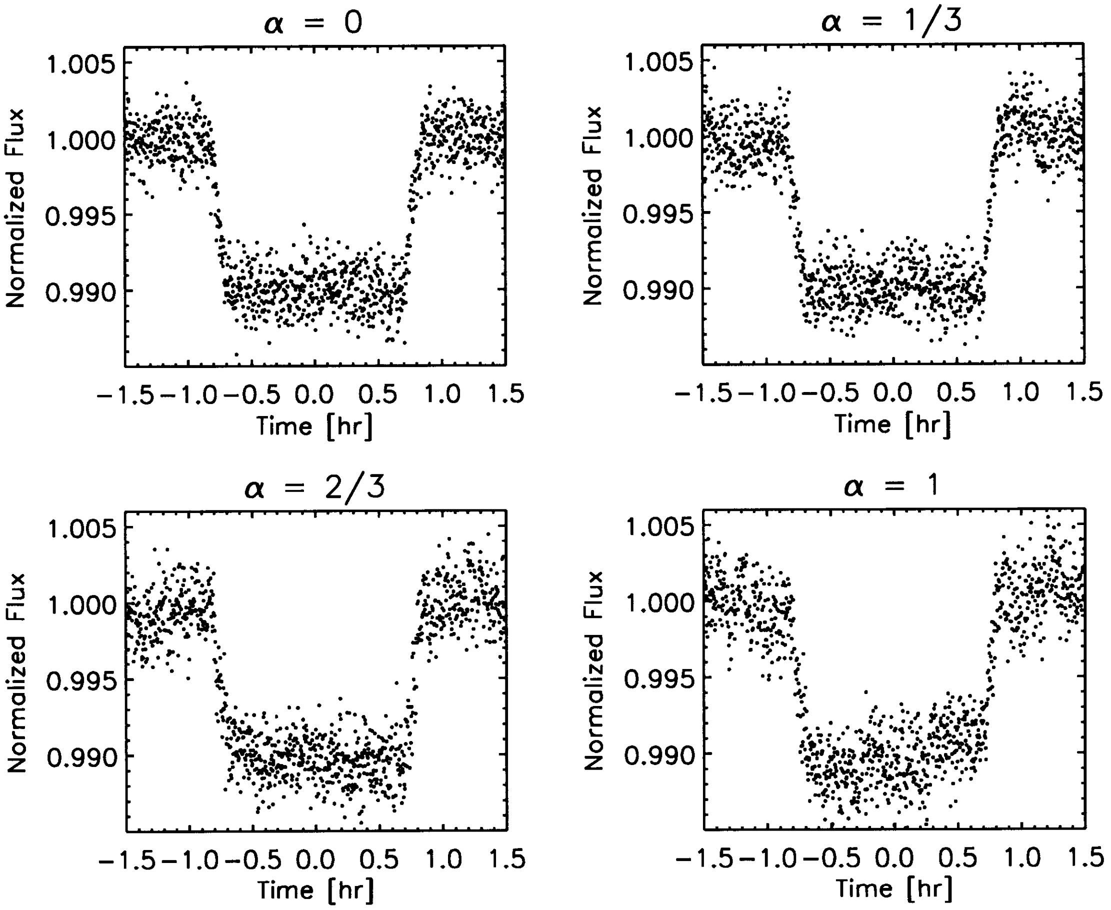

Figure 3-5 Examples of simulated transit light curves with different ratios $\alpha = \mathrm{rms}_r/\mathrm{rms}_w$ between the rms values of the correlated noise component and white noise component.

conditions, and target star.

For a given choice of α, we made 10,000 realizations of the simulated transit light curve with $1/f$ noise. We then constructed two Monte Carlo Markov Chains for t_c starting at the true value of $t_c = 0$. One chain was for the white analysis, with a jump-transition likelihood given by Eqn. (3.6). The other chain was for the wavelet analysis, using Eqn. (3.32) instead. Both chains used the Metropolis-Hastings jump condition, and employed perturbation sizes such that ≈40% of jumps were accepted. Initial numerical experiments showed that the autocorrelation of a given Markov chain for t_c is sharply peaked at zero lag, with the autocorrelation dropping below 0.2 at lag-one. This ensured good convergence with chain lengths of 500 (Tegmark et al. 2004). Chain histograms were also inspected visually to verify that the distribution was smooth. We recorded the median $\hat{t}_c$ and standard deviation $\hat{\sigma}_{t_c}$ for each chain and constructed the statistic $\mathcal{N}$ for each separate analysis (white or wavelet). Finally, we found the median and standard deviation of $\mathcal{N}$ over all 10,000 noise realizations.

Fig. 3-6 shows the resulting distributions of $\mathcal{N}$, for the particular case $\alpha = 1/3$. Table 3.1 gives a collection of results for the choices $\alpha = 0$, $1/3$, $2/3$, and 1. The mean of $\mathcal{N}$ is zero for both the white and wavelet analyses: neither method is biased. This is expected, because all noise sources were described by zero-mean Gaussian distributions. However, the widths of the distributions of $\mathcal{N}$ show that the white analysis underestimates the error in t_c. For a transit light curve constructed with equal parts white and $1/f$ noise ($\alpha = 1$), the white analysis gave an estimate of t_c that differs from the true value by more than 1 σ nearly 80% of the time. The factor by which the white analysis underestimates the error in t_c appears to increase linearly with α. In contrast, for all values of α, the wavelet analysis maintains a unit variance in $\mathcal{N}$, as desired.

The success of the wavelet method is partially attributed to the larger (and more appropriate) error intervals that it returns for $\hat{t}_c$. It is also partly attributable to an improvement in the accuracy of $\hat{t}_c$ itself: the wavelet method tends to produce $\hat{t}_c$ values that are closer to the true t_c. This is shown in the final column in Table (3.1), where we report the percentage of cases in which the analysis method (white or wavelet) produces an estimate of t_c that is closer to the truth. For $\alpha = 1$ the wavelet analysis gives more accurate results 66% of the time.

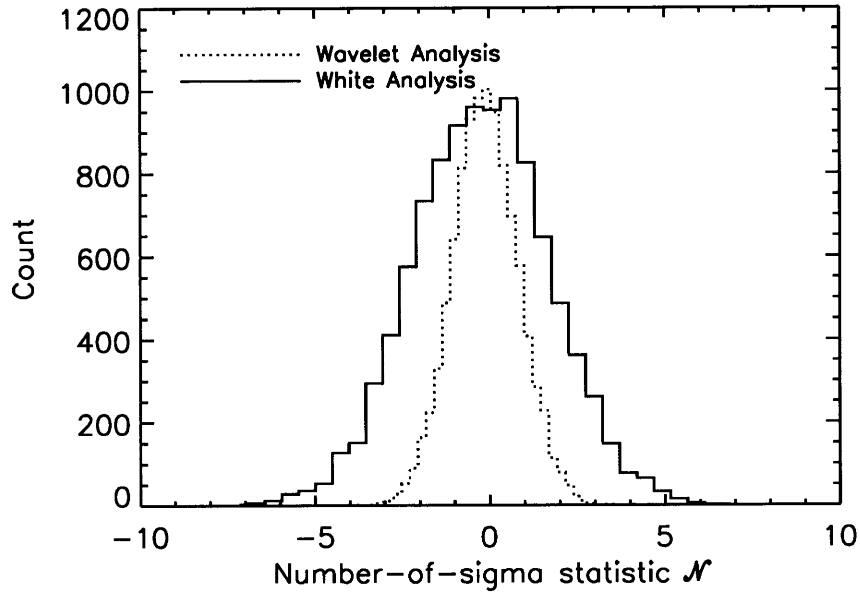

Figure 3-6 Histograms of the number-of-sigma statistic $\mathcal{N}$ for the midtransit time t_c. Each distribution shows the probability of estimating a value for t_c that differs by $\mathcal{N}\sigma$ from the true value. The simulated data were created by adding an idealized transit model to a noise source that is the sum of uncorrelated noise and $1/f$ noise with equal variances ($\alpha = 1$; see the text).

3.4.2 Estimating the midtransit time: Unknown noise parameters

In this section we consider the case in which the noise parameters are not known in advance. Instead the noise parameters must be estimated based on the data. We did this by including the noise parameters as adjustable parameters in the Markov chains. In principle this could be done for all three noise parameters γ, σ_r, and σ_w, but for most of the experiments presented here we restricted the problem to the case $\gamma = 1$. This may be a reasonable simplification, given the preponderance of natural noise sources with $\gamma = 1$ (Press 1978). Some experiments involving noise with $\gamma \neq 1$ are described at the end of this section.

Table 3.1. Estimates of mid-transit time, t_c, from data with known noise properties

Method	α	$\langle \hat{\sigma}_{t_c} \rangle$ [sec]	$\langle \mathcal{N} \rangle$	$\sigma_{\mathcal{N}}$	*prob*($\mathcal{N} > 1$)	*prob*(best)[a]
White	0	4.1	+0.004	0.95	29%	50%
	1/3	4.3	−0.005	1.93	61%	39%
	2/3	5.0	+0.005	3.04	75%	35%
	1	5.9	−0.036	3.82	79%	34%
Wavelet	0	4.0	+0.005	0.95	29%	50%
	1/3	7.2	−0.004	0.93	28%	61%
	2/3	11.5	−0.004	0.94	28%	65%
	1	16.0	−0.001	0.95	29%	66%

[a] The probability that the analysis method (white or wavelet) returns an estimate of t_c that is closer to the true value than the other method.

We also synthesized the noise with a non-wavelet technique, to avoid "stacking the deck" in favor of the wavelet method. We generated the noise in the frequency domain, as follows. We specified the amplitudes of the Fourier coefficients using the assumed functional form of the power spectral density [$\mathcal{S}(f) \propto 1/f$], and drew the phases from a uniform distribution between $-\pi$ and π. The correlated noise in the time domain was found by performing an inverse Fast Fourier Transform. We rescaled the noise such that the rms was α times the specified σ_w. The normally-distributed white noise was then added to the correlated noise to create the total noise. This in turn was added to the idealized transit model.

For each choice of α, we made 10,000 simulated transit light curves and analyzed them with the MCMC method described previously. For the white analysis, the mid-transit time t_c and the single noise parameter σ were estimated using the likelihood defined via Eqn. (3.6). For the wavelet analysis we estimated t_c and the two noise parameters σ_r and σ_w using the likelihood defined in Eqn. (3.32).

Table 3.3 gives the resulting statistics from this experiment, in the same form

Table 3.2. Effect of time sampling on the white analysis

N[a]	Cadence [sec]	$\sigma_{\mathcal{N}}$
256	42.2	1.72
512	21.1	2.04
1024	10.5	2.69
2048	5.27	3.49
4096	2.63	4.39

[a] The number of samples in a 3 hr interval.

as were given in Table 3.1 for the case of known noise parameters. (This table also includes some results from § 3.4.4, which examines two other methods for coping with correlated noise.) Again we find that the wavelet method produces a distribution of $\mathcal{N}$ with unit variance, regardless of α; and again, we find that the white analysis underestimates the error in t_c. In this case the degree of error underestimation is less severe, a consequence of the additional freedom in the noise model to estimate σ from the data. The wavelet method also gives more accurate estimates of t_c than the white method, although the contrast between the two methods is smaller than it was with for the case of known noise parameters.

Our numerical results must be understood to be illustrative, and not universal. They are specific to our choices for the noise parameters and transit parameters. Via further numerical experiments, we found that the width of $\mathcal{N}$ in the white analysis is independent of σ_w, but it does depend on the time sampling. In particular, the width grows larger as the time sampling becomes finer (see Table 3.2). This can be understood as a consequence of the long-range correlations. The white analysis assumes that the increased number of data points will lead to enhanced precision, whereas in reality, the correlations negate the benefit of finer time sampling.

Table 3.4 gives the results of additional experiments with $\gamma \neq 1$. In those cases we created simulated noise with $\gamma \neq 1$ but in the course of the analysis we assumed $\gamma = 1$.

Table 3.3. Estimates of t_c from data with unknown noise properties

Method	α	$\langle\hat{\sigma}_{t_c}\rangle$ [sec]	$\langle\mathcal{N}\rangle$	$\sigma_{\mathcal{N}}$	$prob(\mathcal{N} > 1)$	$prob$(better)[a]
White	0	4.0	−0.011	0.97	31%	–
	1/3	4.2	+0.010	1.70	57%	–
	2/3	4.9	+0.012	2.69	73%	–
	1	5.8	+0.023	3.28	78%	–
Wavelet	0	4.5	−0.009	0.90	26%	50%
	1/3	6.9	−0.003	1.03	33%	56%
	2/3	11.2	−0.005	1.07	35%	57%
	1	15.7	−0.007	1.09	36%	57%
Time-averaging	0	4.4	−0.006	0.88	26%	50%
	1/3	6.8	+0.009	1.15	36%	50%
	2/3	11.6	−0.012	1.24	40%	50%
	1	17.6	+0.007	1.21	38%	50%
Residual-permutation	0	3.5	−0.012	1.16	37%	50%
	1/3	6.6	+0.013	1.24	37%	50%
	2/3	11.8	−0.014	1.28	38%	49%
	1	17.3	+0.008	1.30	38%	48%

[a] The probability that the analysis method returns an estimate of t_c that is closer to the true value than the white analysis.

The correlated noise fraction was set to $\alpha = 1/2$ for these tests. The results show that even when γ is falsely assumed to be unity, the wavelet analysis still produces better estimates of t_c and more reliable error bars than the white analysis.

3.4.3 Runtime analysis of the time-domain method

Having established the superiority of the wavelet method over the white method, we wish to show that the wavelet method is also preferable to the more straightforward approach of computing the likelihood function in the time domain with a non-diagonal covariance matrix. The likelihood in this case is given by Eqn. (3.8).

Table 3.4. Estimates of t_c from data with unknown noise properties

Method	γ[a]	$\langle\hat{\sigma}_{t_c}\rangle$ [sec]	$\langle\mathcal{N}\rangle$	$\sigma_{\mathcal{N}}$	$prob(\mathcal{N} > 1)$	$prob$(best)[b]
White	0.5	4.5	−0.025	1.34	47%	50%
	1.5	4.6	+0.020	3.10	77%	32%
Wavelet	0.5	6.7	−0.021	0.97	30%	50%
	1.5	6.9	+0.002	1.17	39%	68%

[a] The spectral exponent of the Power Spectral Density, $S(f) \propto 1/f^{\gamma}$.

[b] The probability that the analysis method (white or wavelet) returns an estimate of t_c that is closer to the true value than the other method.

The time-domain calculation and the use of the covariance matrix raised two questions. First, how well can we estimate the autocovariance $R(\tau)$ from a single time series? Second, how much content of the resulting covariance matrix needs to be retained in the likelihood calculation for reliable parameter estimation? The answer to the first question depends on whether we wish to utilize the sample autocorrelation as the estimator of $R(\tau)$ or instead use a parametric model (such as an ARMA model) for the autocorrelation. In either case, our ability to estimate the autocorrelation improves with number of data samples contributing to its calculation. The second question is important because retaining the full covariance matrix would cause the computation time to scale as $O(N^2)$ and in many cases the analysis would be prohibitively slow. The second question may be reframed as: what is the minimum number of lags L that needs to be considered in computing the truncated χ^2 of Eqn. (3.9), in order to give unit variance in the number-of-sigma statistic for each model parameter? The time-complexity of the truncated likelihood calculation is $O(NL)$. If $L \leq 5$ then the time-domain method and the wavelet method may have comparable computational time-complexity, while for larger L the wavelet method would offer significant advantage.

We addressed these questions by repeating the experiments of the previous sections using a likelihood function based on the truncated χ^2 statistic. We assumed that the parameters of the noise model were known, as in § 3.4.1. The noise was synthesized in the wavelet domain, with $\gamma = 1$, $\sigma_w = 0.00135$, and α set equal to 1/3 or 2/3. The parameters of the transit model and the time series were the same as in § 3.4.1. We calculated the "exact" autocovariance function $R(l)$ at integer lag l for a given α by averaging sample autocovariances over 50,000 noise realizations. Fig. 3-7 plots the autocorrelation $[R(l)/R(0)]$ as a function of lag for $\alpha = 1/3,\ 2/3$. We constructed the stationary covariance $\Sigma_{ij} = R(|i-j|)$ and computed its inverse $(\Sigma^{-1})_{ij}$ for use in Eqn. (3.9).

Then we used the MCMC method to find estimates and errors for the time of midtransit, and calculated the number-of-sigma statistic $\mathcal{N}$ as defined in Eqn. (3.35). In particular, for each simulated transit light curve, we created a Markov chain of $1,000$ links for t_c, using $\chi^2(L)$ in the jump-transition likelihood. We estimated t_c and σ_{t_c}, and calculated $\mathcal{N}$. We did this for $5,000$ realizations and determined $\sigma_{\mathcal{N}}$, the variance in $\mathcal{N}$, across this sample. We repeated this process for different choices of the maximum lag L. Fig. (3-8) shows the dependence of $\sigma_{\mathcal{N}}$ upon the maximum lag L.

The time-domain method works fine, in the sense that when enough non-diagonal elements in the covariance matrix are retained, the parameter estimation is successful. We find that $\sigma_{\mathcal{N}}$ approaches unity as $L^{-\beta}$ with $\beta = 0.15,\ 0.25$ for $\alpha = 1/3,\ 2/3$, respectively. However, to match the reliability of the wavelet method, a large number of lags must be retained. To reach $\sigma_{\mathcal{N}} = 1.05$, we need $L \approx 50$ for $\alpha = 1/3$ or $L \approx 75$ for $\alpha = 2/3$. In our implementation, the calculation based on the truncated covariance matrix [Eqn. (3.9)] took 30–40 times longer than the calculation based on the wavelet likelihood [Eqn. (3.32)].

This order-of-magnitude penalty in runtime is bad enough, but the real situation may be even worse, because one usually has access to a single noisy estimate of the autocovariance matrix. Or, if one is using an ARMA model, the estimated parameters of the model might be subject to considerable uncertainty as compared to the "exact" autocovariance employed in our numerical experiments. If it is desired to determine the noise parameters simultaneously with the other model parameters, then there is a further penalty associated with inverting the covariance matrix at each step of the calculation for use in Eqn. (3.9), although it may be possible to circumvent that particular problem by modeling the inverse-covariance matrix directly.

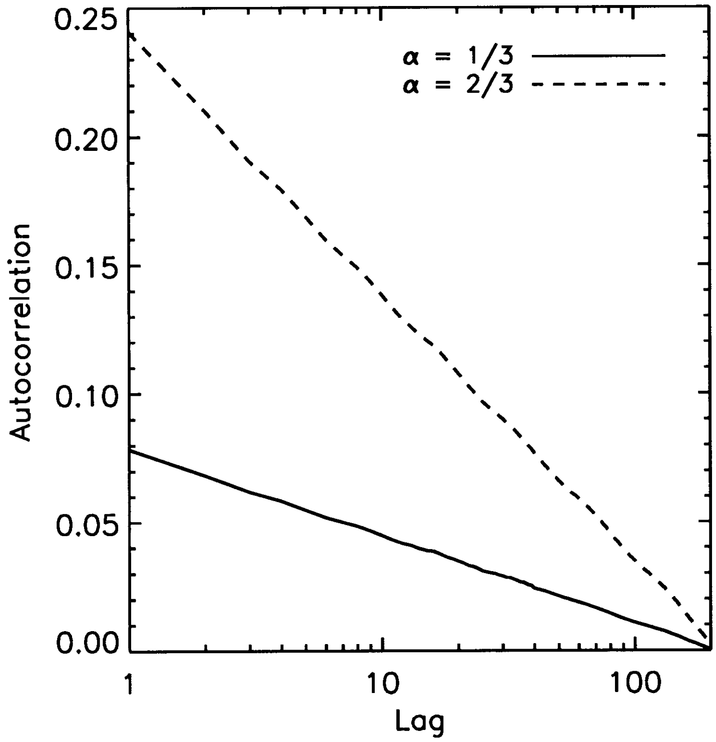

Figure 3-7 Autocorrelation functions of correlated noise. The noise was computed as the sum of white noise with $\sigma_w = 0.00135$ and $1/f$ noise with an rms equal to $\alpha\sigma_w$, for $\alpha = 1/3$ or $2/3$.

3.4.4 Comparison with other methods

In this section we compare the results of the wavelet method to two methods for coping with correlated noise that are drawn from the recent literature on transit photometry. The first of these two methods is the "time averaging" method that was propounded by Pont et al. (2006) and used in various forms by Bakos et al. (2006), Gillon et al. (2006), Winn et al. (2007, 2008, 2009), Gibson et al. (2008), and others.

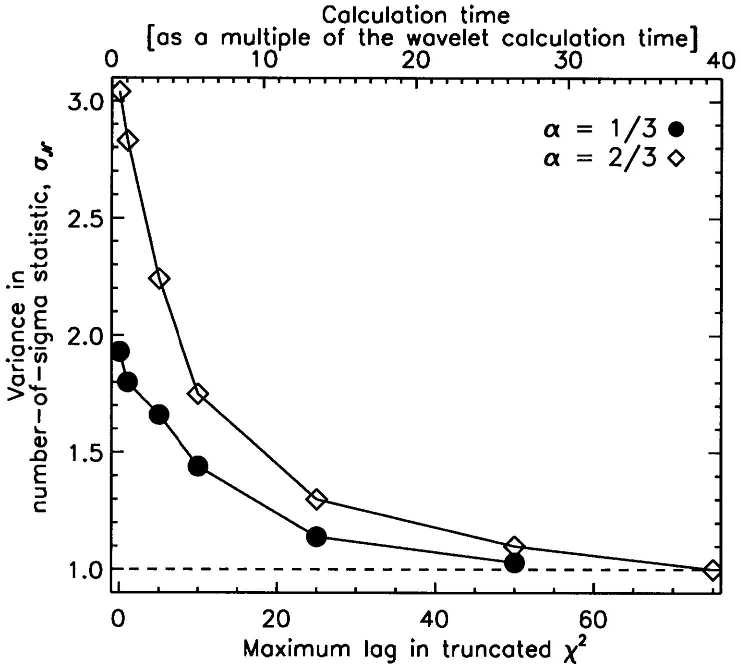

Figure 3-8 Accuracy of the truncated time-domain likelihood in estimating midtransit times. Plotted is the variance in the number-of-sigma statistic $\sigma_{\mathcal{N}}$ for the midtransit time t_c, as a function of the maximum lag in the truncated series. The estimates of t_c were found using the truncated likelihood given in Eqn. (3.9).

In one implementation, the basic idea is to calculate the sample variance of unbinned residuals, $\hat{\sigma}_1^2$, and also the sample variance of the time-averaged residuals, $\hat{\sigma}_n^2$, where every n points have been averaged (creating m time bins). In the absence of correlated noise, we expect

$$\hat{\sigma}_n^2 = \frac{\hat{\sigma}_1^2}{n}\left(\frac{m}{m-1}\right). \tag{3.36}$$

In the presence of correlated noise, $\hat{\sigma}_n^2$ differs from this expectation by a factor $\hat{\beta}_n^2$. The estimator $\hat{\beta}$ is then found by averaging $\hat{\beta}_n$ over a range Δn corresponding to time scales that are judged to be most important. In the case of transit photometry, the duration of ingress or egress is the most relevant time scale (corresponding to averaging time scales on the order of tens of minutes, in our example light curve). A white analysis is then performed, using the noise parameter $\sigma = \beta\sigma_1$ instead of σ_1. This causes the parameter errors $\hat{\sigma}_{p_k}$ to increase by β but does not change the parameter estimates $\hat{p}_k$ themselves.[6]

A second method is the "residual permutation" method that has been used by Jenkins et al. (2002), Moutou et al. (2004), Southworth (2008), Bean et al. (2008), Winn et al. (2008), and others. This method is a variant of a bootstrap analysis, in which the posterior probability distribution for the parameters is based on the collection of results of minimizing χ^2 (assuming white noise) for a large number of synthetic data sets. In the traditional bootstrap analysis the synthetic data sets are produced by scrambling the residuals and adding them to a model light curve, or by drawing data points at random (with replacement) to make a simulated data set with the same number of points as the actual data set. In the residual permutation method, the synthetic data sets are built by performing a cyclic permutation of the time indices of the residuals, and then adding them to the model light curve. In this way, the synthetic data sets have the same bumps, wiggles, and ramps as the actual data, but they are translated in time. The parameter errors are given by the widths of the distributions in the parameters that are estimated from all the different realizations of the synthetic data, and they are usually larger than the parameter errors returned by a purely white analysis.

As before, we limited the scope of the comparison to the estimation of t_c and its uncertainty. We created 5,000 realizations of a noise source with $\gamma = 1$ and a given value of α (either 0, 1/3, 2/3, or 1). We used each of the two approximate methods (time-averaging and residual-permutation) to calculate $\hat{\beta}$ and its uncertainty based on each of the 5,000 noise realizations. Then we found the median and standard deviation of $\hat{\beta}/\beta$ over all 5,000 realizations. Table (3.3) presents the results of this experiment.

Both methods, time-averaging and residual-permutation, gave more reliable uncertainties than the white method. However they both underestimated the true uncertainties by approximately 15-30%. Furthermore, neither method provided more

---

[6] Alternatively one may assign an error to each data point equal to the quadrature sum of the measurement error and an extra term σ_r (Pont et al. 2006). For cases in which the errors in the data points are all equal or nearly equal, these methods are equivalent. When the errors are not all the same, it is more appropriate to use the quadrature-sum approach of Pont et al. (2006). In this paper all our examples involve homogeneous errors.

accurate estimates of t_c than did the white method. For the time-averaging method as we have implemented it, this result is not surprising, for that method differs from the white method only in the inflation of the error bars by some factor β. The parameter values that maximize the likelihood function were unchanged.

3.4.5 Alternative noise models

We have shown the wavelet method to work well in the presence of $1/f^\gamma$ noise. Although this family of noise processes encompasses a wide range of possibilities, the universe of possible correlated noise processes is much larger. In this section we test the wavelet method using simulated data that has correlated noise of a completely different character. In particular, we focus on a process with exclusively short-term correlations, described by one of the aforementioned autoregressive moving-average (ARMA) class of parametric noise models. In this way we test our method on a noise process that is complementary to the longer-range correlations present in $1/f^\gamma$ noise, and we also make contact between our method and the large body of statistical literature on ARMA models.

For $1/f^\gamma$ noise we have shown that time-domain parameter estimation techniques are slow. However, if the noise has exclusively short-range correlations, the autocorrelation function will decay with lag more rapidly than a power law, and the truncated-χ^2 likelihood [Eqn. (3.9)] may become computationally efficient. ARMA models provide a convenient analytic framework for parameterizing such processes. For a detailed review of ARMA models and their use in statistical inference, see Box & Jenkins (1976). Applications of ARMA models to astrophysical problems have been described by in Koen & Lombard (1993), Konig & Timmer (1997) and Timmer et al. (2000).

To see how the wavelet method performs on data with short-range correlations we constructed synthetic transit data in which the noise is described by a single-parameter autoregressive [AR(1; ψ)] model. An AR(1; ψ) process $\epsilon(t_i)$ is defined by the recursive relation

$$\epsilon(t_i) = \eta(t_i) + \psi\epsilon(t_{i-1}) \tag{3.37}$$

where $\eta(t_i)$ is an uncorrelated Gaussian process with width parameter σ and ψ is the sole autoregressive parameter. The autocorrelation $\gamma(l)$ for an AR(1; ψ) process is

$$\gamma(l) = \frac{\sigma^2}{1-\psi^2}\psi^l. \tag{3.38}$$

An AR(1;ψ) process is stationary so long as $0 < \psi < 1$ (Box & Jenkins 1976). The decay length of the autocorrelation function grows as ψ is increased from zero to one. Figure (3-9) plots the autocorrelation function of a process that is an additive combination of an AR(1; $\psi = 0.95$) process and a white noise process. The noise in our synthetic transit light curves was the sum of this AR(1; $\psi = 0.95$) process, and white noise, with $\alpha = 1/2$ (see Fig 3-9). With these choices, the white method underestimates the error in t_c, while at the same time the synthetic data look realistic.

We proceeded with the MCMC method as described previously to estimate the time of mid-transit. All four methods assessed in the previous section were included in this analysis, for comparison. Table 3.5 gives the results. The wavelet method produces more reliable error estimates than the white method. However, the wavelet method no longer stands out as superior to the time-averaging method or the residual-permutation method; all three of these methods give similar results. This illustrates the broader point that using any of these methods is much better than ignoring the noise correlations. The results also show that although the wavelet method is specifically tuned to deal with $1/f^\gamma$ noise, it is still useful in the presence of noise with shorter-range correlations.

It is beyond the scope of this paper to test the applicability of the wavelet method on more general ARMA processes. Instead we suggest the following approach when confronted with real data [see also Beran (1994)]. Calculate the sample autocorrelation, and power spectral density, based on the out-of-transit data or the residuals to an optimized transit model. For stationary processes these two indicators are related

Table 3.5. Estimates of t_c from data with autoregressive correlated noise

Method	$\langle \hat{\sigma}_{t_c} \rangle$ [sec]	$\langle \mathcal{N} \rangle$	$\sigma_{\mathcal{N}}$	$prob(\mathcal{N} > 1)$	$prob$(better)[a]
White	4.5	−0.010	2.50	70%	–
Wavelet	8.7	−0.016	1.33	44%	51%
Time-averaging	9.9	−0.010	1.25	40%	49%
Residual-permutation	10.2	−0.010	1.23	38%	51%

[a] The probability that the analysis method returns an estimate of t_c that is closer to the true value than the white analysis.

as described in § 3.2. Short-memory, ARMA-like processes can be identified by large autocorrelations at small lags or by finite power spectral density at zero frequency. Long-memory processes ($1/f^{\gamma}$) can be identified by possibly small but non-vanishing autocorrelation at longer lags. Processes with short-range correlations could be analyzed with an ARMA model of the covariance matrix [see Box & Jenkins (1976)], or the truncated-lag covariance matrix, although a wavelet-based analysis may be sufficient as well. Long-memory processes are best analyzed with the wavelet method as described in this paper.

It should also be noted that extensions of ARMA models have been developed to mimic long-memory, $1/f^{\gamma}$ processes. In particular, fractional autoregressive integrated moving-average models (ARFIMA) describe "nearly" $1/f^{\gamma}$ stationary processes, according to the criterion described by Beran (1994). As is the case with ARMA models, ARFIMA models enjoy analytic forms for the likelihood in the time-domain. Alas, as noted by Wornell (1996) and Beran (1994), the straightforward calculation of this likelihood is computationally expensive and potentially unstable. For $1/f^{\gamma}$ processes, the wavelet method is probably a better choice than any time-domain method for calculating the likelihood.

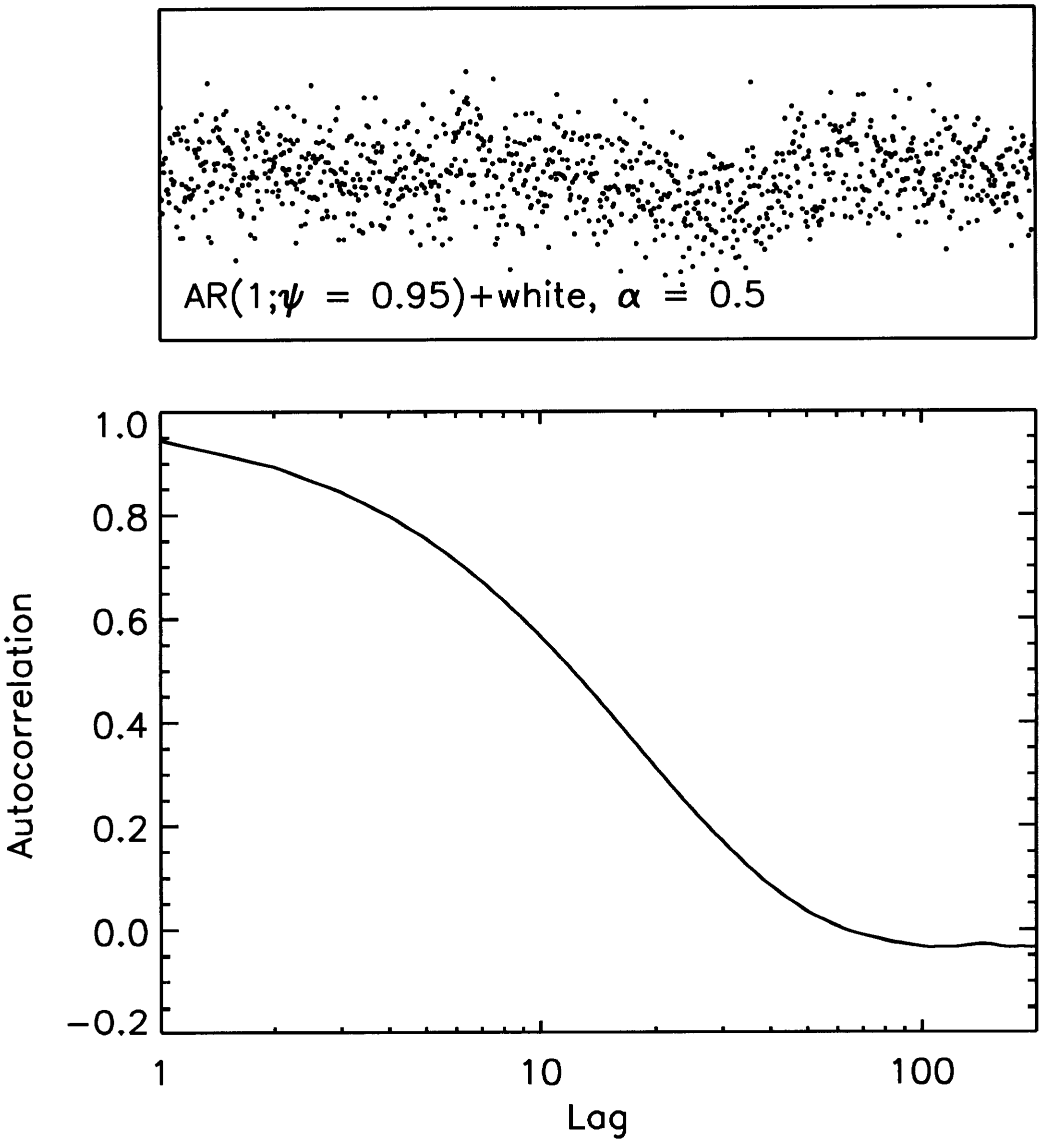

Figure 3-9 An example of an autoregressive noise process with complementary characteristics to a $1/f^{\gamma}$ process. The top panel shows the sum of an AR(1) process with $\psi = 0.95$ and white noise. The correlated and uncorrelated components have equal variances ($\alpha = 0.5$).

3.4.6 Transit timing variations estimated from a collection of light curves

We present here an illustrative calculation that is relevant to the goal of detecting planets or satellites through the perturbations they produce on the sequence of midtransit times of a known transiting planet. Typically an observer would fit the midtransit times $t_{c,i}$, to a model in which the transits are strictly periodic:

$$t_{c,i} \;=\; t_{c,0} + E_i P \tag{3.39}$$

for some integers E_i and constants $t_{c,0}$ and P. Then, the residuals would be computed by subtracting the best-fit model from the data, and a test for anomalies would be performed by assessing the likelihood of obtaining those residuals if the linear model were correct. Assuming there are N data points with normally-distributed, independent errors, the likelihood is given by a χ^2-distribution, $prob(\chi^2, N_{\rm dof})$, where

$$\chi^2 = \sum_i \left[\frac{t_{c,i} - (t_{c,0} + E_i P)}{\sigma_{t_{c,i}}} \right]^2 \tag{3.40}$$

and $N_{\rm dof} = N - 2$ is the number of degrees of freedom. Values of χ^2 with a low probability of occurrence indicate the linear model is deficient, that there are significant anomalies in the timing data, and that further observations are warranted.

We produced 10 simulated light curves of transits of the particular planet GJ 436b, a Neptune-sized planet transiting an M dwarf (Butler et al. 2004, Gillon et al. 2007) which has been the subject of several transit-timing studies (see, e.g., Ribas et al. 2008, Alonso et al. 2008, Coughlin et al. 2008). Our chosen parameters were $R_p/R_\star = 0.084$, $a/R_\star = 12.25$, $i = 85.94$ deg, and $P = 2.644$ d. This gives $\delta = 0.007$, $T = 1$ hr, and $\tau = 0.24$ hr. We chose limb-darkening parameters as appropriate for the SDSS r band (Claret 2004). We assumed that 10 consecutive transits were observed, in each case giving 512 uniformly-sampled flux measurements over 2.5 hours centered on the transit time. Noise was synthesized in the Fourier domain (as in § 3.4.2), with a white component $\sigma_w = 0.001$ and a $1/f$ component with rms 0.0005 ($\alpha = 1/2$). The

Table 3.6. Linear fits to estimated midtransit times

Method	Fitted Period / True Period	$\hat{\chi}^2/N_{\rm dof}$	$prob(\chi^2 < \hat{\chi}^2)$
White	$1.00000071 \pm 0.00000043$	2.25	98%
Wavelet	$1.00000048 \pm 0.00000077$	0.93	51%

10 simulated light curves are plotted in Fig. (3-10). Visually, they resemble the best light curves that have been obtained for this system.

To estimate the midtransit time of each simulated light curve, we performed a wavelet analysis and a white analysis, allowing only the midtransit time and the noise parameters to vary while fixing the other parameter values at their true values. We used the same MCMC technique that was described in § 3.4.2. Each analysis method produced a collection of 10 midtransit times and error bars. These 10 data points were then fitted to the linear model of Eqn. (3.39). Fig. (3-11) shows the residuals of the linear fit (observed − calculated). Table 3.6 gives the best-fit period for each analysis (wavelet or white), along with the associated values of χ^2.

As was expected from the results of § 3.4.2, the white analysis gave error bars that are too small, particularly for epochs 4 and 7. As a result, the practitioner of the white analysis would have rejected the hypothesis of a constant orbital period with 98% confidence. In addition, the white analysis gave an estimate for the orbital period that is more than 1σ away from the true value, which might have complicated the planning and execution of future observations. The wavelet method, in contrast, neither underestimated nor overestimated the errors, giving $\chi^2 \approx N_{\rm dof}$ in excellent agreement with the hypothesis of a constant orbital period. The wavelet method also gave an estimate for the orbital period within 1σ of the true value.

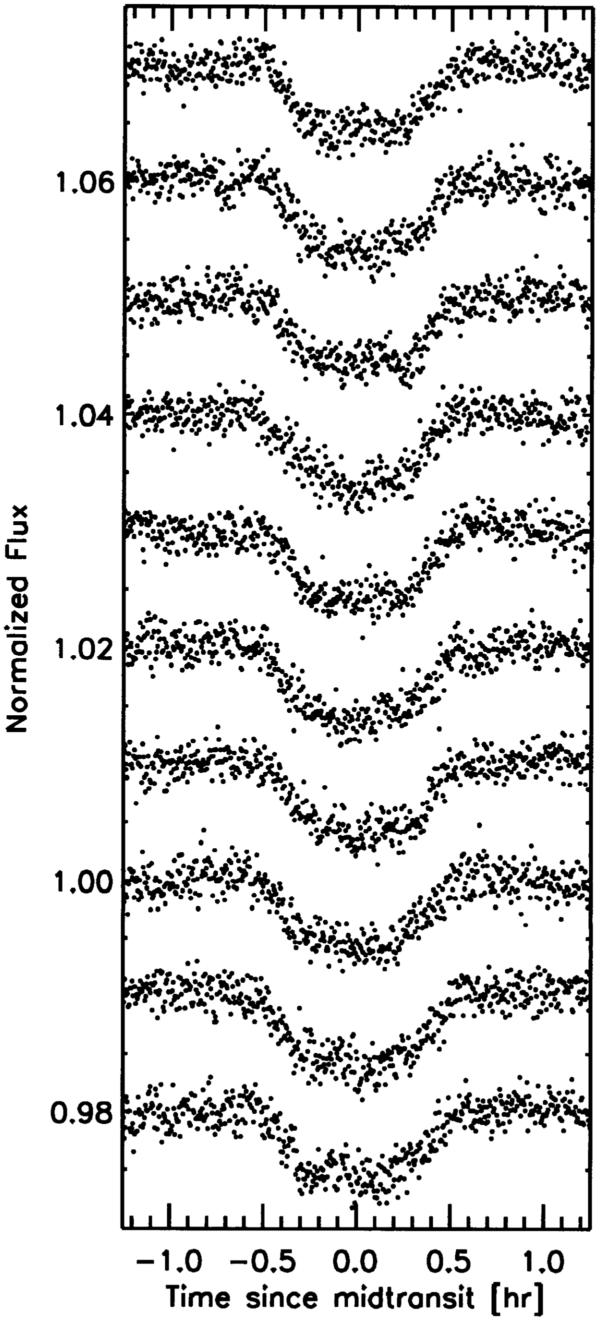

Figure 3-10 Simulated transit observations of the "Hot Neptune" GJ 436. Arbitrary vertical offsets have been applied to the light curves, to separate them on the page.

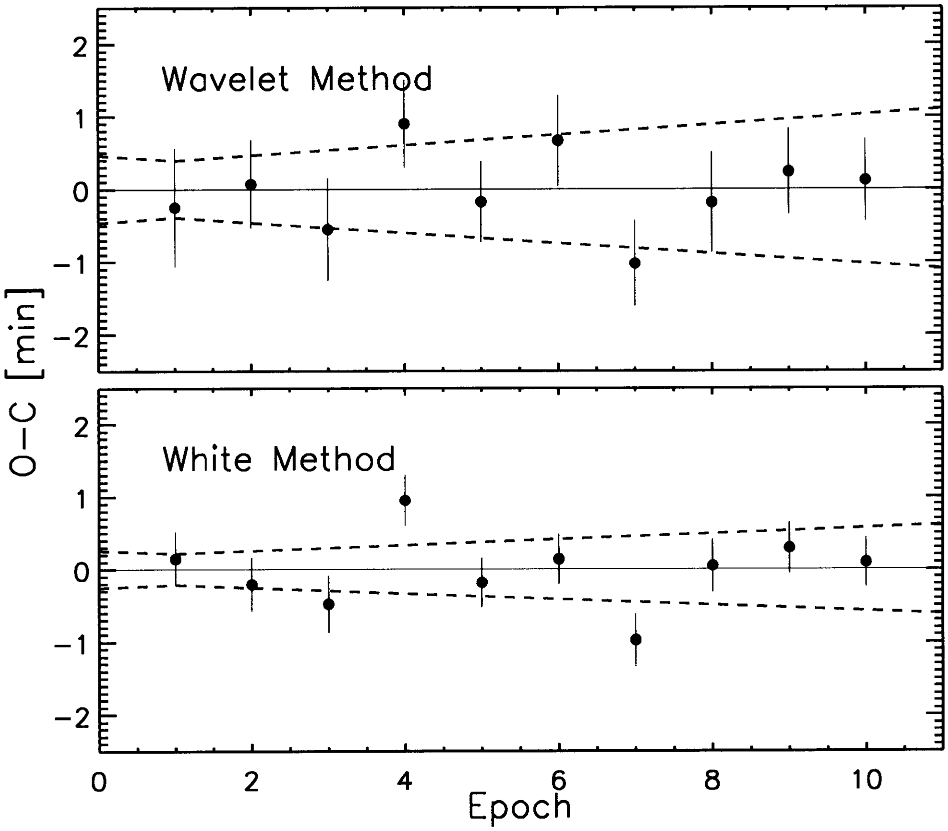

Figure 3-11 Transit timing variations estimated from simulated transit observations of GJ 436b. Each panel shows the residuals (observed − calculated) of a linear fit to the estimated midtransit times. The midtransit times were estimated with a wavelet analysis and also with a white analysis, as described in the text. The dashed lines indicate the 1 σ errors in the linear model.

3.4.7 Estimation of multiple parameters

Thus far we have focused exclusively on the determination of the midtransit time, in the interest of simplicity. However, there is no obstacle to using the wavelet method to estimate multiple parameters, even when there are strong degeneracies among them. In this section we test and illustrate the ability of the wavelet method to solve for all the parameters of a transit light curve, along with the noise parameters.

We modeled the transit as in §§ 3.4.1 and 3.4.2. The noise was synthesized in the frequency domain (as in § 3.4.2), using $\sigma_w = 0.0045$, $\gamma = 1$, and $\alpha = 1/2$. The resulting simulated light curve is the upper time series in Fig. 3-12. We used the MCMC method to estimate the transit parameters $\{R_p/R_\star, a/R_\star, i, t_c\}$ and the noise parameters $\{\sigma_r, \sigma_w\}$ (again fixing $\gamma = 1$ for simplicity). The likelihood was evaluated with either the wavelet method [Eqn. (3.32)] or the white method [Eqn. (3.6)].

Fig. 3-13 displays the results of this analysis in the form of the posterior distribution for the case of t_c, and the joint posterior confidence regions for the other cases. The wavelet method gives larger (and more appropriate) confidence regions than the white analysis. In accordance with our previous findings, the white analysis underestimates the error in t_c and gives an estimate of t_c that differs from the true value by more than 1σ. The wavelet method gives better agreement. Both analyses give an estimate for $R_p/R_\star$ that is smaller than the true value of 0.15, but in the case of the white analysis, this shift is deemed significant, thereby ruling out the correct answer with more than 95% confidence. In the wavelet analysis, the true value of $R_p/R_\star$ is well within the 68% confidence region. Both the wavelet and white analyses give accurate values of $a/R_\star$ and the inclination, and the wavelet method reports larger errors. As shown in Fig. (3-13), the wavelet method was successful at identifying the parameters (α and σ_w) of the underlying $1/f$ noise process.

Fig. 3-12 shows the best-fitting transit model, and also illustrates the action of the "whitening" filter that was described in § 3.3.4. The jagged line plotted over the upper time series is the best estimate of the $1/f$ contribution to the noise, found by applying the whitening filter [Eqn. (3.29)] to the data using the estimated noise parameters.

The lower time series is the whitened data, in which the $1/f$ component has been subtracted. Finally, in Fig. 3-14 we compare the estimated $1/f$ noise component with the actual $1/f$ component used to generate the data. Possibly, by isolating the correlated component in this way, and investigating its relation to other observable parameters, the physical origin of the noise could be identified and understood.

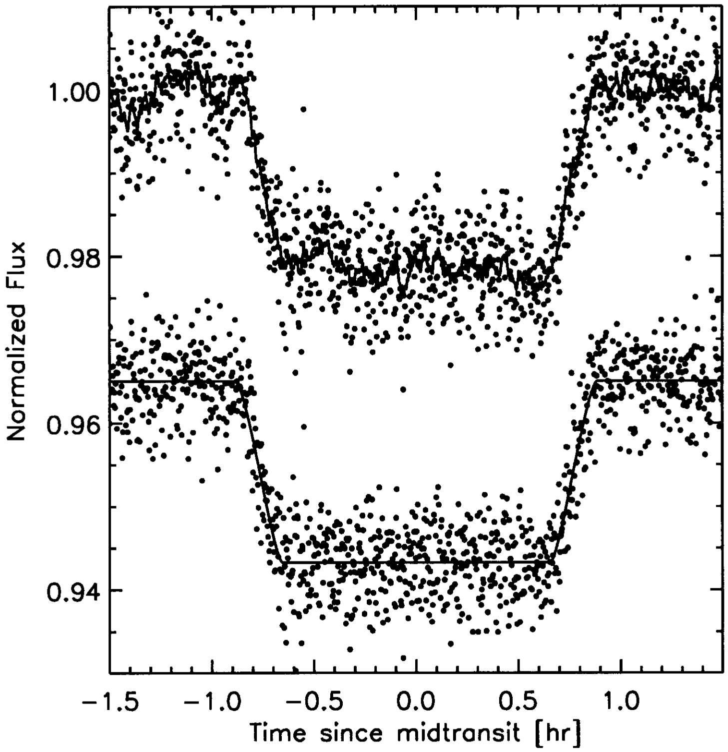

Figure 3-12 Wavelet analysis of a single simulated transit light curve. *Top.*— Simulated light curve with correlated noise. The jagged line is the best-fitting transit model plus the best-fitting model of the $1/f$ component of the noise. *Bottom.*— Simulated light curve after applying the whitening filter of Eqn. (3.29), using the noise parameters estimated from the wavelet analysis. The solid line is the best-fitting transit model.

3.5 Summary and Discussion

In this paper we have introduced a technique for parameter estimation based on fitting a parametric model to a time series that may be contaminated by temporally correlated noise with a $1/f^{\gamma}$ power spectral density. The essence of the technique

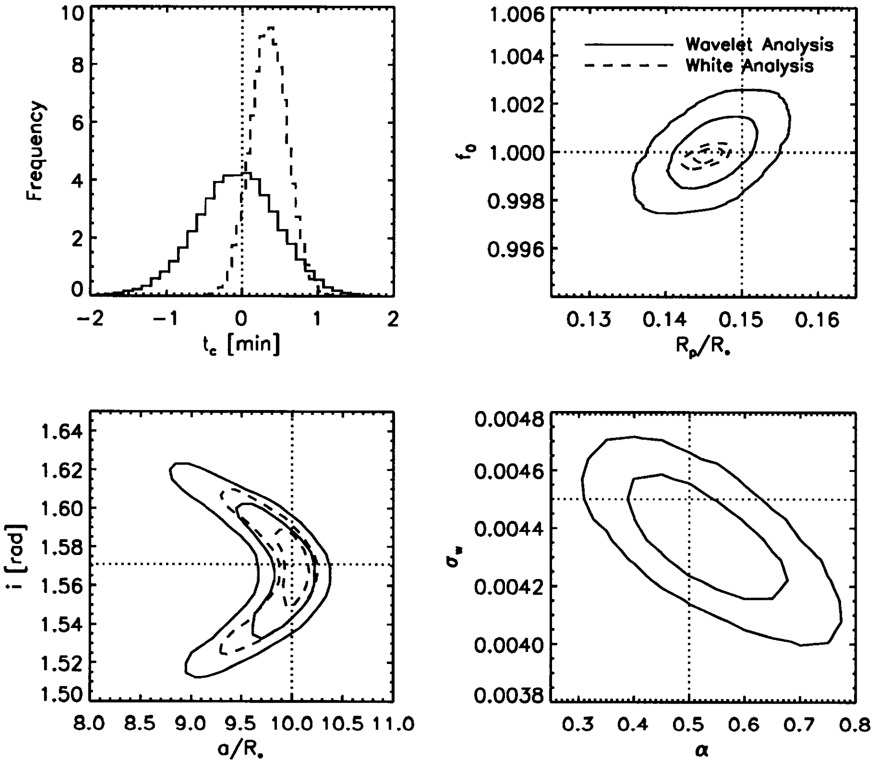

Figure 3-13 Results of parameter estimation for the simulated light curve of Fig. 3-12. Results for both the wavelet method (solid lines) and the white method (dashed lines) are compared. The upper left panel shows the posterior distribution for the midtransit time. The other panels show confidence contours (68.3% and 95.4%) of the joint posterior distribution of two parameters. The true parameter values are indicated by dotted lines.

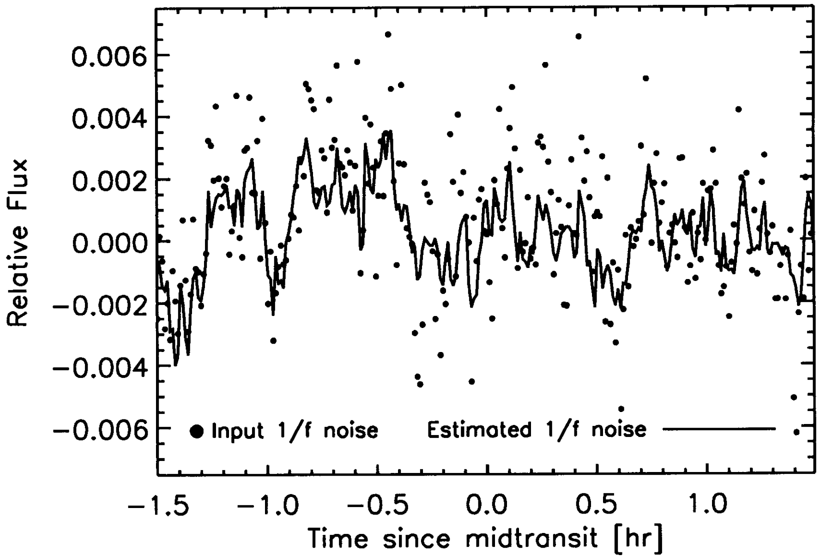

Figure 3-14 Isolating the correlated component. Plotted are the actual and estimated $1/f$ components of the noise in the simulated light curve plotted in Fig. (3-12). The estimated $1/f$ signal was found by applying the wavelet filter, Eqn. (3.29), to the residuals.

is to calculate the likelihood function in a wavelet basis. This is advantageous because a broad class of realistic noise processes produce a nearly diagonal covariance matrix in the wavelet basis, and because fast methods for computing wavelet transforms are available. We have tested and illustrated this technique, and compared it to other techniques, using numerical experiments involving simulated photometric observations of exoplanetary transits.

For convenience we summarize the likelihood calculation here:

- Given the N data points $y(t_i)$ obtained at evenly-spaced times t_i, subtract the model $f_i(t_i;\vec{p})$ with model parameters $\vec{p}$ to form the N residuals $r(t_i) \equiv y(t_i) - f(t_i;\vec{p})$.

- If N is not a multiple of a power of two, either truncate the time series or enlarge it by padding it with zeros, until $N = n_0 2^M$ for some $n_0 > 0$, $M > 0$.

- Apply the Fast Wavelet Transform (FWT) to the residuals to obtain $n_0(2^M - 1)$ wavelet coefficients r_n^m and n_0 scaling coefficients $\bar{r}_n^1$.

- For stationary, Gaussian noise built from an additive combination of uncorrelated and correlated noise (with Power Spectral Density $\mathcal{S}(f) \propto 1/f^{\gamma}$), the likelihood for the residuals $r(t_i)$ is given by

$$\begin{aligned} \mathcal{L} &= \left\{ \prod_{m=2}^{M} \prod_{n=1}^{n_0 2^{m-1}} \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left[-\frac{(r_n^m)^2}{2\sigma_W^2} \right] \right\} \\ &\times \left\{ \prod_{n=1}^{n_0} \frac{1}{\sqrt{2\pi\sigma_S^2}} \exp\left[-\frac{(\bar{r}_n^1)^2}{2\sigma_S^2} \right] \right\} \end{aligned} \tag{3.41}$$

  where

$$\sigma_W^2 = \sigma_r^2 2^{-\gamma m} + \sigma_w^2 \tag{3.42}$$

$$\sigma_S^2 = \sigma_r^2 2^{-\gamma} g(\gamma) + \sigma_w^2 \tag{3.43}$$

  for some noise parameters $\sigma_w > 0$, $\sigma_r > 0$ and $g(\gamma) = O(1)$ [e.g., $g(1) \approx 0.72$].

The calculation entails the multiplication of N terms and has an overall time-complexity of $O(N)$. With this prescription for the likelihood function, the parameters may be optimized using any number of traditional algorithms. For example, the likelihood may be used in the jump-transition probability in a Monte Carlo Markov Chain analysis, as we have done in this work.

Among the premises of this technique are that the correlations among the wavelet and scaling coefficients are small enough to be negligible. In fact, the magnitude of the correlations at different scales and times are dependent on the choice of wavelet basis and the spectral index γ describing the power spectral density of the correlated component of the noise. We have chosen for our experiments the Daubechies 4th-order wavelet basis which seems well-suited to the cases we considered. A perhaps more serious limitation is that the noise should be stationary. Real noise is often nonstationary. For example, photometric observations are noisier during periods of poor weather, and even in good conditions there may be more noise at the beginning or end of the night when the target is observed through the largest airmass. It is possible that this limitation could be overcome with more elaborate noise models, or by analyzing the time series in separate segments; future work on these topics may be warranted.

Apart from the utility of the wavelet method, we draw the following conclusions based on the numerical experiments of § 3. First, any analysis that ignores possible correlated errors (a "white" analysis in our terminology) is suspect, and any 2–3σ results from such an analysis should be regarded as provisional at best. As shown in §§ 3.4.1, 3.4.2, and 3.4.6, even data that appear "good" on visual inspection and that are dominated by uncorrelated noise may give parameter errors that are underestimated by a factor of 2–3 in a white analysis. Second, using any of the methods described in 3.4.4 (the wavelet method, the time-averaging method, or the residual-permutation method) is preferable to ignoring correlated noise altogether.

Throughout this work our main application has been estimation of the parameters of a single time series or a few such time series, especially determining the midtransit times of transit light curves. One potentially important application that we have not discussed is the *detection* of transits in a database of time-series photometry of many stars. Photometric surveys such as the ground-based HAT (Bakos et al. 2007) and SuperWASP (Pollacco et al. 2006), and space-based missions such as Corot (Baglin et al. 2003) and Kepler (Borucki et al. 2003) produce tens to hundreds of thousands of time series, spanning much longer intervals than the transit durations. It seems likely that the parameters of a noise model could be very well constrained using these vast databases, and that the application of a wavelet-based whitening filter could facilitate the detection of transits and the elimination of statistical false positives. Popular techniques for dealing with correlated noise in large photometric databases are those of Tamuz et al. (2005), Kovács et al. (2005), and Pont et al. (2006). A priority for future work is to compare these methods with a wavelet-based method, by experimenting with realistic survey data.

We are grateful to Frederic Pont for a very detailed and constructive critique of an early version of this manuscript. We also thank Scott Gaudi and Jason Eastman for helpful comments. This work was partly supported by the NASA Origins program (grant no. NNX09AB33G).

Bibliography

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567

Alonso, R., Barbieri, M., Rabus, M., Deeg, H. J., Belmonte, J. A., & Almenara, J. M. 2008, A&A, 487, L5

Baglin, A. and the COROT team 2003, Advances in Space Research, 31, 345

Bakos, G. Á., et al. 2006, ApJ, 650, 1160

Bakos, G. Á., et al. 2007, ApJ, 656, 552

Bean, J. L., et al. 2008, A&A, 486, 1039

Beran, J. 1994, Statistics for Long-Memory Processes (London: Chapman & Hall)

Bevington, P. R., & Robinson, D. K. 2003, Data reduction and error analysis for the physical sciences, 3rd ed. (Boston, MA: McGraw-Hill)

Borucki, W. J., et al. 2003, Proc. SPIE, 4854, 129

Bouchy, F., Pont, F., Melo, C., Santos, N. C., Mayor, M., Queloz, D., & Udry, S. 2005, A&A, 431, 1105

Box, G. E. P., & Jenkins, G. M. 1976, Holden-Day Series in Time Series Analysis, Revised ed., San Francisco: Holden-Day, 1976,

Bracewell, R. The Fourier Transform and Its Applications. McGraw-Hill Electrical and Electronic Engineering Series, Terman, F., Ed. New York: McGraw-Hill, 1965, pg. 115

Burke, C. J., et al. 2007, ApJ, 671, 2115

Butler, R. P., Vogt, S. S., Marcy, G. W., Fischer, D. A., Wright, J. T., Henry, G. W., Laughlin, G., & Lissauer, J. J. 2004, ApJ, 617, 580

Carter, J. A., Yee, J. C., Eastman, J., Gaudi, B. S., & Winn, J. N. 2008, ApJ, 689, 499

Claret, A. 2004, A&A, 428, 1001

Collier Cameron, A., et al. 2007, MNRAS, 380, 1230

Coughlin, J. L., Stringfellow, G. S., Becker, A. C., López-Morales, M., Mezzalira, F., & Krajci, T. 2008, ApJ, 689, L149

Daubechies, I. 1988, Communications on Pure and Applied Mathematics, 41, 7, 909

Fadili, M.J. & Bullmore, E.T. 2002, *Neuroimage*, 15, 217

Gibson, N. P., et al. 2008, A&A, 492, 603

Gillon, M., Pont, F., Moutou, C., Bouchy, F., Courbin, F., Sohy, S., & Magain, P. 2006, A&A, 459, 249

Gillon, M., et al. 2007, A&A, 472, L13

Gillon, M., et al. 2009, A&A, 496, 259

Giménez, A. 2007, A&A, 474, 1049

Gould, A. 2003, ArXiv Astrophysics e-prints, arXiv:astro-ph/0310577

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with *Mathematica* Support (Cambridge Univ. Press, UK).

Haar, A., Zur Theorie der Orthogonalen Funktionensysteme., Math. Annal., Vol. 69, pp. 331-371, 1910

Holman, M. J., & Murray, N. W. 2005, Science, 307, 1288

Holman, M. J., et al. 2006, ApJ, 652, 1715

Jenkins, J. M., Caldwell, D. A., & Borucki, W. J. 2002, ApJ, 564, 495

Johnson, J. A., Winn, J. N., Cabrera, N. E., & Carter, J. A. 2009, ApJ, 692, L100

Kipping, D. M. 2009, MNRAS, 392, 181

Koen, C., & Lombard, F. 1993, MNRAS, 263, 287

Konig, M., & Timmer, J. 1997, A&AS, 124, 589

Kovács, G., Bakos, G., & Noyes, R. W. 2005, MNRAS, 356, 557

Mallat, S. G. 1989, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 674

Mallat, S. A Wavelet Tour of Signal Processing, London:AP Professional, 1997

Mandel, K., & Agol, E. 2002, ApJ, 580, L171

Percival D.B., & A.T. Walden. Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques. Cambridge: Cambridge University Press, 1993

Pollacco, D. L., et al. 2006, PASP, 118, 1407

Pont, F., Zucker, S., & Queloz, D. 2006, MNRAS, 373, 231

Press, W. H. 1978, Comments on Astrophysics, 7, 103

Press, William H., Teukolsky, Saul A., Vetterling, William T., & Flannery, Brian P. 2007, Numerical Recipes: The Art of Scientific Computing, 3rd ed. (New York: Cambridge University Press)

Ribas, I., Font-Ribera, A., & Beaulieu, J.-P. 2008, ApJ, 677, L59

Southworth, J. 2008, MNRAS, 386, 1644

Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466

Teolis, A. Computational Signal Processing with Wavelets. Boston:Birkhäuser, 1998

Timmer, J., Schwarz, U., Voss, H. U., Wardinski, I., Belloni, T., Hasinger, G., van der Klis, M., & Kurths, J. 2000, Phys. Rev. E, 61, 1342

Udalski, A., Szymanski, M. K., Kubiak, M., Pietrzynski, G., Soszynski, I., Zebrun, K., Szewczyk, O., & Wyrzykowski, L. 2004, Acta Astronomica, 54, 313

Winn, J. N., et al. 2007, AJ, 133, 1828

Winn, J. N., et al. 2008, ApJ, 683, 1076

Winn, J. N., Holman, M. J., Carter, J. A., Torres, G., Osip, D. J., & Beatty, T. 2009, AJ, 137, 3826

Wornell, G. Signal Processing with Fractals: A Wavelet-Based Approach. Prentice Hall Signal Processing Series, Oppenheim, A., Ed. Upper Saddle River, NJ:Prentice Hall Press, 1996

Chapter 4

Near-infrared transit photometry of the exoplanet HD 149026b

4.1 Introduction

Since its discovery by Sato et al. (2005), HD 149026b has been one of the most closely scrutinized planets outside the Solar system. It is a close-in gas giant, orbiting a G star with a period of only 2.5 d. Observations of transits (Sato et al. 2005, Charbonneau et al. 2006, Winn et al. 2008b, Nutzman et al. 2008), in combination with observations of radial-velocity variations of the parent star (Sato et al. 2005), have shown that the planet has approximately Saturn's mass but is considerably denser, despite the intense irradiation from the parent star that should inflate the planet and lower its density. There is consensus among theorists that the reason for the "shrunken radius" is a highly metal-enriched composition, although the total metal mass, its distribution within the planet, and the reason for the enrichment are debated [Sato et al. (2005), Fortney et al. (2006), Ikoma et al. (2006), Broeg & Wuchterl (2007), Burrows et al. (2007)]. The total metal mass, for example, ranges from 60 $M_\oplus$ to 114 $M_\oplus$ among the possible models. The latter estimate would represent 80% of the total mass of the planet.

The planet's outer atmosphere is also of interest, given the possibly unusual composition and the strong heating from the parent star. Models by Fortney et al. (2005)

indicated the possibility of a very hot stratosphere as a result of gaseous TiO and VO opacity. By using the *Spitzer Space Telescope* to observe a planetary occultation, Harrington et al. (2007) found the planet's 8 μm brightness temperature to be much larger than the temperature that one would expect based on thermal equilibrium with the incident stellar radiation. This may be the result of the predicted TiO and VO heating, although the details of whether and where these absorbers actually condense in the atmosphere are not yet understood (Fortney et al. 2005, Harrington et al. 2007).

Fundamental to all these discussions are the measurements of the mass and radius of HD 149026b. These measurements are limited by the uncertainties in the stellar mass and radius. One way to improve the situation is to observe transits with greater photometric precision than has been possible before. As shown by Seager & Mallen-Ornelas (2003), with a good light curve and Kepler's third law, one may determine the stellar mean density. If the mean density is known precisely enough, it is a key constraint that can be combined with the other stellar observables (parallax, apparent magnitude, effective temperature, metallicity, etc.) and stellar-evolutionary models to determine the stellar mass and radius. This technique has been put into practice for many other systems [see, e.g., Sozzetti et al. (2007), Holman et al. (2007), Torres et al. (2008)] but never to advantage for HD 149026b because of the limited precision of prior determinations of $\rho_\star$ (Winn et al. 2008b, Nutzman et al. 2008). Observers must cope with the small transit depth of 2.5 mmag (smaller than any other transiting exoplanet by a factor of two) and the paucity of suitable comparison stars within the field of view of most telescopes.

In this chapter we present observations of transits of HD 149026b with the Near Infrared Camera and Multi-Object Spectrometer [NICMOS, Thompson (1992)] on board the *Hubble Space Telescope* (*HST*). We chose this instrument because a high precision in relative photometry is possible even without using comparison stars (Gilliland 2006, Swain et al. 2008) and because the reduced stellar limb-darkening at near-infrared wavelengths is advantageous for the light-curve analysis (Carter et al. 2008, Pál et al. 2008). We have organized this chapter as follows. In § 4.2 we describe the observations and data reductions leading to the final photometric time series. In § 4.3 we describe our photometric model and the results of the NICMOS light-curve analysis. In § 4.4, we describe how the light-curve results were incorporated into stellar-evolutionary models to determine the parameters of the HD 149026 system. In § 4.5, the light analysis is repeated using not only the NICMOS data but also the most precise light curves that have been obtained at optical and mid-infrared wavelengths. In § 4.6, all the available transit times are analyzed to produce a refined transit ephemeris and to search for possible period variations that could be indicative of additional bodies in the HD 149026 system (Holman & Murray 2005, Agol et al. 2005, Ford & Holman 2007). Finally, in § 4.7, we discuss the possible implications of our observations and analysis.

4.2 Observations and Reductions

We observed HD 149026 on four occasions ("visits" in *HST* parlance) when transits were predicted to occur, on 2007 Dec 22, 2007 Dec 24, 2008 Feb 08, and 2008 Mar 20. Each visit consisted of five orbits spanning a transit. Between each pair of orbits is an observing gap of approximately 45 minutes, when *HST*'s vision is blocked by the Earth. The visits were scheduled in such a manner that the combined data set provides complete phase coverage of the transit, including redundant coverage of the critical ingress and egress phases. In particular, visits 1 and 3 covered the ingress phase, and visit 2 covered both ingress and egress phases. Visit 4 captured the beginning of egress.

We used Camera 3 of the NICMOS detector, a 256 × 256 HgCdTe array with a field of view of $51.2'' \times 51.2''$. We used the G141 grism filter, which is centered at 1.4μm, spans 0.8 μm and is roughly equivalent to H band. An exposure was obtained every 13 s. The camera was operated in "MULTIACCUM" mode, wherein five nondestructive readouts are recorded during a single exposure, and the first readout is subtracted from the final readout. After accounting for overheads, the effective integration time was 4 s per exposure. We deliberately defocused the instrument to give a full-width at half-maximum (FWHM) of approximately 5 pixels in the cross-dispersion direction. This was done for two reasons: firstly, when focused, camera 3 undersamples the point-spread–function (PSF) of point sources; and secondly, the detector pixels exhibit intra-pixel sensitivity variations as large as 30%. Defocusing the images causes the PSF to be well-sampled and averages over the intra-pixel sensitivity variations.

Approximately 220 exposures were collected during each *HST* orbit. Experience with *HST* has shown that photometric stability is relatively poor during the first orbit of a given visit. Our observations were scheduled under the assumption that the first orbit from each visit would not be utilized, and indeed we ended up omitting the first-orbit data from our analysis. At the start of each visit, we obtained a single non-dispersed image using a narrow filter centered at 1.66 μm in order to establish the pixel position corresponding to zero dispersion. We then adopted the relation from the *HST* Data Handbook for NICMOS[1],

$$\lambda(\Delta x) = -0.007992\ \Delta x + 1.401 \tag{4.1}$$

where $\lambda(\Delta x)$ is the wavelength (measured in μm), and Δx is the x coordinate (measured in pixels) relative to the center of the undispersed image.

For completeness, we performed the steps of flat-fielding, background subtraction, and pixel flagging, as described below; however, it is noteworthy that these steps in the data reduction made very little difference in the aperture photometry or in the final results. Flat-field correction for grism images is not straightforward and is not done as part of the standard NICMOS pipeline reductions, because the appropriate flat field depends both upon wavelength and upon the position of the source in the non-dispersed image. To accomplish the flat-field correction, we obtained seven flat fields, each using a narrow bandpass within the G141 bandpass, and fitted the data for each pixel with a quadratic function of wavelength $C[\lambda; x, y]$. We then applied $C[\lambda(\Delta x); x, y]$ as a multiplicative correction to each of our science images, using

---

[1] `http://www.stsci.edu/hst/nicmos/documents/handbooks/DataHandbookv7/`

the wavelength-coordinate relation $\lambda(\Delta x)$ that was determined from the single non-dispersed image. The background level was estimated in each image based on the counts in a relatively clean region of the detector (away from the spectral trace) and subtracted from the entire image. To identify bad pixels, all images from a given orbit were used to create a time series of counts specific to each pixel. Pixels showing an anomalously large variance were flagged. The list of flagged pixels was appended to the list of hot or cold pixels that were identified in the standard NICMOS pipeline reductions, and the values of all of those bad pixels were replaced by interpolated values of the neighboring good pixels.

Aperture photometry was performed on the first-order spectrum, using a simple sum of the counts within a rectangular box centered on the spectral trace. The box had a width of 20 pixels in the cross-dispersion direction (the y direction), which was four times the FWHM of the PSF. The box had a length of 120 pixels in the dispersion direction (the x direction), which was long enough the capture the entire first-order spectrum.

At this stage the data had been reduced to a single number per image: the total number of counts in the aperture (the "flux"). We examined the resulting time series. As expected, the data collected during the first orbits of each visit showed flux variations that were both larger in amplitude and different in their time-dependence than the variations observed in subsequent orbits. The first-orbit data were excluded from subsequent analysis. In addition, we excluded the 10 exposures near the beginning and end of each orbit sequence, because they showed strong flux variations that are probably due to "Earth shine." After these exclusions, there remained 800, 820, 800, and 792 good data points in visits 1, 2, 3, and 4, respectively.

4.3 NICMOS Light-Curve Analysis

Fig. 4-1 shows the time series of the aperture-summed flux, after dividing by the mean flux. In each panel, the zero point of the x-axis is the expected mid-transit time. The flux decrement during the transit is identifiable, but this decrement is superimposed on at least two other sources of variability: orbit-to-orbit discontinuities, and smooth intra-orbital variability showing a consistent pattern among all orbits of a given visit. The intra-orbital variability has been seen by all other investigators attempting precise *HST* photometry of single bright stars, since the pioneering work by Brown et al. (2001), and the flux discontinuities have been seen by other investigators using NICMOS (see, e.g., Swain et al. 2008). The origins of these systematic effects have not been established. The orbit-to-orbit consistency of the smooth variations suggests a phenomenon that is a function of the phase of the telescope's orbit around the Earth, such as thermal cycling or scattered light. The discontinuities between orbits suggest a non-repeating event associated with the re-aquisition of the target star after each Earth occultation, such as pointing changes or positional shifts of the grism filter.

Ideally, the underlying physical processes giving rise to these systematic effects could be ascertained, and this understanding would lead to either the recognition of an improved method for deriving the photometric signal or a physical model that could be used to correct the aperture-summed flux. Given that we do not yet have such knowledge, what can be done? The intra-orbital variations are very well-described by a smooth function of the *HST* orbital phase; following other investigators we used a smooth function with several adjustable parameters as an *ad hoc* model for this variation. The parameters of this model are fairly well constrained by the out-of-transit data, for which all variations are assumed to be systematic effects.

The inter-orbital discontinuities are more problematic. To investigate the systematic effects, we examined the spectral trace on each image. Specifically we computed the flux-weighted mean y position as a function of x, giving a curve $y(x)$ representing the centroid of the spectral trace (the "footprint" of the spectrum on the detector). We also estimated the orientation of the spectral trace relative to the detector edges, by performing a linear fit to the previously calculated function $y(x)$. Figure 4-2 shows the results. Within a single orbit, the position and orientation of the spectral trace are relatively constant, as compared to the larger movements that are observed between orbits. The largest variations of the spectral trace (its position and width) seem to coincide with the largest discontinuities in the flux time series.

Given the correlations that are observed between the properties of the spectral trace and the aperture-summed flux, the approach taken by Swain et al. (2008) and other investigators is to "decorrelate" the flux against a number of measured parameters ("state variables") such as the mean y position, cross-dispersion width, orientation angle, and so forth. One way to achieve this decorrelation is to fit linear functions of the state variables to the out-of-transit data, for which all time variations are expected to be due to the systematic effects. Then the best-fitting parameters are used to correct all of the data. Alternatively, one could fit for the linear functions of the state variables simultaneously with the parameters describing the transit light curve.

We attempted both of these procedures and found that while they do reduce the amplitude of the systematic effects, they still leave highly significant systematic variations. We also find this procedure to be undesirable because it is not clear which parameters to include in the fit; because the fitted parameters are highly correlated (the state variables do not vary independently); and because we have no justification for the assumption of a linear function for any of these parameters, without an understanding of the underlying physical effect. An example of a possibly relevant physical effect that would not necessarily be described by a linear function is intra-pixel sensitivity variation, which could lead to a function that is periodic in the pixel coordinates of the spectral trace.

We attempted to fit numerous physically-motivated models (based on the premise of intra-pixel sensitivity variations, among others), and did not find any such model that provided a good fit to the data while also having only a few, nondegenerate adjustable parameters. Ultimately we gave up on attempting to correct the intra-orbital discontinuities based on *a priori* information. Instead we included in our model an adjustable multiplicative factor specific to each orbit. This might seem devastating to the goal of analyzing the transit light curve, but this is not so. It means that we cannot make use of the relative flux between orbits to derive the transit depth; but as we will show, the truly precious information is in the duration of ingress or egress, which is much less vulnerable to the problem of flux discontinuities. In addition, four of the orbits spanned a full ingress or egress. Data from those four orbits does provide useful information about the transit depth, because the discontinuities appear between orbits and not within orbits.

All together, our model for the flux variation due to systematic effects is

$$f_{\rm sys}(t) = f_o^v \left\{1 + c_0^v \phi(t) + c_1^v [\phi(t)]^2 + c_2^v [\phi(t)]^3 \right\}, \tag{4.2}$$

where the v index specifies the visit number (1–4), the o index specifies the orbit number (1–4) within each visit after omitting the first orbit, the 16 numbers f_o^v are the multiplicative factors describing the inter-orbital discontinuities, $\phi(t)$ is the *HST* orbital phase at time t, and the 12 numbers c_i^v (3 per visit) are constants specifying a polynomial function of ϕ that describes the intra-orbital variation. The *HST* orbital phase was defined as

$$\phi(t) \equiv \frac{(t - \langle t \rangle) \mod P_{HST}}{P_{HST}}, \tag{4.3}$$

where $P_{HST} = 1.5975$ hours is the orbital period of the *HST* around the Earth and $\langle t \rangle$ is the midpoint of each orbit's observations. The choice of a polynomial, as opposed to some other smoothly varying function, was arbitrary. We also tried using sinusoidal functions with an angular frequency of $2\pi/P_{HST}$, with no significant differences in any of the results described below.

For the transit model, we used the analytic formulas of Mandel & Agol (2002). Our parameters were the planet-to-star radius ratio $(R_p/R_\star)$, the cosine of the orbital inclination $(\cos i)$, the semimajor axis in units of the stellar radius $(a/R_\star)$, the mid-transit time t_c^v, and the two coefficients u_1 and u_2 of a quadratic limb-darkening law,

$$\frac{I_\mu}{I_1} = 1 - u_1(1-\mu) - u_2(1-\mu)^2 \tag{4.4}$$

where μ is the cosine of the angle between the observer and the normal to the stellar

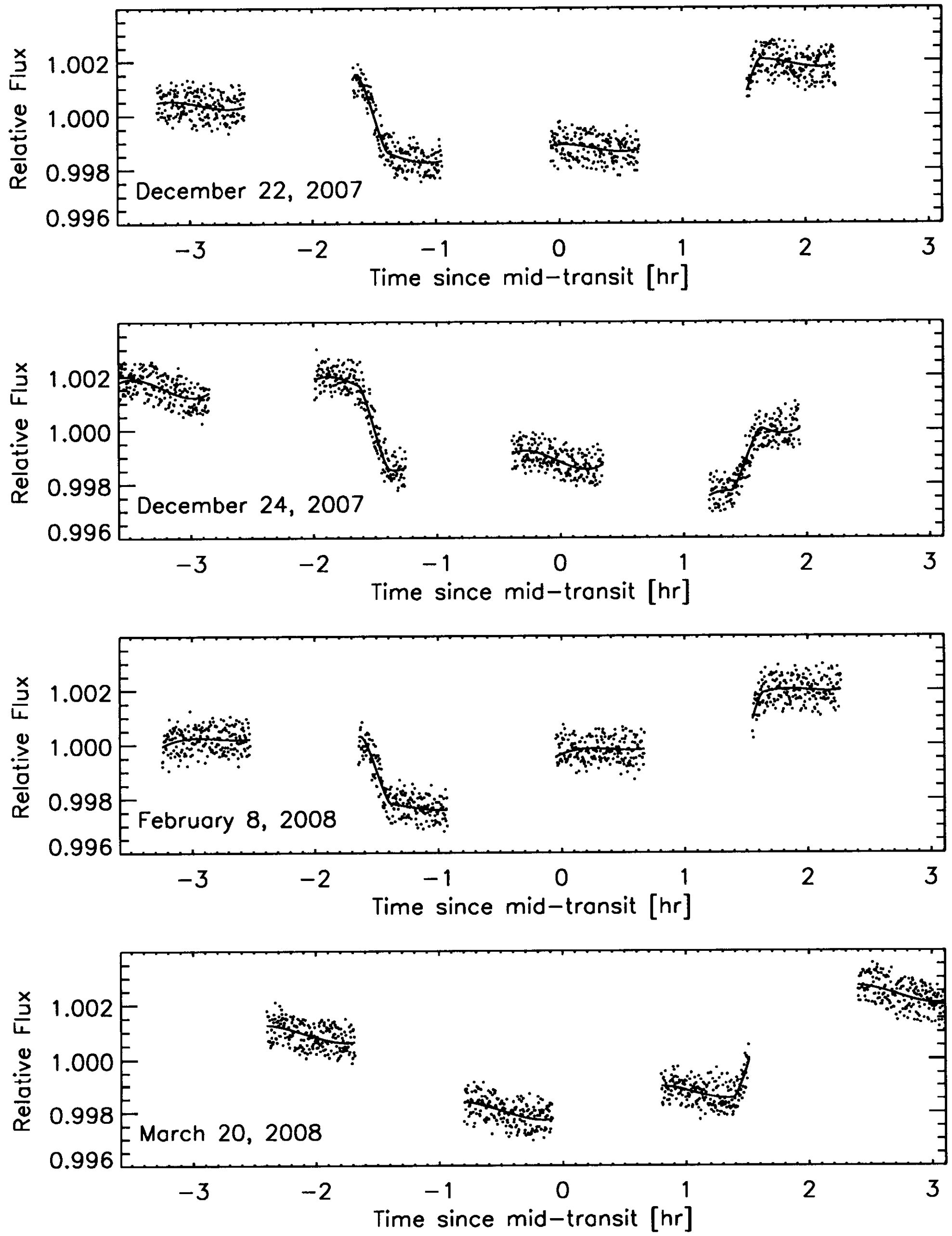

Figure 4-1 NICMOS photometry (1.1–2.0 μm) of HD 149026b of 4 transits, with interruptions due to Earth occultations. Plotted are the results of simple aperture photometry. The observed variations are a combination of the transit signal and systematic effects (intra-orbital variations and inter-orbital discontinuities). The solid curve is the best-fitting model that accounts for both the transit signal and systematic effects.

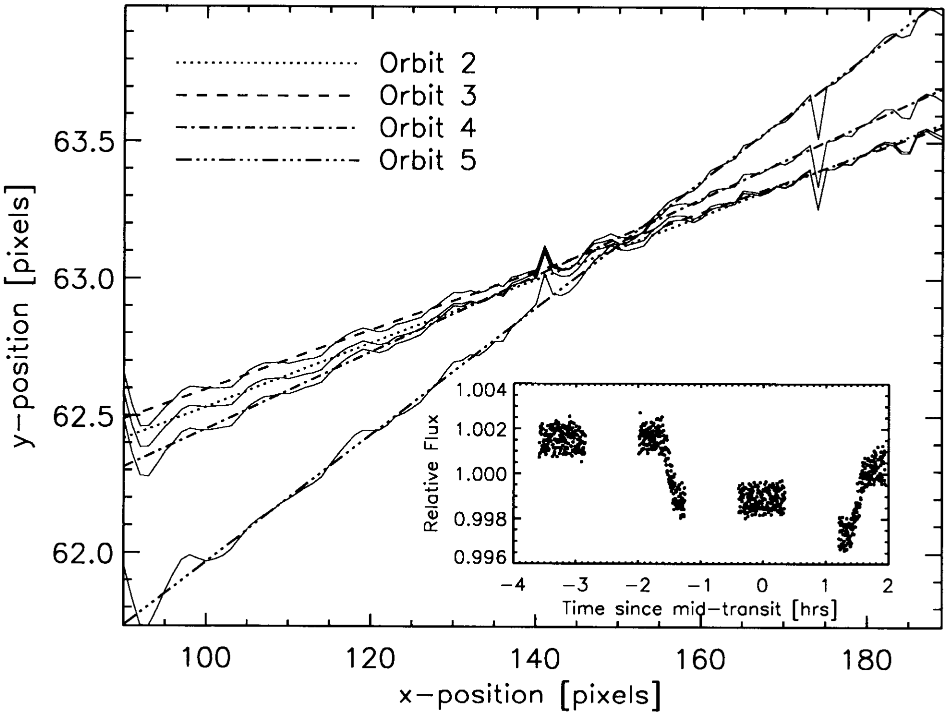

Figure 4-2 Illustration of inter-orbital variations of the spectral trace. The solid curves are the flux-weighted mean y position of the first-order spectrum as a function of x. Overplotted are linear fits to $y(x)$. The inset figure shows the measured light curve after dividing out the intra-orbital variations correlated with *HST* orbital phase. The largest rotation of the spectral trace (at the fifth orbit) coincides with the largest discontinuity in the flux time series.

surface and I_μ is the specific intensity as a function of μ. We allowed u_1 and u_2 to vary freely, subject to the conditions $u_1 + u_2 < 1$, $u_1 + u_2 > 0$, and $u_1 > 0$. These conditions require the brightness profile to be everywhere positive and monotonically decreasing from limb to center. In practice, the fitting parameters were actually

$$
\begin{aligned}
u_1' &= u_1 \cos 40^\circ - u_2 \sin 40^\circ && (4.5) \\
u_2' &= u_1 \sin 40^\circ + u_2 \cos 40^\circ && (4.6)
\end{aligned}
$$

because u_1' and u_2' are weakly correlated, unlike u_1 and u_2 (Pál 2008). In computing the transit light curve we assumed the orbit to be circular, consistent with the findings of Sato et al. (2005) and Madhusudhan & Winn (2008). We held the orbital period fixed at the value $P = 2.87588$ days based on the results of Winn et al. (2008b). Here the period is used only to relate the measured transit durations and $a/R_\star$. The fractional error in P is approximately 10^4 times smaller than the fractional error in $a/R_\star$ and is safely ignored (Carter et al. 2008).

Our complete model of the photometric time series was the product of the transit model and $f_{\rm sys}(t)$ (Eq. 4.2). We fitted simultaneously for the parameter set $R_p/R_\star$, $\cos i$, $a/R_\star$, u_1', u_2', $\{t_c^v\}$, $\{f_o^v\}$, and $\{c_i^v\}$). The polynomial describing intra-orbital variations was specific to each visit, and the flux discontinuities were specific to each orbit, but the transit parameters were required to be consistent across all orbits and visits. We performed a least-squares fit to the unbinned data using a box-constrained Levenberg-Marquardt algorithm (Levenberg 1944, Marquardt 1963, Lourakis 2004) utilizing the Jacobian calculation of Pál (2008). Box constraints were needed to enforce the restrictions on the limb darkening parameters u_1 and u_2. The goodness-of-fit statistic was

$$
\chi^2 = \sum_{v=1}^{4} \sum_{i=1}^{N_v} \left(\frac{f_{\rm obs}^v(t_i) - f_{\rm calc}^v(t_i)}{\sigma_v} \right)^2 \qquad (4.7)
$$

where $f_{mod}^v(i)$ the calculated flux at the time of the i^{th} data point during visit v, $f_{obs}^v(i)$ is the i^{th} flux measurement during visit v, N_v is the number of data points in visit v, and σ_v was assumed to be a constant at this step. The solid curve in Fig. 4-1 shows the best-fitting model. The root-mean-square (rms) residual between the data and the best-fitting model is 440 parts per million (ppm). This is approximately 2.2 times the expected noise level calculated within the NICMOS calibration pipeline (which is dominated by photon noise). Fig. 4-3 shows histograms of the residuals for all the data and for each visit individually. The residuals are not Gaussian; the flattened peak in the histograms indicates a nonzero kurtosis[2].

Evidently the noise is not photon-limited, and is not Gaussian, but at least it does not appear to be strongly correlated in time. We assessed the degree of temporal correlations ("red noise") in two ways. First, we binned the residuals in time by a factor N ranging from 1 to 100, and calculated the standard deviation σ_N of the binned residuals. The results are shown in Fig. 4-4. They follow closely the expectation of independent random numbers, $\sigma_N = \sigma_1 \; N^{-1/2} \; [M/(M-1)]^{1/2}$, where M is the number of bins. Second, we calculated the Allan (1964) variance $\sigma_A^2(l)$ of the residuals, defined as

$$\sigma_A^2(l) = \frac{1}{2(N+1-2l)} \sum_{i=0}^{N-2l} \left(\frac{1}{l} \sum_{j=0}^{l-1} r_{i+j} - r_{i+j+l} \right)^2 \tag{4.8}$$

where r_k denotes the residual of the kth data point, N is the number of data points, and l is the lag. The Allan variance is commonly used in the time metrology literature to assess $1/f$ noise. For independent residuals, one expects $\sigma_A^2(l) \approx \sigma_A^2(0)/l$. The results for our data, also shown in Fig. 4-4, satisfy this expectation. There is no readily identifiable time-correlated component in the time series of residuals.

Plotted in Fig. 4-5 is the measured flux after dividing by the optimized function

---

[2] One may wonder about the effect of the apparent non-Gaussianity of the noise, shown in Fig. 3. To investigate this issue we used an Edgeworth expansion to create a new χ^2-like statistic that accounts for the skewness and kurtosis of the residuals (see, e.g., Amendola et al. 1996). Using this different fitting statistic, we found that the best-fit parameter values were unchanged. This was not surprising because bias is expected to arise from skewness (as opposed to kurtosis) and the skewness of the residuals is very small. However, the confidence intervals are affected by the kurtosis. We found that accounting for kurtosis leads to error bars that are *smaller* than the error bars quoted here, but only by a modest amount ($\lesssim 20\%$). For simplicity, the results quoted in this chapter are based on the standard χ^2 statistic given in Eq. (7).

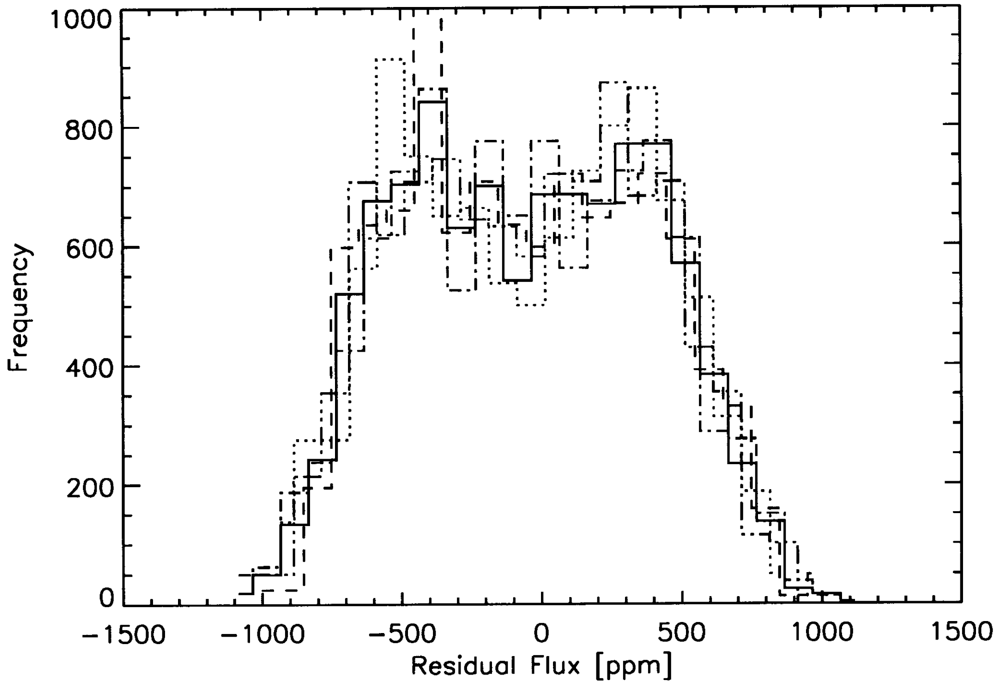

Figure 4-3 Histograms of the residuals between the data and the best-fitting model. The solid line is the histogram based on all the data. The other lines are for data specific to visit 1 (dotted), visit 2 (dashed), visit 3 (dash-dot), and visit 4 (dash-dot-dot-dot).

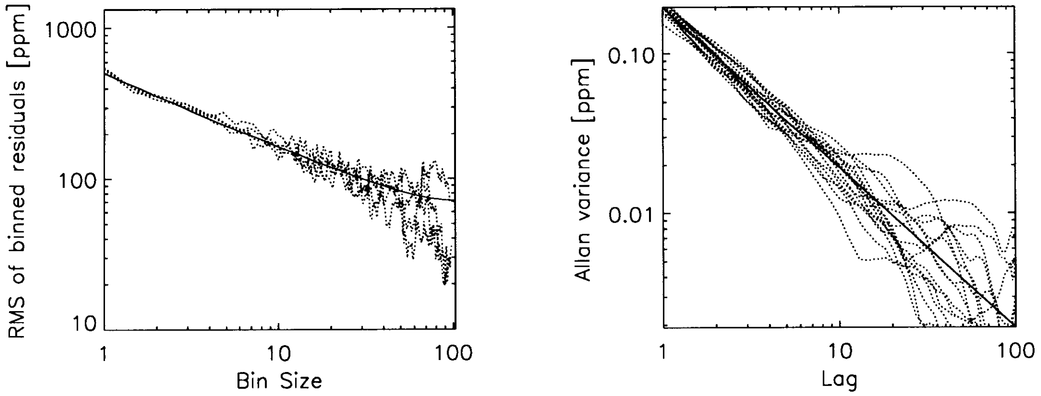

Figure 4-4 Assessment of correlated noise. Left panel: The rms of time-binned residuals, as a function of bin size, for each of the 4 visits. Right panel: The Allan variance of the residuals, as a function of lag, for each of the 16 orbits. The dotted lines are the results of the calculations based on the data, and the solid lines show the expected trend if the noise were uncorrelated.

$f_{\rm sys}(t)$. This represents our best effort to correct for the systematic effects. In Fig. 4-6, we show the results of combining the data from all the visits (after correcting for systematic effects) into a single transit light curve. In this composite light curve, the median time between samples is 7.2 s. Finally, in Fig. 4-7 we show a time-binned version of the composite light curve to allow a visual comparison with the best previously-measured light curves, at optical and mid-infrared wavelengths.

To determine the "allowed range" of each parameter—or, more precisely, the *a posteriori* joint probability distribution of all the parameter values—we employed the Markov chain Monte Carlo (MCMC) technique (see, e.g., Winn et al. 2007; Burke et al. 2007). We produced 8 chains of length 7.4×10^6 using a Gibbs sampler and a Metropolis-Hastings jump condition such that each parameter had an effective chain length of roughly 2×10^5. This was accomplished by adjusting the scale of the jump-function distribution such that the probability of accepting a jump is roughly uniform and equal to approximately 40% across all parameters. To establish initial estimates of parameter uncertainties, a preliminary Monte Carlo bootstrap analysis was performed, based on the Levenberg-Marquardt least squares minimization; then, the MCMC initial conditions were drawn from normal distributions with widths equal to five times these initial error estimates. The first 25% of each chain was trimmed, and then all the chains were concatenated. Each parameter had a Gelman & Rubin (1992) R statistic smaller than 1.01, a sign of good convergence of the posterior parameter distributions. For each parameter, the values at each link of the chain were sorted. To describe the results, we report the median (50%) value, along with the interval between the 15.85% and 84.15% levels (the 68.3% confidence interval). The results are given in Column 2 of Table (4.1).

4.3.1 Results from NICMOS photometric analysis

For the orbital inclination, we find $i = 84.55^{+0.35}_{-0.81}$ deg. This is in agreement with the independent estimate of $i = 85.4^{+0.9}_{-0.8}$ deg by Nutzman et al. (2008), using the 8 μm channel of the Infrared Array Camera (IRAC) aboard the *Spitzer Space Telescope*. For the normalized orbital distance, we find $a/R_\star = 6.01^{+0.17}_{-0.23}$. This is also in agreement

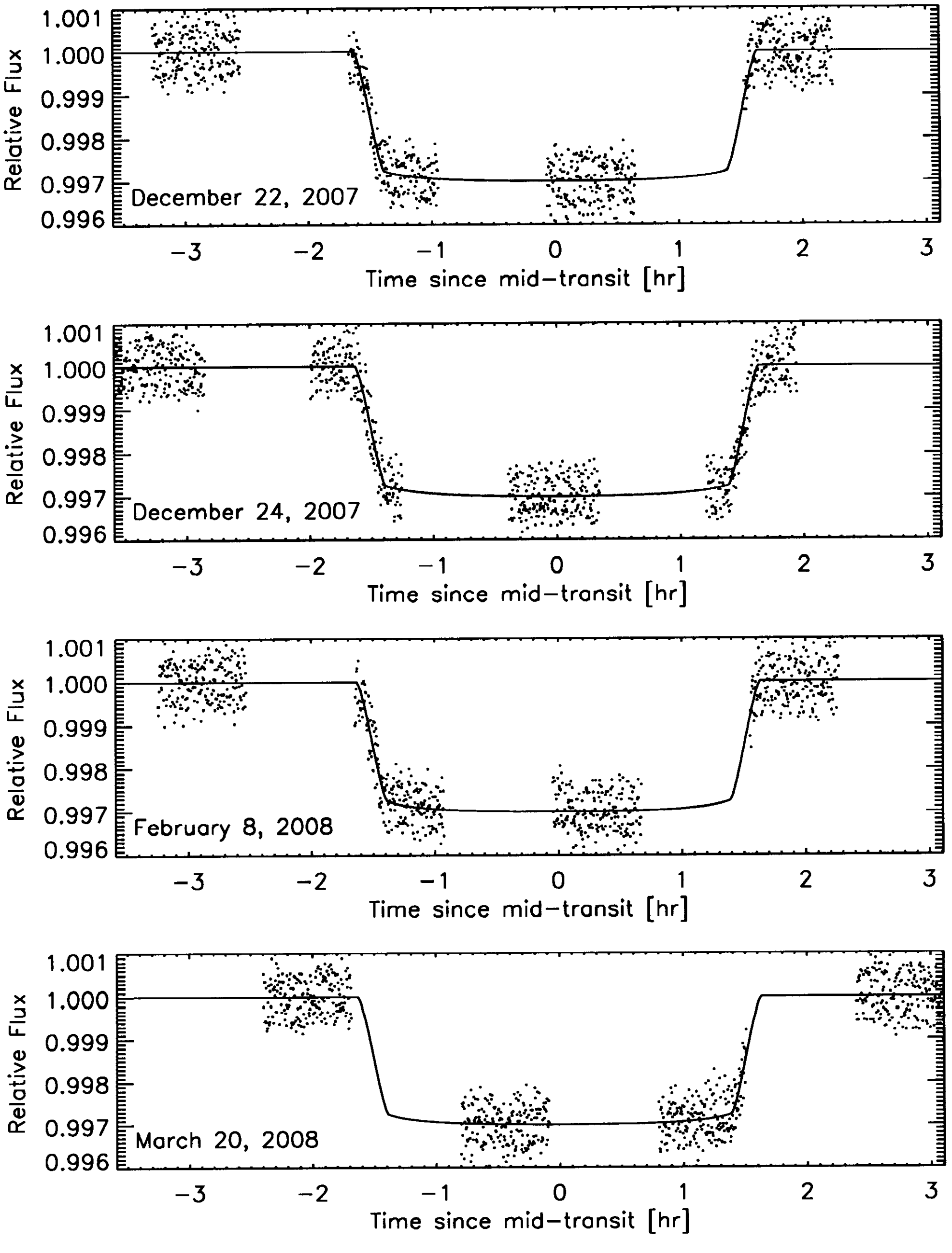

Figure 4-5 NICMOS photometry (1.1–2.0 μm) of 4 transits of HD 149026b, after correcting for systematic effects. In each panel, the solid line shows the best-fitting model.

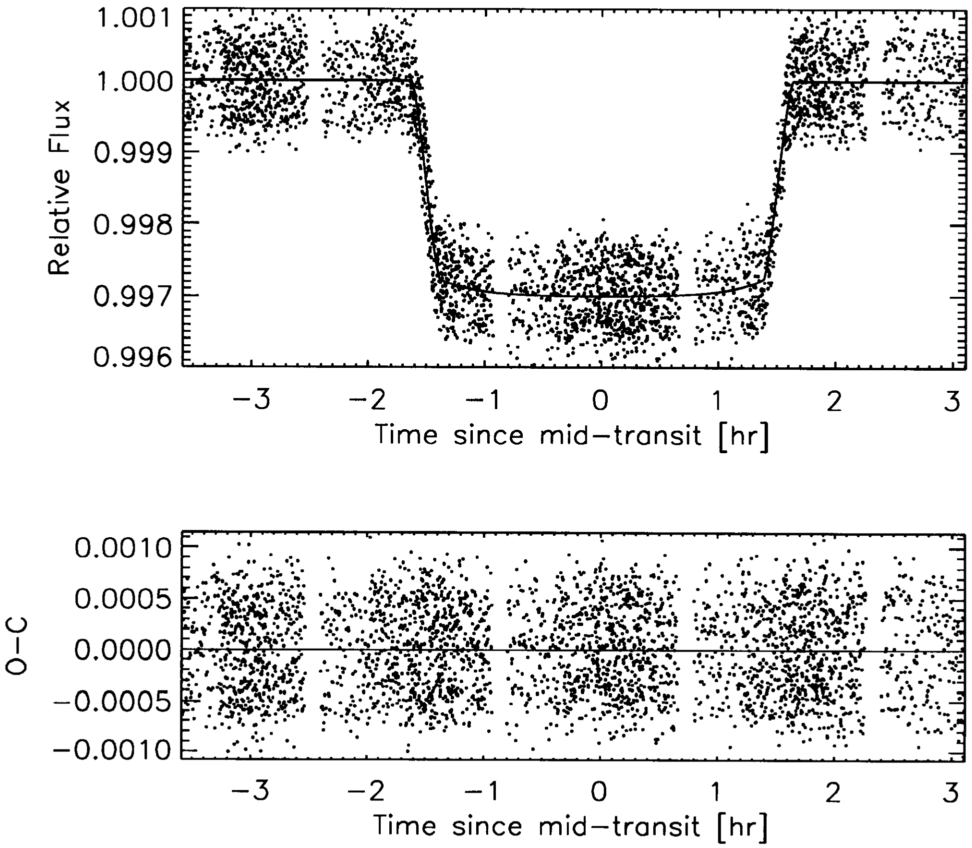

Figure 4-6 NICMOS transit light curve (1.1–2.0 μm) of HD 149026b. The data from 4 transits have been superimposed, after correcting for systematic effects. The solid line shows the best-fitting model.

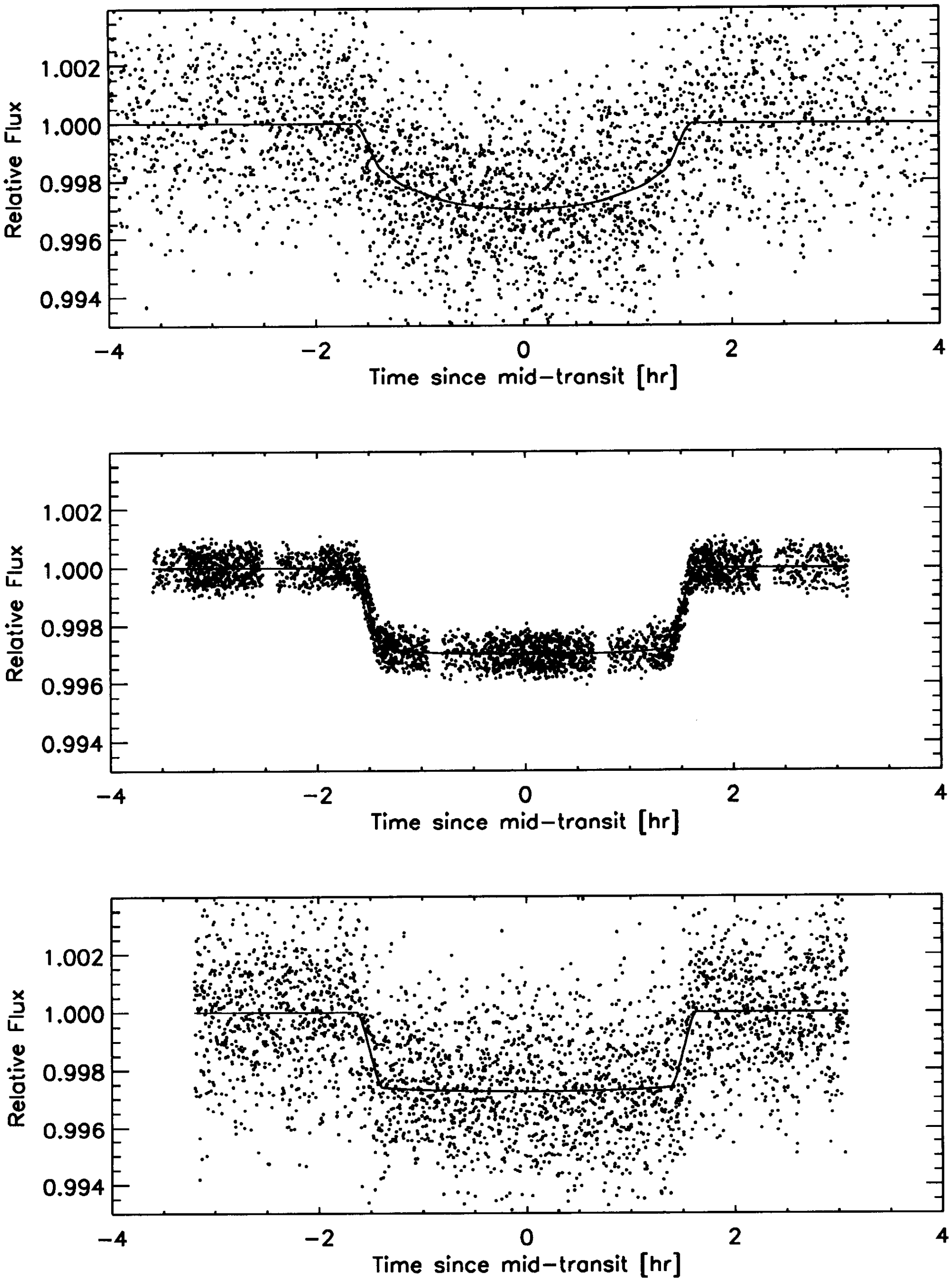

Figure 4-7 Comparison of the best available transit light curves of HD 149026. Top panel: optical [Stromgren $(b+y)/2$] photometry from Sato et al. (2005) and Winn et al. (2008b), with a time sampling of 8.6 s and an rms residual of 2017 ppm. Middle panel: near-infrared [1.1–2.0 μm] photometry from this work, with a time sampling of 7.2 s and an rms residual of 440 ppm. Bottom panel: mid-infrared [8 μm] photometry from Nutzman et al. (2008), with a time sampling of 7.4 s and an rms residual of 1854 ppm. The solid lines show the best-fitting model, which was calculated as described in § 4.5.

with the results of Nutzman et al. (2008), who found $a/R_\star = 6.20^{+0.28}_{-0.63}$. The new result is more precise, which (as we will show in § 4.4) leads to tighter constraints on the stellar mass and radius. This is important because the uncertainties in the stellar properties have been the limiting factors in the analysis of this system. Based on the preceding results, we find the impact parameter (defined as $b = a\cos i/R_\star$) to be $0.571^{+0.044}_{-0.038}$. This is the tightest such constraint that has been achieved for HD 149026b. The earliest measurements of the impact parameter were consistent with zero (Sato et al. 2005, Charbonneau et al. 2006, Winn et al. 2008b), a situation leading to strong degeneracies among the transit parameters (Carter et al. 2008). More recently, Nutzman et al. (2008) found $b = 0.62^{+0.08}_{-0.24}$, consistent with the new and more precise result. The increased precision in $a/R_\star$ and b is a direct consequence of the improved precision with which the ingress (and egress) duration is known (Carter et al. 2008). In this sense, the greatest value of the NICMOS data is in the good coverage of the ingress and egress phases.

The enhanced precision of the NICMOS data relative to previous data sets does not lead to a correspondingly enhanced precision in the planet-to-star radius ratio. This is because we allowed the time series from each orbit to have its own adjustable flux multiplier. Consequently, all of the information about $R_p/R_\star$ comes only from those orbits that span an entire ingress or egress event. However, the value of $R_p/R_\star$ that we derive is at least comparable in precision to previous determinations. We find $R_p/R_\star = 0.05416^{+0.00091}_{-0.00070}$. Interestingly this is larger by 2σ than the values derived previously, which were based on optical and mid-infrared data. Winn et al. (2008b) found $R_p/R_\star = 0.0491^{+0.0018}_{-0.0005}$ based on Stromgren $(b+y)/2$ photometry, and Nutzman et al. (2008) found $R_p/R_\star = 0.05158 \pm 0.00077$ based on 8 μm photometry.

Since we do not understand all of our noise sources with a physically-grounded model, we cannot rule out the possibility that the discrepancy between our result and the previous results is due to a faulty model of the systematic effects. The culprit would probably need to be the polynomial function of orbital phase. We do find that our result is unaffected if we replace the polynomial function of ϕ with trigonometric functions, as mentioned previously; and we also find that our results are unchanged if we use a linear limb darkening law or fix the quadratic limb darkening parameters to those tabulated by Claret (2000). These tests do not prove that our results are valid but they do suggest that our analysis procedure is robust to changes in the functional form of the model. However, to the extent that the intra-orbital variations are not strictly repeatable within a given visit, our model would produce biased results. Fig. 4-8 shows the data after dividing by the optimized values of f_o^v (the orbit-specific flux multipliers) and dividing by the optimized transit model. The purpose of the divisions is to isolate the intra-orbital systematic effects, which do appear consistent among the orbits within a given visit.

Another possibility for the discrepancy in the transit depth between our near-infrared result and the previous optical and mid-infrared results is differential absorption due to constituents in the outer, optically-thin portion of the planet's atmosphere. This interpretation is the basis of the "transmission spectroscopy" technique for identifying constituents of exoplanetary atmospheres pioneered by Charbonneau et al. (2002). We consider this possibility at some length in sections 4.5 and 4.7.

We find that the center-to-limb variation is less pronounced than was expected based on the tabulated limb darkening coefficients of Claret (2000). Fig. 4-9 shows the confidence contours in the u_1, u_2 parameter space. The open square corresponds to the tabulated values for H band (for a star with $T_{\rm eff} = 6250~K$, $\log g_\star = 4.5$, $[\mathrm{Fe/H}] = 0.3$, matching the properties of HD 149026). The tabulated values are excluded with >95% confidence.

Two quantities intrinsic to the star and planet that may be calculated directly in terms of observables are the surface gravity of the planet, and the mean density of the star. The surface gravity of the planet is calculated as (Southworth et al. 2007)

$$g_p = \frac{2\pi}{P}\frac{K}{(R_p/a)^2 \sin i}, \tag{4.9}$$

where K is the semiamplitude of the stellar radial-velocity variation (43.3 ± 1.2 m s^{-1}; Sato et al. 2005), and R_p/a and $\sin i$ are derived from our MCMC analysis. We find $\log g_p = 3.132^{+0.029}_{-0.035}$ where g_p is in cgs units. The mean stellar density $\rho_\star$ is calculated

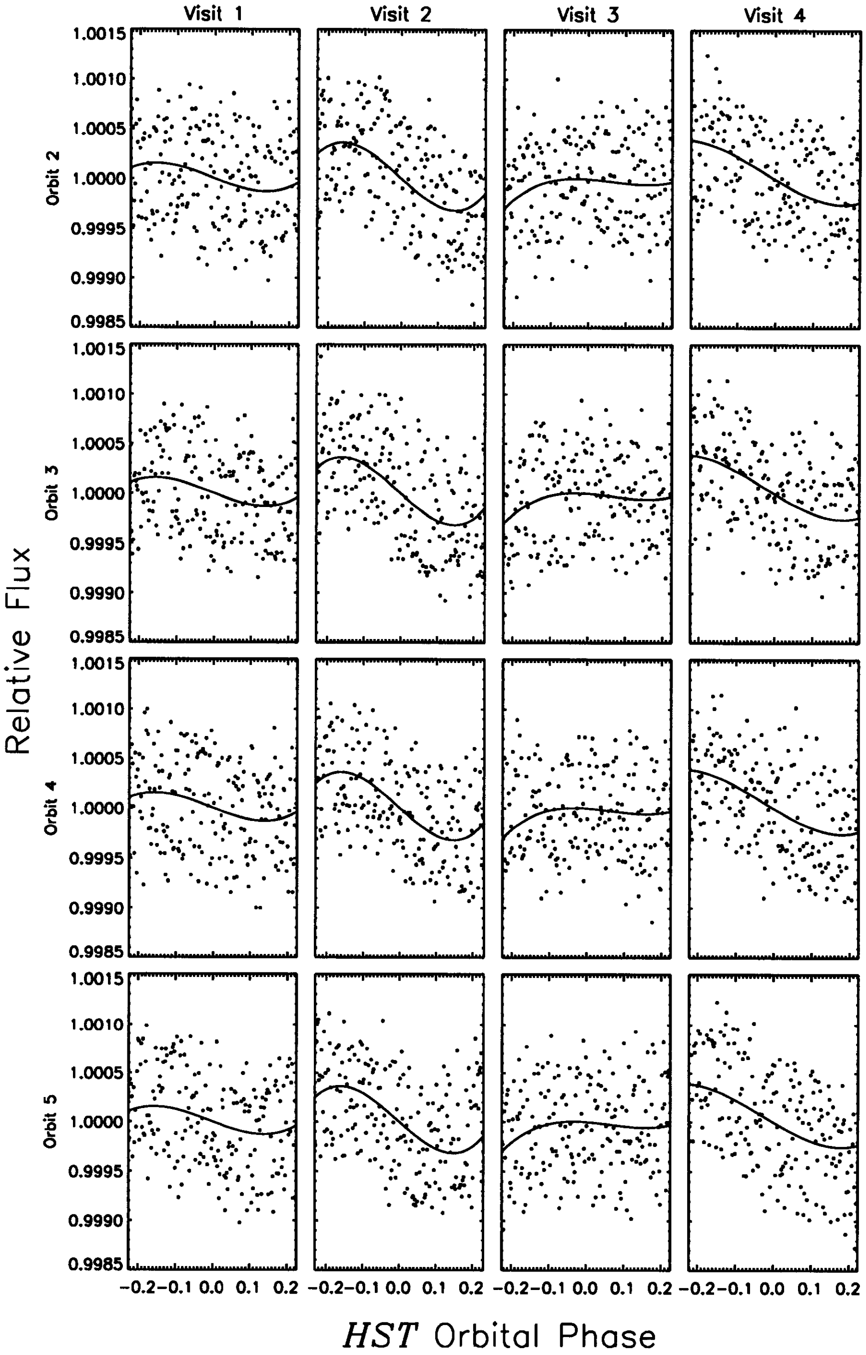

Figure 4-8 Isolation of the intra-orbital variations. The flux time series has been divided by the optimized flux multipliers f_o^v and by the optimized transit model. The remaining variation appears to present a consistent pattern among all orbits within a given visit, as assumed in our model. The solid line is the optimized model. Each column shows data from all orbits of a given visit. Each row shows orbits arranged from first to last in rows from top to bottom, respectively.

as (Seager & Mallen-Ornelas 2003, Sozzetti et al. 2007)

$$\rho_\star = \frac{3\pi}{GP^2}\left(\frac{a}{R_\star}\right)^3 - \rho_p\left(\frac{R_p}{R_\star}\right)^3 \tag{4.10}$$

where ρ_p is the mean density of the exoplanet. We may neglect the correction term involving the planetary density as $\rho_\star \sim \rho_p$, $(R_p/R_\star)^3 \sim 1 \times 10^{-4}$ and the fractional uncertainty in $a/R_\star$ is larger than the fractional uncertainty in $R_p/R_\star$ by a factor of 2.

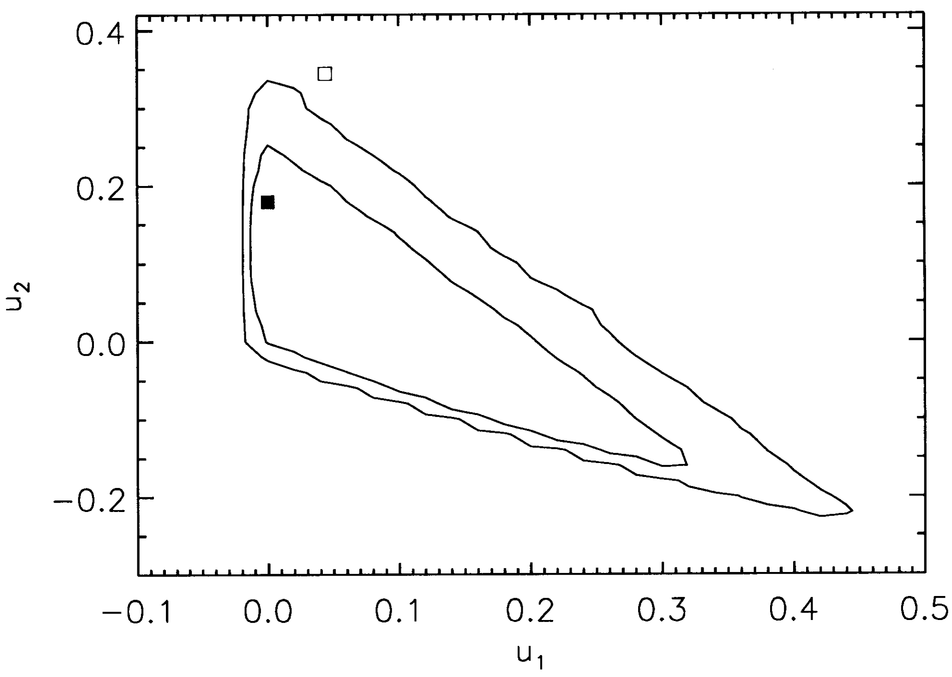

Figure 4-9 Results for the limb-darkening parameters u_1 and u_2. The contours are the 68% and 95% confidence regions as determined by the MCMC analysis of the photometric time series. The solid square is the minimum-χ^2 solution, $u_1 = 0.0$, $u_2 = 0.1789$. The open square marks the tabulated values of Claret (2000) for H band ($u_1 = 0.044$, $u_2 = 0.344$).

4.4 Stellar Parameters

The basic inputs to models of the planetary interior are the planetary mass M_p and radius R_p, in units of grams and kilometers, or in units of Jupiter's mass and radius.

Transit photometry and Doppler velocimetry alone do not determine these quantities. Additional information about the star must be introduced. Several techniques for estimating the stellar mass $M_\star$ and radius $R_\star$ were reviewed by Winn et al. (2008b). We chose to estimate $M_\star$ and $R_\star$ using stellar-evolution models that are constrained by the best available, relevant, observable properties of the star: the mean density $0.497^{+0.042}_{-0.057}$ g cm^{-3} determined from our light-curve analysis, the absolute magnitude $M_V = 3.65 \pm 0.12$ derived from the *Hipparcos* parallax and apparent magnitude [$\pi = 12.59 \pm 0.70$ mas, $V = 8.15 \pm 0.02$; van Leeuwen (2007)], effective temperature [$T_{\rm eff} = 6160 \pm 50$ K, a weighted mean of the results from Sato et al. (2005) and Masana et al. (2006)], and metallicity [0.36 ± 0.08, from Sato et al. (2005) with a more conservative error bar]. We chose not to use the spectroscopically-determined stellar surface gravity [$\log g_\star = 4.26 \pm 0.07$; Sato et al. (2005)] because the photometrically-determined value of $\rho_\star$ provides an effectively tighter constraint, and because spectroscopically-determined surface gravities have been found to be susceptible to systematic error (see, e.g., Winn et al. 2008a, Johns-Krull et al. (2008)).

Following the procedure of Torres, Winn, & Holman (2008), we employed Yonsei-Yale stellar models[3] (Yi et al. 2001, Demarque et al. 2004). Model isochrones were interpolated in both age and metallicity, for metallicities [Fe/H] ranging from 0.28 to 0.43 and for ages ranging from 0.1 to 14 Gyr, in steps of 0.1 Gyr. Fig. 4-10 shows several of these theoretical isochrones, along with some of the observational constraints. The upper left panel illustrates the constraint due to the spectroscopically-determined surface gravity, even though we did not actually apply that constraint, as explained above. It is evident that the constraint due to $\rho_\star$ is stronger.

The isochrones were interpolated to provide a fine grid in stellar mass (with a step size of 0.005 $M_\odot$). We then assumed that the likelihood of each point on the

---

[3] We chose the Y^2 models mainly because of the convenient form in which they are publicly available. Other stellar-evolutionary models are available, and other investigators have examined the sensitivity of results such as ours to the choice of model. For HD 149026 in particular, Southworth (2008) found that the Y^2 models and independent models by Claret (2007) gave results for $M_\star$ and $R_\star$ that agreed to within 1%. Another set of publicly available models, the Padova models of Girardi et al. (2000), are not computed for the high metallicity observed for HD 149026; but at zero metallicity, at least, both Torres et al. (2008) and Southworth (2008) found that the Padova models give results for the stellar mass and radius that are also within 1% of the Y^2 results.

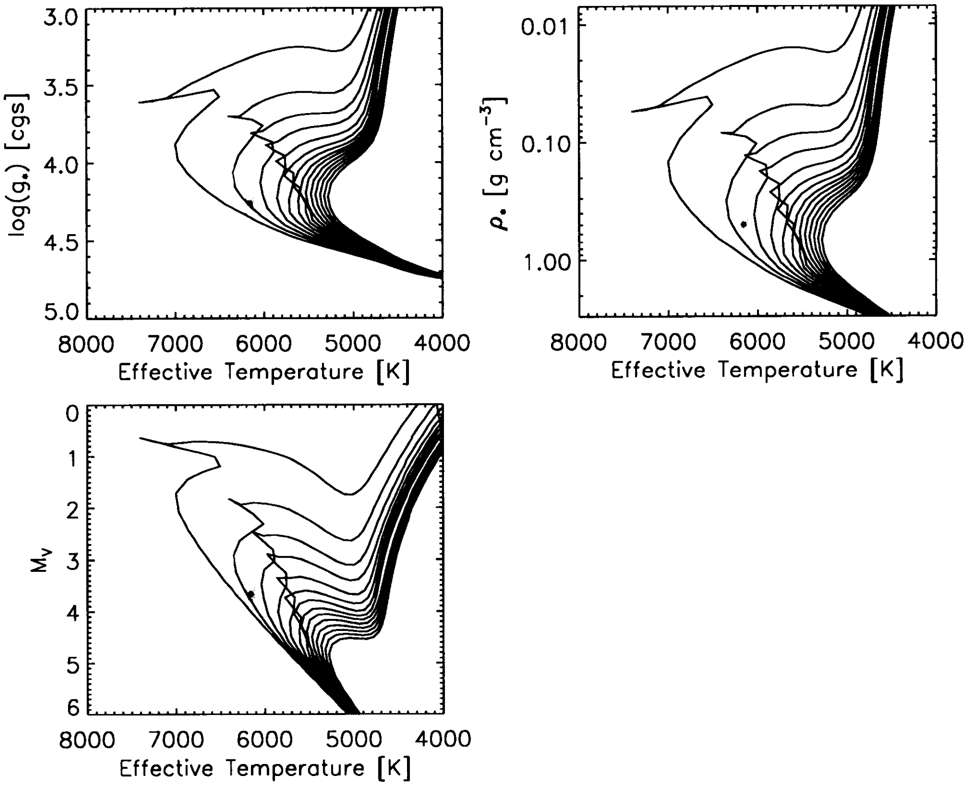

Figure 4-10 Stellar-evolutionary model isochrones, from the Yonsei-Yale series by Yi et al. (2001). The points and shaded boxes represent the observationally-determined values and 1σ errors. Here, surface gravity is determined spectroscopically (Sato et al. 2005), M_V is derived from *Hipparcos* parallax and V magnitudes, and $\rho_\star$ is determined photometrically from the transit light curve. Isochrones are shown for ages of 1 to 13 Gyr (from left to right) in steps of 1 Gyr for a fixed stellar metallicity of [Fe/H]= 0.36.

interpolated isochrones was proportional to $\exp(-\chi_\star^2/2)$, where

$$\chi_*^2 = \left(\frac{\Delta[\mathrm{Fe/H}]}{\sigma_{[\mathrm{Fe/H}]}}\right)^2 + \left(\frac{\Delta T_{\mathrm{eff}}}{\sigma_{T_{\mathrm{eff}}}}\right)^2 + \left(\frac{\Delta M_V}{\sigma_{M_V}}\right)^2 + \left(\frac{\Delta \rho_\star}{\sigma_{\rho_\star}}\right)^2, \tag{4.11}$$

and the Δ quantities denote the differences between the observed and calculated values. The asymmetry in the error distribution for $\rho_\star$ was taken into account. Additionally, the likelihood was taken to be proportional to an Salpeter initial mass function, $\xi(M) \propto M^{-(1+x)}$ with $x = 1.35$ (Salpeter 1955). The joint probability function, P, was taken to be proportional to the likelihood, viz.,

$$P(R_\star, M_\star, T_{\mathrm{eff}}, \log g_\star, M_V, [\mathrm{Fe/H}], \rho_\star, \mathrm{Age}) \propto \exp(-\chi_*^2/2). \tag{4.12}$$

For a given parameter X_0 from this list, we calculated the cumulative distribution function (CDF) by numerically evaluating

$$D_{X_0}(x) = \int_{-\infty}^{x} dX_0 \int_{-\infty}^{\infty} dX_1 \cdots \int_{-\infty}^{\infty} dX_N \;\; P(X_0, X_1, \cdots, X_N). \tag{4.13}$$

For each parameter, we record the values of x for which the CDF takes the values 15.85%, 50%, and 84.15%. The 50% level (the median) is reported as the "best-fit value" and the interval between the 84.15% and 15.85% levels is reported as the 68.3% (1σ) confidence interval.[4]

Table (4.1) reports the best-fit stellar parameters and confidence intervals. We find the stellar radius to be $R_\star = 1.541^{+0.046}_{-0.042}\ R_\odot$. This is larger than (but in agreement with) the previous estimates of $R_\star = 1.46 \pm 0.10\ R_\odot$ by Sato et al. (2005), and $R_\star =$

---

[4] Although our procedure was inspired by the work of Torres et al. (2008) and is similar in almost all respects, there is one significant difference. The best fit values reported by Torres et al. (2008) were those that minimized χ^2 as in our analysis. The difference is that Torres et al. (2008) estimated the 1σ errors in the stellar properties based on the total span of the calculated values that gave agreement within 1σ with the observables. Effectively, they assumed a uniform error distribution for each observable, rather than a Gaussian error distribution as we have done. Consequently, our method produces smaller error intervals in the stellar properties. Caution would dictate that larger error intervals are desirable, especially since we are relying on the theoretical isochrones that surely have some unaccounted-for systematic errors. However, using a uniform error distribution for the observables is an arbitrary way to inflate the output errors, and the true error distribution for the observables is probably closer to Gaussian. For these reasons we chose our approach and emphasize the caveat that our results place complete trust in the Y^2 isochrones.

$1.497 \pm 0.069\ R_\odot$) by Nutzman et al. (2008). By combining the derived distribution for $R_\star$ with the photometrically-determined distribution for $R_p/R_\star$, we find the planetary radius to be $R_p = 0.813^{+0.027}_{-0.025}\ R_{\rm Jup}$. This is larger than any previous result. Using optical photometry, Sato et al. (2005) found $0.725 \pm 0.050\ R_{\rm Jup}$, Charbonneau et al. (2006) found $0.726 \pm 0.064\ R_{\rm Jup}$, and Winn et al. (2008b) found $0.71 \pm 0.05\ R_{\rm Jup}$. Using mid-infrared photometry, Nutzman et al. (2008) found $0.755 \pm 0.040\ R_{\rm Jup}$. It is important to note that these determinations were not wholly independent, and therefore should not be combined into a weighted average. They all used many common inputs for the stellar properties, and the analyses of optical photometry all included a common subset of at least 3 light curves.

As mentioned previously, we did not apply any constraint to the models based on the spectroscopically determined value of $\log g_\star$. However, given our results for $M_\star$ and $R_\star$ we computed the implied value of $\log g_\star$, finding $\log g_\star = 4.189^{+0.020}_{-0.021}$. This in agreement with, and is more precise than, the spectroscopically-determined value of $\log g_\star = 4.26 \pm 0.07$ (Sato et al. 2005).

4.5 Joint Analysis with Optical and Mid-Infrared Light Curves

Transit observations of HD 149026b have now been made at optical wavelengths (Sato et al. 2005, Charbonneau et al. 2006, Winn et al. 2008b), near-infrared wavelengths (this work), and mid-infrared wavelengths (Nutzman et al. 2008). In this section we repeat our analysis on all of these data, in order to bring all of these data to bear on the determination of the system parameters, while seeking possible wavelength variations in the planet-to-star radius ratio.

It is reasonable to require consistency across these data in the parameters relating to the orbital configuration of the transit, such as the inclination angle and normalized semi-major axis. However, the inferred planet-to-star radius ratio is a wavelength-dependent quantity, depending on the opacity of the exoplanetary atmosphere and the emergent flux from the planetary nightside (which is expected to be unimportant). With this in mind, we performed a joint analysis of all of the data, requiring consistency in $\cos i$ and $a/R_\star$ but allowing $R_p/R_\star$ to take separate values for each of the three types of data: optical, near-infrared, and mid-infrared.

Specifically, we fitted our NICMOS data, the 8 μm IRAC time series of Nutzman et al. (2008), and the 8 light curves obtained in the Strömgren $(b+y)/2$ band by Sato et al. (2005) and Winn et al. (2008b). Our photometric model for the NICMOS data, including the associated systematic effects, has already been described. For the data sets presented by other authors, we followed those authors' prescriptions to account for systematic errors. For the $(b+y)/2$ data, we corrected the data by allowing the out-of-transit flux to be a linear function of time. For the 8 μm data, we modeled the time-variable sensitivity of the detector (the "ramp") as a multiplicative correction, $f_{\rm sys} = a_0 + a_1 \log(t - t_0) + a_2 \log^2(t - t_0)$, where t_0 is the time immediately prior to the start of the observation.

We performed an MCMC analysis of this joint data set. The free parameters relating to the transit model were the three values of $R_p/R_\star$ (corresponding to the ratios measured at approximately 0.5 μm, 1.5 μm, and 8.0 μm); the geometric parameters $\cos i$ and $a/R_\star$; the quadratic limb-darkening coefficients for the NICMOS light curve; the linear limb-darkening coefficients for the optical and infrared light curves (for which the precision of the data do not justify the more accurate quadratic law); and the mid-transit times for the NICMOS data and the IRAC data.[5] We also fitted for the ramp-correction terms for the IRAC data and the parameters relating to the flux offsets and intra-orbital variations for the NICMOS data. Six chains of length 9×10^6 were created, representing approximately 2×10^5 correlation-lengths per parameter. These were concatenated after removing the first 25% of each chain. The Gelman-Rubin R statistic was smaller than 1.01 for each parameter. We then repeated the analyses that were described in § 4.3 and § 4.4 to determine the stellar, planetary, and orbital parameters, based on this joint analysis. The results are

---

[5] To keep the number of parameters as small as possible, we did not vary the optical mid-transit times or baseline correction parameters at this stage, having found that they are uncorrelated with the other parameters of interest.

tabulated in Table (4.1).

The results for the geometric parameters are hardly changed from the NICMOS-only analysis, a reflection of the greater precision of the NICMOS light curve. The planet-to-star radius ratio was found to be larger for the NICMOS data than for the other bandpasses, as was already evident from the comparison of our NICMOS-only analysis to previously published analyses. Fig. 4-11 shows the variation in $(R_p/R_\star)^2$ with wavelength. The quantity $(R_p/R_\star)^2$ is essentially the transit depth, or fractional loss of light during the total phase of the transit, after "removing" the effects of limb darkening. The radius ratios that we derive for the mid-infrared and optical data are in agreement with those reported previously. For the IRAC data we find $R_p/R_\star = 0.05188^{+0.00084}_{-0.00086}$ as compared to the value 0.05158 ± 0.00077 reported by Nutzman et al. (2008). For the $(b+y)/2$ data we find $R_p/R_\star = 0.05070^{+0.00058}_{-0.00088}$ as compared to the value $0.0491^{+0.0018}_{-0.0005}$ found by Winn et al. (2008b). The precision in the optical $R_p/R_\star$ has been increased because the NICMOS data pins down all of the other parameters that are correlated with $R_p/R_\star$.

4.6 Ephemeris and transit timing

The NICMOS-only analysis resulted in the measurement of four distinct mid-transit times, with uncertainties smaller than 45 s. These are given in Table (4.2). We pooled together all of the independent measurements of mid-transit times from Winn et al. (2008b), Nutzman et al. (2008), and this work, to derive a new transit ephemeris. We fitted the times to a linear function of the integral epoch E,

$$T_c(E) = T_c(0) + EP \tag{4.14}$$

where P is the period and $T_c(0)$ is the mid-transit time at some fiducial epoch. The results were $T_c(0) = 2454456.78761 \pm 0.00014$ HJD and $P = 2.8758911 \pm 0.0000025$ days. The linear fit had $\chi^2 = 20.16$ and 14 degrees of freedom. This is a marginally unacceptable fit. The formal probability to find a value of χ^2 this large is 15%.

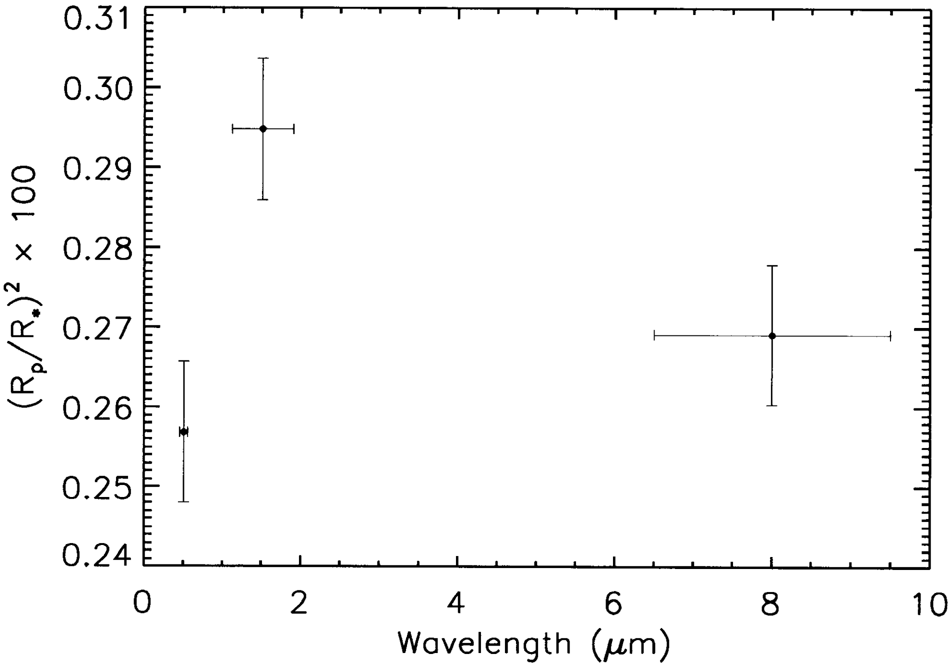

Figure 4-11 The planet-to-star area ratio, $(R_p/R_\star)^2$, as a function of observing wavelength, based on a joint fit to the NICMOS data, the $(b+y)/2$ data of Winn et al. (2008), and the IRAC data of Nutzman et al. (2008). The horizontal error bars show the approximate wavelength range of each bandpass.

Further transit observations are needed to distinguish the possibilities of a genuine period variation, a statistical fluke, and underestimated timing errors. Fig. 4-12 shows the O–C (observed minus calculated) timing diagram.

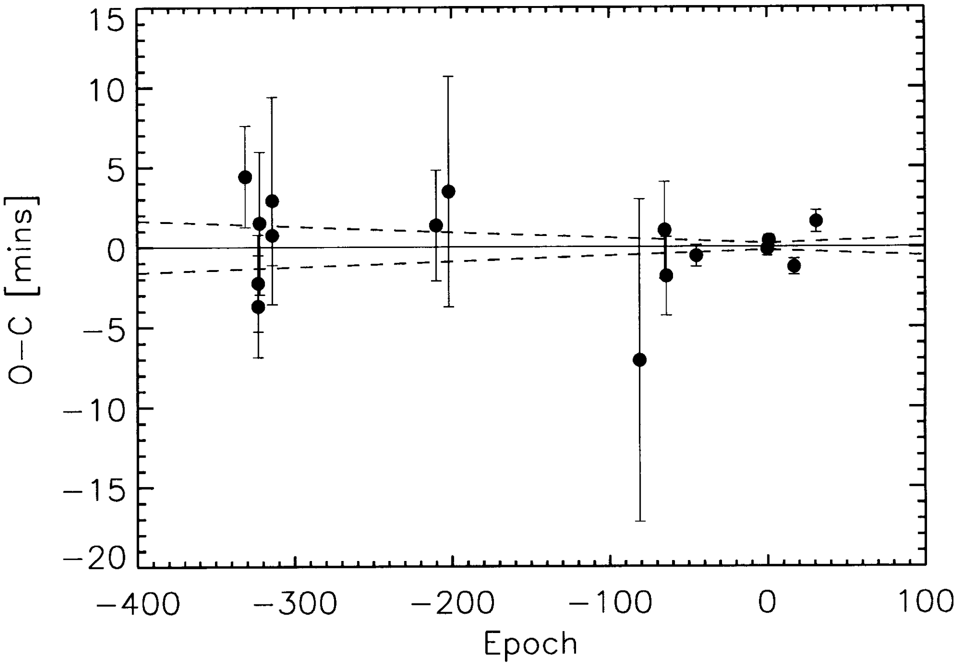

Figure 4-12 Transit-timing variations for HD 149026b. The differences between observed and calculated mid-transit times are plotted. The last 4 points represent the new NICMOS observations. The dashed lines show the 1σ range in the calculated times according to the linear ephemeris presented in § 4.6.

4.7 Discussion of broadband results

We have presented observations of four transits of HD 149026b at near-infrared wavelengths with the *HST* NICMOS detector. The NICMOS data place the strongest constraints yet on the geometrical system parameters. In particular, the increased precision of the measurement of the normalized semi-major axis ($a/R_\star$) leads to an improved estimate of the mean stellar density, which was then coupled with stellar evolution models to constrain the stellar mass and radius. Improved knowledge of the stellar mass and radius leads to greater precision in the planetary mass and radius.

We have found a larger stellar radius, and a larger planet-to-star radius ratio, than previous estimates. As a result of these two factors, we have also found the planetary radius to be larger than previously thought. The planet has "grown" by about 7%.

Despite this increase, our results are still consistent with the contention that HD 149026b is highly enriched in heavy elements. It is still smaller than expected for a hydrogen-helium planet with the given mass and degree of stellar irradiation (Burrows et al. 2007). For comparison, the tabulated models by Fortney et al. (2007) predict a 1.3 $R_{\rm Jup}$ hydrogen-helium HD 149026b at an age of 1 Gyr. A variety of models have been developed to estimate the heavy-element content of HD 149026b (Sato et al. 2005, Fortney et al. 2006, Ikoma et al. 2006, Burrows et al. 2007), most of which suppose that the metals are confined to an inner core of material beneath a hydrogen-helium envelope. Other physical considerations in these models include the equation of state for heavy elements at core pressures, atmospheric opacities and the upper boundary condition where energy is delivered from the star. To determine a revised estimate for the heavy-element content, we used the tabulated models provided by Sato et al. (2005) and Fortney et al. (2007), and interpolated the tabulated results as appropriate for the planetary radius, planetary mass, and degree of irradiation that follow from the parameters determined from the NICMOS data. We find a core mass in the range of $45 - 70$ $M_{\oplus}$, depending on assumed stellar age and core density. Thus, the interpretation of the planet as highly enriched is unaffected, although the required amount of enrichment is slightly reduced.

It is also interesting that the planet-to-star area ratio, $(R_p/R_\star)^2$, was found to be 2–3σ larger in the NICMOS band (1.1–2.0 μm) than in the optical band (0.45–0.55 μm) or mid-infrared bands (6.5–9.5 μm), while the results for the latter two bands are in agreement. Caution dictates that this discrepancy should not be over-interpreted. It is possible that the discrepancy is at least partly the result of unresolved systematic errors in any of the data sets. We have already noted that the noise in the NICMOS data exceeds the photon noise level by a factor of 2, and is not well understood.

However, it is also worth considering that this wavelength-dependent variation represents selective absorption by constituents in the outermost layer of the planet's atmosphere. Molecules with strong absorption bands at near-infrared wavelengths would cause the transit to appear deeper at those wavelengths. Strong bands are expected for the common molecules CO, H_2O, and CH_4 (Brown 2001, Hubbard et al. 2001, Seager & Sasselov 2000). If this were the case, then a detailed analysis of the NICMOS spectrophotometry—breaking it down into smaller wavelength bins, as opposed to summing the entire first-order spectrum—might be used to identify some constituents of the planet's atmosphere. In addition, more care would be needed in choosing which radius to use in the comparison with models of the planet's interior. It is beyond the scope of this chapter to analyze the wavelength dependence of the transit depth across the NICMOS band, or to compute a realistic atmospheric model to see if the contrast between the optical, near-infrared, and mid-infrared results can be accommodated. We can, however, perform an order-of-magnitude calculation to check on the plausibility of this interpretation.

Let $z(\tau)$ be the height in the planet's atmosphere at which the optical depth is τ for a path from the star to the observer, as diagrammed in Fig. 4-13. This height is measured relative to an atmospheric base radius R_0, where the planet is optically thick at all relevant wavelengths. The height $z(\tau)$ depends, in part, on the wavelength-dependent opacity and the density profile of the atmosphere. We define $R_p(\lambda)$ as $R_0 + z(\tau = 1)$ and $\delta \equiv (R_p/R_\star)^2$. If we assume that $z(1) \ll R_0$, then the transit depth is approximately linear in $z(1)$:

$$\delta = \frac{[R_0 + z(1)]^2}{R_\star^2} \approx \left(\frac{R_0}{R_\star}\right)^2 \left[1 + \frac{2z(1)}{R_0}\right]. \tag{4.15}$$

Next, we consider the difference in δ as measured in two distinct wavelength bands:

$$\delta_1 - \delta_2 \approx \frac{2R_0}{R_\star^2}\left[z_1(1) - z_2(1)\right] = \frac{2R_0}{R_\star^2}\delta z \tag{4.16}$$

where we have defined the height difference $\delta z \equiv z_1(1) - z_2(1)$. The height difference δz reflects differing levels of absorption in the two bands. Solving for δz in Eqn. (4.16) we find

$$\delta z = \frac{1}{2}\left[\frac{\delta_1 - \delta_2}{(R_0/R_\star)^2}\right] R_0. \tag{4.17}$$

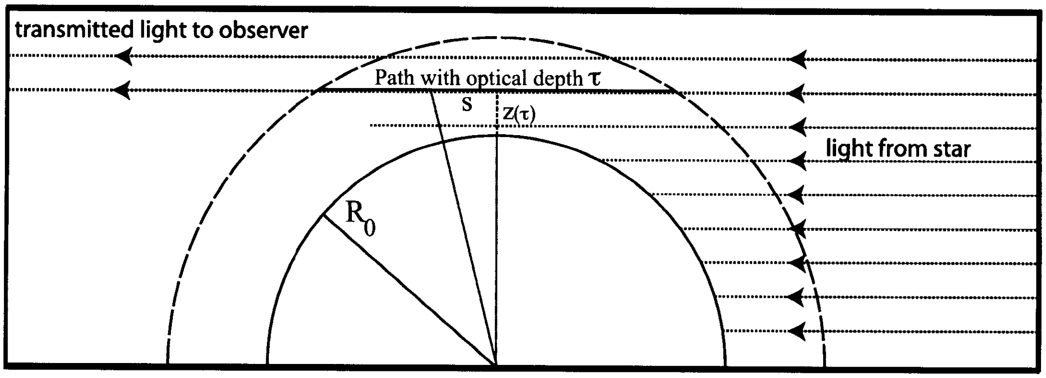

Figure 4-13 Illustration of wavelength-dependent absorption. Shown are some rays that skirt the planetary atmosphere on their way to Earth. At some height $z(\tau)$ above the fiducial radius R_0, the ray (parameterized by path length s) has an optical depth of τ (solid line). Light that follows paths with $z < z(\tau = 1)$ are mainly absorbed. The height $z(\tau = 1)$ corresponding to optical depth of unity depends on wavelength, giving rise to a wavelength-dependent transit radius.

For HD 149026b, to evaluate the idea that the larger near-infrared measurement of δ is due to molecular absorption, we assume that R_0 is the optically-derived radius ($R_0 = 0.757\ R_{\rm Jup}$) because the optical spectrum is expected to show comparatively weak absorption features (Brown 2001, Seager & Sasselov 2000). Using Eqn. (4.17), the 3.5×10^{-4} difference in δ between near-infrared and mid-infrared wavelengths implies $\delta z \approx 2500$ km (4% of R_0). To judge if this is realistic, we are interested in expressing δz in units of the pressure scale height, for which an order-of-magnitude expression is $H = kT/\mu m_p g_p$, where T is a representative atmospheric temperature, μ is the mean molecular weight of the atmosphere, m_p is the proton mass and g_p is the surface gravity. Using the surface gravity that was determined from our analysis of the optical light curve ($g_p = 1535$ cm s^{-2}), and assuming a H_2–He atmosphere with $T = 2300$ K [as measured at 8 μm by Harrington et al. (2007)], we find $H \approx 530$ km, and $N_H \approx 5$. If instead we use the planet's predicted temperature at thermal equilibrium with the incident stellar radiation ($T = 1700$ K), we find $H \approx 400$ km and $N_H \approx 6$.

If we assume further that the absorbers have an exponential density profile,

$$\rho(z) \;=\; \rho(0)\exp\left(-\frac{z}{h}\right), \tag{4.18}$$

where h is the density scale height, then we may express $z(\tau)$ in terms of the opacity σ of the absorbing molecules, as follows. By integrating the optical depth τ across the optical path at height $z \ll R_0$ (as illustrated in Fig. 4-13), where a change of path length ds results in a change in optical depth $d\tau$ as

$$d\tau = -s\sigma\rho(0)\frac{\exp\left[-\frac{R_0}{h}\left(\sqrt{(1+z/R_0)^2+(s/R_0)^2}-1\right)\right]}{\sqrt{(1+z/R_0)^2+(s/R_0)^2}}\frac{ds}{R_0}, \tag{4.19}$$

we find

$$\frac{z(\tau)}{h} \;\approx\; \ln\left(\frac{2h\sigma\rho(0)}{\tau}\right) - 1. \tag{4.20}$$

In general, h may be different for each atmospheric constituent. If we assume that the components are uniformly mixed throughout the atmosphere, then h is independent of composition, and the difference between two heights $z_1(\tau)$ and $z_2(\tau)$ at two distinct wavelength bands with different opacities σ_1 and σ_2 can be written independently of the optical depth, as

$$\delta z = z_1 - z_2 \;\approx\; h\ln\left(\frac{\sigma_1}{\sigma_2}\right). \tag{4.21}$$

If we assume further that the temperature scale height is large compared to the pressure scale height, then $h \approx H$ and $N_H \approx \ln(\sigma_1/\sigma_2)$.

We may now judge the plausibility of this interpretation with reference to the typical opacities and widths of molecular absorption features. For strong molecular bands and atomic lines, the ratio of the in-band opacity to the nearby continuum opacity may be as large as 10^4 (Brown 2001, Seager & Sasselov 2000), yielding a maximum height difference of $N_H \approx 10$ scale heights within the absorption band. The NICMOS band from $1.1 - 2.0$ μm includes strong rotation-vibration molecular absorption bands due to H_2O, CO and CH_4. An example of a very strong absorption band is a water band centered at 1.4 μm, spanning approximately 10% of the effective filter width. If we assume that this is the dominant spectral feature in this band, then the result of $N_H \approx 5$ across the entire bandpass translates into $N_H \approx 50$ within the bandpass of the absorption feature. This is larger than the criterion $N_H \approx 10$ mentioned above.

Therefore this interpretation seems to require significantly more opaque or broader-band absorption features than are seen in the models. In one sense the result of the order-of-magntiude calculation is discouraging, as it may make it seem more likely that the discrepancy in depths is due to systematic errors. On the other hand, if the noise were well-understood and the discrepancy could be confidently proclaimed, then it would be the sign of new and interesting atmospheric physics that is not described in the standard models.

Some priorities for progress on this issue include an examination of the wavelength-dependence of the transit across the NICMOS band (see the next section), and the observation of the system with other NICMOS grisms, which are reputed to be more stable than the G141 grism used here.

4.8 Transmission spectroscopy

To further assess the larger-than-expected transit depth seen with the broadband light curve, we performed transmission spectroscopy over the NICMOS bandpass from $1 - 2$ μm. To this end, we divided the rectangular aperture described in § 4.2 along the dispersion axis into twenty-four uniformly sized sub-apertures. Twenty-four wavelength "channels," as constructed here, correspond to a sub-aperture width approximately equal to the full-width-at-half-maximum of the defocused stellar PSF. Each sub-aperture is indexed by a wavelength which is calculated via the relation in Eqn. (4.1) where now Δx is the x coordinate of the sub-aperture center relative to the center of the undispersed image. We perform aperture photometry on each sub-aperture as described in § 4.2 generating twenty-four transit light curves for each of four transit observations. Total counts are suppressed at either end of the wavelength range as the G141 filter's transmission has sharp but not perfectly vertical profiles at the edges.

We fit, via Levenberg-Marquardt, a transit light curve to each wavelength channel and for each visit. We require that the global orbital parameters describing the transit (inclination, $a/R_\star$, time of mid-transit) be consistent across all wavelengths and visits, but we allow the transit depth, out-of-transit flux, and systematic correction parameters (f_o^v, c_0^v, c_1^v, c_2^v, c_3^v) to vary freely with wavelength. We fixed the limb-darkening quadratic parameters to those values found from the broadband analysis[6].

Figure (4-14) shows the systematic-corrected light curves spanning the wavelength range of the NICMOS G141 filter. It is obvious from this figure that the transit depth is a strong function of wavelength. The variation is nonlinear in that there are banded features near 1.2 and 1.6 micron. Such features are suggestive of absorption features due to molecular absorption [as may have been identified by Swain et al. (2008)]. Figure (4-11) plots the transit depth, with error bars estimated from χ^2, versus wavelength. The minimum absorption level of the spectrum is consistent with the depth measured in the optical [at 0.5μm, the 68% confidence region is indicated by the blue bar in Fig. (4-14)]. This result is encouraging as the transit depth measured at 0.5μm should correspond to the minimum transit radius, R_0 (Brown 2001). However, the strength of the absorption near 1.2 and 1.6 micron is unphysically large compared to expectations making standard assumptions. In particular, using the formalism from the previous section, the feature at 1.2 micron corresponds to an opacity source for which $N_H > 15$.

Investigating further, we computed an expected transmission spectrum for HD

---

[6] While there is no reason to suspect the limb darkening parameters are constant for all wavelengths in the NICMOS band, we find that fixing u_1 and u_2 across all wavelengths has a negligible effect on the transmission spectrum as compared to wavelength dependent models. For example, we experimented with limb darkening models for which u_1 and/or u_2 varied linearly with wavelength and observed no significant difference in the transmission spectrum.

149026b following the formalism provided by Brown (2001). In detail, we populated our model atmosphere with the most important molecules in the NICMOS band as determined by their respective wavelength dependent opacities. These molecules are H_2, He, N_2, CO, H_2O, CH_4, and NH_3. Opacities for each molecular species are specified using the values tabulated in the HITRAN database (Rothman et al. 1998). Atmospheric height z is related to temperature and pressure by solving the equation of hydrostatic equilibrium assuming a pressure-temperature profile for HD 149026b as specified by Fortney et al. (2006). We then calculated partial pressures and densities at several atmospheric heights for the mixture of molecules assuming chemical equilibrium. We then integrated the total wavelength-dependent opacity, $\sigma_{\text{tot}}(\lambda)$, along an optical path tangential to the exoplanetary surface at height z, as diagrammed in Fig. (4-13). The transit depth at wavelength λ is defined by the atmospheric height $z(\tau)$ for which the optical depth is unity, $\tau = 1$, where $d\tau = -\sigma_{\text{tot}}(\lambda)ds$. The red spectrum in Fig. (4-11) gives this calculated transmission spectrum at high spectral resolution over the NICMOS G141 wavelength range. The green curve shows this same spectrum after convolution with the expected PSF for our defocused observation.

The model transmission spectrum is clearly incompatible with the measured transmission spectrum. The strong features at 1.2 and 1.6 micron occur at wavelengths where the model spectrum suggests no significant absorption should occur. The absorption feature near 1.4 micron can be partly attributed to absorption by water, however, the measured transit depth is still too large to be completely explained. Currently, we have no model for the absorption features suggested by the data, however, it is possible the features seen are as a result of a yet-to-be-identified wavelength dependent systematic with the G141 grism. Given the unusual systematics already seen with this grism and filter, intuition would suggest this interpretation is plausible. On the other hand, our atmospheric model may be too simplistic. While we have tried to incorporate the most likely molecular atmospheric physics in our model transmission spectra, it may be the case that the strong absorption is due to non-molecular absorption such as that due to Rayleigh scattering or collision-induced absorption by

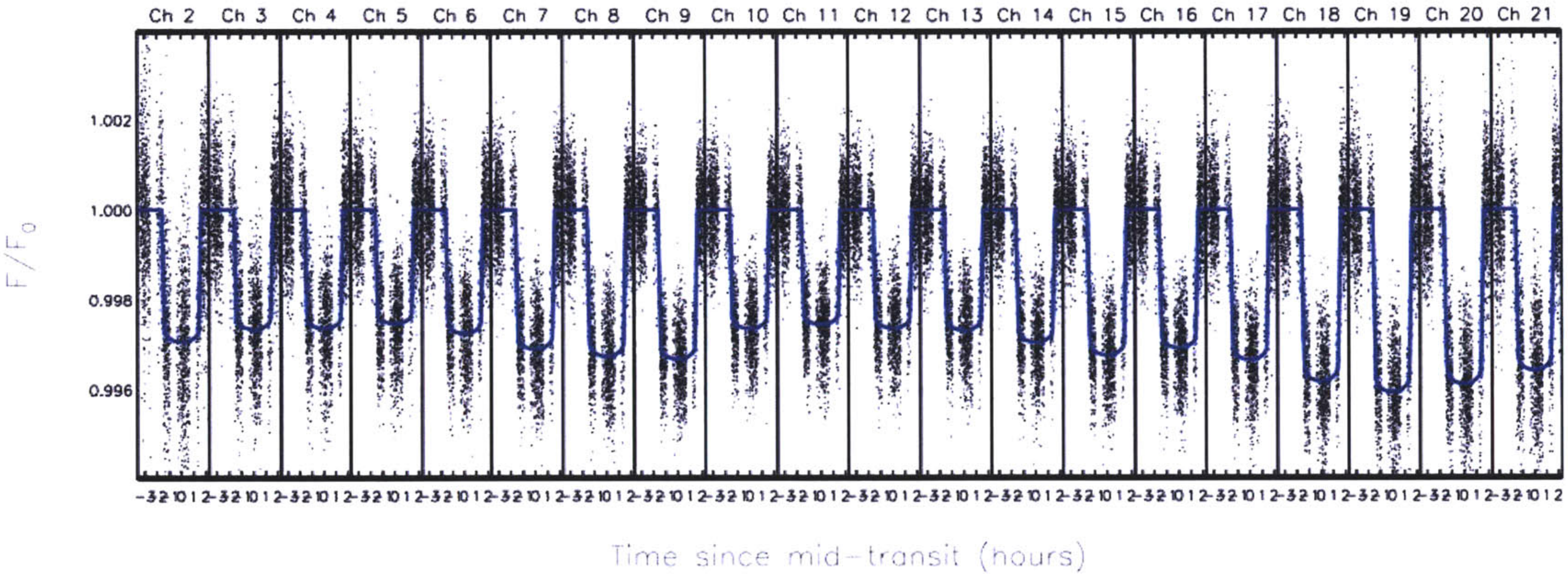

Figure 4-14 Measured transit light curves of HD 149026b at the same transit epoch over twenty-four uniformly distributed wavelength channels covering the NICMOS G141 1.1 – 2.0 μm bandpass. The solid blue curves give the best fit transit model. The data have been corrected for systematics due to intra-orbital variation and inter-orbital offsets. The channel numbers increase from blue to red wavelengths.

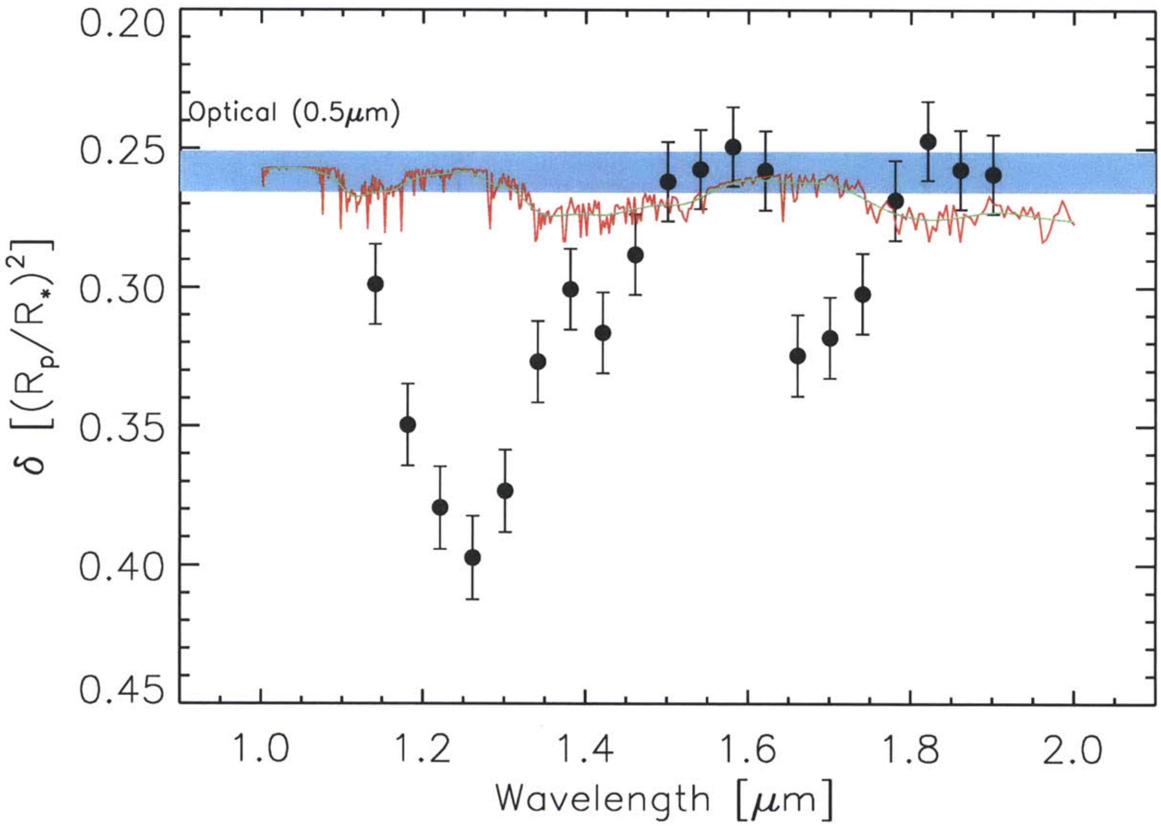

Figure 4-15 The transmission spectrum from $1.1-2.0$ μm for HD 149026b. The black points are the transit depths estimated from a global analysis of twenty-four transit light curves for each of four transit epochs over distinct wavelengths as measured with NICMOS. The blue bar represents the 68% confidence interval in the transit depth from a independent analysis of optical (0.5 μm) transit data. The red curve is the expected transmission spectrum, at high spectral resolution, modeled as described in § 4.8. The green curve is the same model spectrum convolved with the expected PSF for the NICMOS observation.

H_2. Basic scaling arguments show that these effects are unlikely to completely account for the anomalously large transit depth (Brown 2001). More likely is that the strong absorber, should it exist, is atomic in nature. We expect the atmosphere to be quite diffuse at the extreme height implicated by the measured transmission spectrum at 1.2 μm, for example. At such heights, molecules should be rare. A carefully tailored distribution of ionized particles may yield continuum or wide band absorption in the NICMOS band, possibly accounting for the transit depth features at 1.2 and 1.6 μm. We do not speculate further on such atomic absorption as it is beyond the scope of this chapter. Another possibility accounting for the stronger than expected absorption is that we may have simply misrepresented the opacities of the rotation-vibration modes of the specified molecular families. Indeed, the HITRAN database is not suited for use at the high temperatures expected in the exoplanet atmosphere. Our model will need to be recalculated using the forthcoming high temperature complement to HITRAN, HITEMP (Rothman, private communication) and possibly additional high temperature line-lists (e.g., Nassar & Bernath 2003).

While the transmission spectrum in Fig. (4-11) is intriguing, little may be constrained from it about the atmosphere of HD 149026b assuming only simple atmospheric models as described above. We intend to revisit our analysis discussed in this section with new observations and models as soon as they become available.

We thank G. Torres for helpful discussions concerning the determination of stellar parameters. We thank L. Rothman for helpful discussions concerning the HITRAN molecular opacity database. This work was supported by NASA grant HST-GO-11165 from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

		Simultaneous Fit		
Parameter	NICMOS only	NICMOS	*Spitzer* 8 μm	$(b+y)/2$
Wavelength range [μm]	1.1–2.0	1.1–2.0	6.5–9.5	0.45–0.55
Cadence [secs][a]	7.2	7.2	4.1	8.6
Normalized error[b]	0.00044	0.00044	0.0026	0.0020
$R_p/R_\star$	$0.05416^{+0.00091}_{-0.00070}$	$0.05430^{+0.00085}_{-0.00078}$	$0.05188^{+0.00084}_{-0.00086}$	$0.05070^{+0.00058}_{-0.00088}$
$(R_p/R_\star)^2 \times 100$	$0.2933^{+0.0099}_{-0.0076}$	$0.2949^{+0.0092}_{-0.0085}$	$0.2692^{+0.0087}_{-0.0089}$	$0.2570^{+0.0059}_{-0.0089}$
i [deg]	$84.55^{+0.35}_{-0.81}$	$84.50^{+0.60}_{-0.52}$		
$a/R_\star$	$6.01^{+0.17}_{-0.23}$	$5.99^{+0.21}_{-0.18}$		
$b \equiv a \cos i/R_\star$	$0.571^{+0.044}_{-0.038}$	$0.574^{+0.035}_{-0.045}$		
Ingress Duration [hr][c]	$0.241^{+0.012}_{-0.013}$	$0.243^{+0.012}_{-0.013}$	$0.232^{+0.011}_{-0.012}$	$0.227^{+0.011}_{-0.012}$
Transit Duration [hr][d]	$3.24^{+0.14}_{-0.15}$	$3.25^{+0.14}_{-0.15}$	$3.24^{+0.14}_{-0.15}$	$3.23^{+0.14}_{-0.15}$
$M_\star$ [$M_\odot$]	$1.345^{+0.020}_{-0.020}$	$1.34^{+0.020}_{-0.020}$		
$R_\star$ [$R_\odot$]	$1.541^{+0.046}_{-0.042}$	$1.534^{+0.049}_{-0.047}$		
$\rho_\star$ [g cm^{-3}]	$0.497^{+0.042}_{-0.057}$	$0.492^{+0.052}_{-0.044}$		
$\log g_\star$ [cgs]	$4.189^{+0.020}_{-0.021}$	$4.192^{+0.022}_{-0.022}$		
Distance [pc]	$83.0^{+2.8}_{-2.7}$	$82.6^{+2.8}_{-2.8}$		
Stellar Age [Gyr]	$2.6^{+0.3}_{-0.2}$	$2.6^{+0.2}_{-0.2}$		
$L_\star$ [$L_\odot$]	$3.03^{+0.20}_{-0.18}$	$3.00^{+0.20}_{-0.19}$		
M_p [$M_{\rm Jup}$][e]	$0.368^{+0.013}_{-0.014}$	$0.366^{+0.014}_{-0.013}$		
R_p [$R_{\rm Jup}$]	$0.813^{+0.027}_{-0.025}$	$0.811^{+0.029}_{-0.027}$	$0.775^{+0.028}_{-0.027}$	$0.757^{+0.025}_{-0.027}$
ρ_p [g cm^{-3}]	$0.85^{+0.10}_{-0.09}$	$0.85^{+0.10}_{-0.09}$	$0.98^{+0.11}_{-0.11}$	$1.05^{+0.11}_{-0.12}$
$\log g_p$ [cgs][e]	$3.132^{+0.029}_{-0.035}$	$3.127^{+0.033}_{-0.029}$	$3.166^{+0.034}_{-0.030}$	$3.186^{+0.032}_{-0.030}$

Table 4.1 **System Parameters of HD 149026.** [a]Defined as the median time interval between data points in the composite (phase-folded) light curve. [b]Defined as σ_F/F where σ_F is the rms residual between the data and best-fitting model, and F is the out-of-transit flux. [c]Defined as the time between first and second contacts, or between third and fourth contacts. (In our model these durations must be equal.) [d]Defined as the time between the first and fourth contacts. [e]Using $K = 43.3 \pm 1.2$ m s^{-1}, from Sato et al. (2005).

Epoch	Mid-transit time [HJD]	Error
0	2454456.78751	0.00030
1	2454459.66379	0.00023
17	2454505.67688	0.00034
31	2454545.94133	0.00047

Table 4.2 Mid-transit times, based on the NICMOS data.

Bibliography

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567

Amendola, L. 1996, MNRAS, 283, 983

Broeg, C., & Wuchterl, G. 2007, MNRAS, 376, L62

Brown, T. M. 2001, ApJ, 553, 1006

Burke, C. J., et al. 2007, ApJ, 671, 2115

Burrows, A., Hubeny, I., Budaj, J., & Hubbard, W. B. 2007, ApJ, 661, 502

Carter, J. A., Yee, J. C., Eastman, J., Gaudi, B. S., & Winn, J. N. 2008, ApJ, 689, 499

Charbonneau, D., et al. 2006, ApJ, 636, 445

Charbonneau, D., Brown, T. M., Burrows, A., & Laughlin, G. 2007, Protostars and Planets V, 701

Claret, A. 2000, A&A, 363, 1081

Claret, A. 2007, A&A, 467, 1389

Demarque, P., Woo, J.-H., Kim, Y.-C., & Yi, S. K. 2004, ApJS, 155, 667

Ford, E. B., & Holman, M. J. 2007, ApJ, 664, L51

Fortney, J. J. 2005, MNRAS, 364, 649

Fortney, J. J., Saumon, D., Marley, M. S., Lodders, K., & Freedman, R. S. 2006, ApJ, 642, 495

Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, ApJ, 659, 1661

Fortney, J. J., Saumon, D., Marley, M. S., Lodders, K., & Freedman, R. S. 2006, ApJ, 642, 495

Gelman, A., Rubin D. B. 1992, Statistical Science, 7, 457

Gilliland, R. L. 2006, The 2005 HST Calibration Workshop: Hubble After the Transition to Two-Gyro Mode, 162

Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, A&AS, 141, 371

Harrington, J., Luszcz, S., Seager, S., Deming, D., & Richardson, L. J. 2007, Nature, 447, 691

Holman, M. J., & Murray, N. W. 2005, Science, 307, 1288

Hubbard, W. B., Fortney, J. J., Lunine, J. I., Burrows, A., Sudarsky, D., & Pinto, P. 2001, ApJ, 560, 413

Ikoma, M., Guillot, T., Genda, H., Tanigawa, T., & Ida, S. 2006, ApJ, 650, 1150

Johns-Krull, C. M., et al. 2008, ApJ, 677, 657

Levenberg, K., 1944, The Quarterly of Applied Mathematics, 2, 164

Lourakis, M. I. A., 2004, `http://www.ics.forth.gr/~lourakis/levmar`

Mandel, K., & Agol, E. 2002, ApJ, 580, L171

Masana, E., Jordi, C., & Ribas, I. 2006, VizieR Online Data Catalog, 345, 735

Marquardt, D., 1963, SIAM Journal on Applied Mathematics, 11, 431

Nassar, R., & Bernath, P. 2003, Journal of Quantitative Spectroscopy and Radiative Transfer, 82, 279

Nutzman, P., Charbonneau, D., Winn, J. N., Knutson, H. A., Fortney, J. J., Holman, M. J., & Agol, E. 2008, ArXiv e-prints, 805, arXiv:0805.0777

Pál, A. 2008, MNRAS, 390, 281

Rothman, L. S., et al. 1998, Journal of Quantitative Spectroscopy and Radiative Transfer, 60, 665

Salpeter, E. E. 1955, ApJ, 121, 161

Seager, S., & Mallén-Ornelas, G. 2003, ApJ, 585, 1038

Seager, S., & Sasselov, D. D. 2000, ApJ, 537, 916

Sato, B., et al. 2005, ApJ, 633, 465

Southworth, J. 2008, arXiv:0811.3277

Southworth, J., Wheatley, P. J., & Sams, G. 2007, MNRAS, 379, L11

Sozzetti, A., Torres, G., Charbonneau, D., Latham, D. W., Holman, M. J., Winn, J. N., Laird, J. B., & O'Donovan, F. T. 2007, ApJ, 664, 1190

Swain, M. R., Vasisht, G., & Tinetti, G. 2008, ArXiv e-prints, 802, arXiv:0802.1030

Thompson, R. 1992, Space Science Reviews, 61, 69

Torres, G., Winn, J. N., & Holman, M. J. 2008, ApJ, 677, 1324

van Leeuwen, F. 2007, A&A, 474, 653

Winn, J. N., et al. 2008, ApJ, 683, 1076

Winn, J. N., Holman, M. J., & Roussanova, A. 2007, ApJ, 657, 1098

Winn, J. N., Henry, G. W., Torres, G., & Holman, M. J. 2008, ApJ, 675, 1531

Yi, S., Demarque, P., Kim, Y.-C., Lee, Y.-W., Ree, C. H., Lejeune, T., & Barnes, S. 2001, ApJS, 136, 417

Chapter 5

An Empirical Upper Limit on the Oblateness of an Exoplanet

5.1 Introduction

To first order, an exoplanet is shaped like a sphere. The spherical approximation is appropriate for a non-rotating, self-gravitating, fluid planet in isolation. In general though, the shape of an exoplanet is affected by two components: one owing to external gravitational influences and the other to exoplanetary rotation. The surficial shape of tidally influenced objects is due, in part, to the presence of tidal "bulges" formed in response to a gravitational potential gradient. For example, perturbations to Earth's spherical shape (in its oceans and atmosphere) are as a result of the gravitational tug of the Moon and Sun (Murray & Dermott 2000). An analogous perturbation is expected in the case of "Hot Jupiter" exoplanets, this time as a result of tides raised by the nearby host star. Some satellites, notably the Moon to the Earth and Mercury to the Sun, have "frozen-in" bulges existing in equilibrium with their companion (Murray & Dermott 2000). Exoplanets incapable of establishing permanent bulges, such as is expected with gas giants, are likely to suffer from dissipative effects in their interiors, relinquishing orbital and rotational energy as tides are raised (Goldreich & Soter 1966, Peale 1999, Murray & Dermott 2000). A measure of exoplanetary shape in combination with an understanding of these dissipative timescales (see, e.g., Hut 1981) will give clues to the orbital evolution of exoplanet and star.

Even in isolation from tides, properties intrinsic to the exoplanet shape its surface, primarily rotation (Hubbard 1984, Murray & Dermott 2000). Uniform rotation of a self-gravitating fluid should yield an oblate spheroidal equilibrium shape having minimum radial extent being along the axis of rotation (Eddington 1926). The deviation from sphericity can be dramatic for fast rotators. Saturn and Jupiter, for example, rotate rapidly enough (with rotational periods of 9.9 hr and 10.2 hr, respectively) such that their respective oblateness is plainly evident in full disk images. The oblate spheroidal shape is a valid approximation for planets within our own Solar System (Murray & Dermott 2000) and is a likely model for exoplanets not under strong tidal influence by their host. Ignoring for the moment the effects of external gravitational perturbations, a measurement of planet shape would yield the planetary rotation rate, likely a key observable describing the planet's formation history[1].

For these reasons and more (such as the enhancement in the polarization of scattered starlight by an oblate exoplanet; Sengupta & Maiti 2006) it is of great interest to determine the shapes of planets outside our own Solar System. Fortunately, exoplanetary transit light curves are sensitive to the line-of-sight projected shape of the transiting exoplanet. In particular, it is possible to utilize the transit light curve to constrain the oblateness of the exoplanet, f, defined as

$$f = \frac{R_{\rm eq} - R_{\rm pol}}{R_{\rm eq}} \tag{5.1}$$

in terms of the polar and equatorial radii of the oblate spheriod (Murray & Dermott 2000). Additionally, with the transit light curve, we may constrain the obliquity angle, θ, measured between the orbital spin angular momentum axis and the polar axis. That exoplanetary oblateness could induce an observable light curve signature had been reported first by Seager & Hui (2002) and investigated further by Barnes & Fortney (2003). Both appreciated the need for high precision light curves for an unambiguous detection of an exoplanetary oblateness signature. For example, in order

---

[1] A measure of oblateness is not the only route to a constraint on planetary rotation. See, for example, Speigel et al. (2007) for a novel route.

to detect the rotationally induced oblateness signature in the transit light curve for the case of a synchronously rotating HD 209458b, one would need to measure the light curve to better than 0.1 parts-per-million (ppm) (Barnes & Fortney 2003).

Currently, no transit data is available at this precision, however in the near future, space-based missions, such as the recently launched *Kepler* mission (Borucki et al. 2009), will likely provide many high-precision light curves for short period exoplanets. It is important, therefore, to develop the necessary tools to assist in precisely measuring exoplanetary oblateness from transit light curves. In this chapter, we present a fast, efficient method for the calculation of light curves of oblate exoplanets. We then demonstrate this tool by analyzing seven high-precision transit light curves of the well-studied Hot Jupiter HD 189733b (Agol et al. 2009), constraining, for the first time, parameters relating to an exoplanet's oblateness. Finally, we calculate constraints on oblateness for the eccentric transiting exoplanet HD 80606b (Naef et al. 2001, Moutou et al. 2009) with a single, simulated high precision *Hubble Space Telescope* transit light curve.

This chapter is organized as follows: In § 5.2, we review some physics involved with exoplanetary oblateness, primarily those associated with exoplanetary rotation. Here, we also pause to consider other effects that may compete or mask the signal of planetary oblateness in the transit light curve. In § 5.3, we describe a numerical algorithm to quickly calculate transit light curves of oblate exoplanets via a semi-analytic, quasi-Monte Carlo integration technique. In § 5.4, we present seven *Spitzer* IRAC 8 μm transit observations of HD 189733b. In § 5.4.3, we present results of a Markov chain Monte Carlo (MCMC) analysis of the combined light curve of the seven transit light curves constraining, in addition to standard transit parameters, the *projected* oblateness and obliquity. In § 5.5, we summarize our methods and results, highlighting extensions of the work completed in this chapter. We conclude § 5.5 by asking what may be inferred with regards to oblateness from a single high-precision transit light curve for HD 80606b.

5.2 Physical review

In this section we review the physics needed for interpreting a measurement of exoplanetary oblateness. Our discussion follows Murray & Dermott (2000) pp. 150 ff., Peale (1999) and is similar to the treatments in Seager & Hui (2002) and Barnes & Fortney (2003).

5.2.1 Relevant timescales

The tidal influence of a star on a close exoplanet can result in changing orbital parameters owing to friction in the planet's interior. In particular, tidal oscillation forced by the stellar gravity in a rotating exoplanet's bulk are frictionally damped in such a way that the axis of the tidal "bulge" lags behind the line connecting the centers of planet and host. As a result, a torque is established on the planet, opening a channel of energy and angular momentum transfer (Murray & Dermott 2000, Peale 1999). The final result of this transfer is synchronous rotation of planet and host (i.e., 1:1 spin-orbit synchronization). The characteristic timescale for synchronization is given by

$$\tau \sim Q\left(\frac{R_{\mathrm{mean}}^3}{GM_p}\right)\omega_{\mathrm{prim}}\left(\frac{M_p}{M_\star}\right)^2\left(\frac{a}{R_{\mathrm{mean}}}\right)^6 \tag{5.2}$$

where $R_{\mathrm{mean}} = \sqrt{R_{\mathrm{eq}}R_{\mathrm{pol}}}$, a is the orbital semi-major axis, ω_{prim} is the planet's primordial rotation rate, M_p is the planet mass, $M_\star$ is the stellar mass, and Q is the specific dissipation factor of the tidal oscillator (Goldreich & Soter 1966, Hubbard 1984, Guillot et al. 1996, Murray & Dermott 2000). The dissipation factor Q is poorly known even for Solar System planets (Goldreich & Soter 1966). A detection of oblateness in combination with a precise estimate of the stellar system's age could be used to derive the exoplanetary Q (Seager & Hui 2002). Additional information about the evolutionary state of the orbital system is also encoded in the obliquity angle θ as described as follows. If the planet is in a tidally influenced phase, a non-zero obliquity would bring the tidal "bulge" out of the orbital plane as the planet rotates.

A torque in the perpendicular direction to the orbital plane will be established and, as a result, the obliquity angle will *increase* to some quasi-equilibrium angle related to the rotation rate. Upon reaching this quasi-equilibrium configuration, the obliquity will reduce to zero in a similar time as the time to the march to synchronous rotation (Peale 1999). For a fixed circular orbit, this equilibrium obliquity is $\theta = 0$ and the timescale for zeroed obliquity and synchronous rotation are equivalent (Peale 1999). It is possible, however, to sustain a non-zero obliquity in true equilibrium for exoplanets in a so-called Cassini state. Here, if the stellar and orbital spin axes are initially precessing (by torques set up by other planets, or satellites for example), the final equilibrium state is that in which the spin axes precess at the same rate about a common precession axis. In this case, the equilibrium obliquity may be significantly different from zero (Winn & Holman 2005). Notably, the Earth-Moon system are in a Cassini state (Ward 1975).

Planets with orbital planes tilted relative to the stellar equatorial plane or having non-zero eccentricities tend to coplanarity and circular orbits, with timescales for the associated processes being longer than the timescale of synchronization.[2] (Hut 1981). Planets with large orbital eccentricity (and with close periastron distances), are likely to achieve pseudo-synchronous rotation in that the exoplanet rotates at the instantaneous orbital velocity at periastron (Hut 1981). For HD 80606b, such a configuration seems plausible, however, the dynamics are not so simple owing to the complex interaction between HD 80606 and its nearby companion HD 80607. In particular, the large eccentricity of HD 80606b may have been pumped from a primordial, wide, circular orbit to its current state (Wu and Murray 2003) from gravitational perturbations by HD 80607 (the so-called Kozai effect; Kozai 1962). Spin-orbit obliquity is a trademark consequence of the Kozai process. Spin-orbit misalignment has in fact been detected for HD 80606b (Winn et al. 2009b). The slow rotation rate expected for a pseudo-synchronous HD 80606b (on the order of days) may not be realized owing to these effects. Thus, it may be possible that HD 80606b is rotating fast enough to induce a measurable oblateness. A constraint on

---

[2] Orbital eccentricity couples to obliquity in a more complicated manner (see Peale 1999).

the rotation of HD 80606b will be of great interest to dynamicists.

We may estimate the synchronization time τ for HD 189733b to help in our interpretation of any measurement of oblateness. Using $M_\star = 0.81\ M_\odot$, $M_p = 1.14\ M_{\mathrm{Jupiter}}$, $R_{\mathrm{mean}} = 1.14\ R_{\mathrm{Jupiter}}$, $a = 0.031$ AU (Torres et al. 2008) and $Q = 10^5$, $\omega_{\mathrm{prim}} = 1.7 \times 10^{-4}\ \mathrm{s}^{-1}$ (Jupiter's values; Guillot et al. 1996), we find $\tau \sim 10^6$ yr. It is therefore a likely conclusion that the planet has reached synchronous rotation having compared the calculated τ to the age of HD 189733 ($= 6 \pm 5$ Gyr; Torres et al. 2008). The fact that HD 189733 has a negligible measured orbital eccentricity corroborates this conclusion (Agol et al. 2009). Therefore, we expect HD 189733b to be rotating once every orbital period, with the length of the "exoday" being ≈ 2.2 days. Additionally, the spin-orbit obliquity angle should be near zero. If our theoretical expectations are physically accurate, rotationally induced oblateness of HD 189733b should be small (see below).

5.2.2 Oblateness and rotation

The relationship between oblateness, f, and the rotational period P_{rot} may be established with the following planetary model (following Murray & Dermott 2000). We consider an exoplanet that, as a result of rotation or otherwise, is an oblate spheroid, whose external gravitational potential may be written as

$$V_{\mathrm{gravity}}(r,\theta) = -\frac{GM_p}{r}\left[1 - \sum_{n=2}^{\infty} J_n \left(\frac{R_{\mathrm{eq}}}{r}\right)^n \mathcal{P}_n(\cos\theta)\right] \qquad (5.3)$$

where J_n are the spherical moments of the mass distribution[3] , and $\mathcal{P}_n(\cos\theta)$ is the Legendre polynomial of degree n. For an exoplanet with rotational angular speed $\Omega = 2\pi/P$, the centrifugal potential[4] may be written as

$$V_{\mathrm{centrifugal}}(r,\theta) = \frac{1}{3}\Omega^2 r^2 \left[\mathcal{P}_2(\cos\theta) - 1\right]. \qquad (5.4)$$

---

[3] $J_n = \frac{1}{M_p R_{\mathrm{eq}}} \int_0^{R_{\mathrm{eq}}} \int_{-1}^{+1} r^n \mathcal{P}_n(\mu)\rho(r,\mu) 2\pi r^2 \, d\mu \, dr$

[4] The centrifugal potential is the scalar that describes the conservative "force" felt by otherwise unforced test particles in a uniformly rotating non-inertial reference frame.

For an exoplanet with symmetric hemispheres, $J_{2k+3} = 0$ for all integers $k \geq 0$. We assume that even moments $J_{2(k+1)}$ with $k > 0$ are negligible relative to J_2 so that the total potential may be written as

$$V_{\text{total}}(r,\theta) = -\frac{GM_p}{r} + \left(\frac{GM_p R_{\text{eq}}^2}{r^3} J_2 + \frac{1}{3}\Omega^2 r^2\right) \mathcal{P}_2(\cos\theta) - \frac{1}{3}\Omega^2 r^2 \qquad (5.5)$$

The surface of the planet must lie on an equipotential of this potential. As such, we may relate equatorial and polar radii via $V_{\text{total}}(R_{\text{eq}}, \frac{\pi}{2}) = V_{\text{total}}(R_{\text{pol}}, 0)$:

$$-\frac{GM_p}{R_{\text{eq}}} - \frac{1}{2}\left(\frac{GM_p}{R_{\text{eq}}} J_2 + \Omega^2 R_{\text{eq}}^2\right) = -\frac{GM_p}{R_{\text{eq}}} + \frac{GM_p R_{\text{eq}}^2}{R_{\text{pol}}^3} J_2 \qquad (5.6)$$

which in turn requires that to leading order in f [Eqn. (5.1)]

$$f = \frac{3}{2} J_2 + \frac{1}{2}\frac{\Omega^2 R_{\text{eq}}^3}{GM_p}. \qquad (5.7)$$

Solving for $P_{\text{rot}} = 2\pi/\Omega$, we find

$$\begin{aligned} P_{\text{rot}} &= 2\pi\sqrt{\frac{R_{\text{eq}}^3}{GM_p\,(2f - 3J_2)}} \\ &= 2\pi\sqrt{\frac{R_{\text{mean}}^3}{GM_p\,(2f - 3J_2)\,(1-f)^{3/2}}}. \end{aligned} \qquad (5.8)$$

This equation differs from Seager & Hui (2002) in the addition of the J_2 term and from Barnes & Fortney (2003) in that we did not immediately relate the exoplanetary moment of inertia to J_2, Ω, and f via the Darwin-Radau relation (see § 5.5). For a planet with uniform density it can be shown that (Hubbard 1984)

$$J_2 = \frac{1}{2}\frac{\Omega^2 R_{\text{eq}}^3}{GM_p}. \qquad (5.9)$$

In this case, Eqn. (5.8) becomes

$$P_{\text{rot}} = 2\pi\sqrt{\frac{R_{\text{eq}}^3}{\frac{4}{5}GM_p f}}$$

$$= 2\pi\sqrt{\frac{R_{\rm mean}^3}{\frac{4}{5}GM_pf\,(1-f)^{3/2}}}. \tag{5.10}$$

The J_2 may be measured for Solar System planets by monitoring elliptical orbits of satellites whose orbits precess in response to the aspherical gravitational field of their host. Here the parameter J_2 appears as a parameter competing with the oblateness f in the determination of $P_{\rm rot}$. However, should we have some expectation of the rotational period (based upon the timescales presented above, for example) it may be possible to estimate J_2 with a measurement of oblateness. Table 5.1 tabulates f and J_2 for planets in our Solar System. It should be noted that the uniform density approximation for f [by combining Eqns. (5.9) and (5.7)] is incapable of producing the oblateness seen in Saturn or Jupiter, for example.

A minimum rotational period $P_{\rm min}$ may be set by considering the limit $f = 1$, however, a more physically interesting bound on f is set by the rotational-breakup limit. Rotational breakup will occur for oblateness $f_{\rm break}$ when the outward centrifugal acceleration, $a_{\rm cent}$ is equivalent to the gravitational acceleration, g. Considering the acceleration at the equator ($\theta = \pi/2$, $r = R_{\rm eq}$),

$$\begin{aligned} a_{\rm cent} &= \Omega^2 R_{\rm eq} && (5.11) \\ g &= \frac{GM_p}{R_{\rm eq}^2} - \frac{3}{2}\frac{GM_p}{R_{\rm eq}^2}J_2 \\ &= \frac{GM_p}{R_{\rm eq}^2}\left(1 - \frac{3}{2}J_2\right). && (5.12) \end{aligned}$$

Then, $\frac{\Omega^2 R_{\rm eq}^3}{GM_p} = (1 - 3/2\ J_2)\Omega^2$ at breakup ($a_{\rm cent} = g$). Therefore, by applying Eqn. (5.7), we have

$$f_{\rm breakup} = \frac{1}{2} + \frac{3}{4}J_2. \tag{5.13}$$

Table 5.1. Solar System planet parameters

Planet	Oblateness f	J_2
Mercury	0.00012	0.000060
Venus	0.00009	0.000004
Earth	0.00350	0.001083
Mars	0.00520	0.001960
Jupiter	0.06487	0.014736
Saturn	0.09796	0.016298
Uranus	0.02293	0.003343
Neptune	0.01708	0.003411

References. — Murray & Dermott (2000), Barnes & Fortney (2003), Hubbard (1984)

5.2.3 Competing effects in the transit light curve

Other effects may interfere with the detection of oblateness from transit light curve data or masquerade as a rotationally induced oblateness. In this section, we briefly discuss possible contaminants.

Rayleigh scattering, cloud scattering, refraction, absorption and diffraction will affect the measured transit radius, however will not, to first order, induce a non-zero projected-oblateness when the underlying total potential has spherical equipotentials (Seager & Sasselov 2000, Hubbard et al. 2001, Hui & Seager 2002). Refraction (also referred to by Hui & Seager (2002) as "atmospheric lensing") will weakly affect the transit light curve; the observer's distance to the lens is much greater than the lens' distance to the source, making angular deviations at the lens imperceivable to the observer. Diffraction is likely negligible for the same geometric reasons as with lensing and additionally unlikely due to the incoherent nature of starlight emitted from the stellar surface (Hui & Seager 2002). Inhomogenieties in cloud cover could organize in such a way to generate an oblate signal, however, these inhomogeneities would have to be large in extent to change the light curve appreciably (Hui & Seager 2002).

Companions of the exoplanet may induce an oblateness signal in the transit light curve. Moons, if close to their host but displaced in projection, would introduce higher moments in the projected shape of the exoplanet-moon combination. For the case of HD 189733b, Pont et al. (2007) ruled out orbiting moons larger than 0.8 $R_\oplus$ using a high precision *HST* light curve. Rings around an exoplanet, when viewed at favorable orientations, would likely yield a noticeable oblate-like signal in excess of that expected from rotation alone (Barnes & Fortney 2004, Ohta et al. 2009). Using an oblate model to fit light curves to ringed-exoplanets would therefore have an advantage over a purely spherical model. It is a reasonable assumption that tidally locked planets, such as is likely for HD 189733b, will have ring systems that are in the orbital plane as opposed to being at tilted viewing angles thereby making their detection less likely (Pont et al. 2007).

Finally, stellar effects may contribute to asymmetry or deformation of the transit light curve (see, e.g., Czesla et al. 2009). Stellar oscillations may be present in the data as fluctuations in the received flux at the 10^{-5} level over the course of a transit (Hui & Seager 2002). These fluctuations are likely to manifest themselves in random fashion resulting in "bumps" and "wiggles" impeding a measurement of oblateness or any other transit parameter (see Chapter 3). Stellar variability is present at the 1% level at optical wavelengths for the case of HD 189733b; this variability may be damaging for any attempt at accurate parameter estimation (Pont et al. 2007). Non-radial brightness profiles on the projected surface of the star would naturally lead to asymmetry in the transit light curve, the most dramatic examples of are as a result of star spots (as may have been seen with TrES-1; see Alonso et al. 2004, Rabus et al. 2009 for evidence for and Winn et al. 2007 for a non-detection of star spots). To combat the asymmetries resulting from spots and variability, we may make our observations farther into the infrared. For example, the effect of a star spot on the light curve would scale as the product of the ratio of intensities of the spot to stellar surface and the square of the ratio of spot radius and stellar radius (Silva 2003, Winn et al. 2007). At 8 micron (where the HD 189733b observations were made), a typical solar-like spot would induce an anomaly in the transit light curve at the 5×10^{-6} level [using temperatures and spot sizes from Steinegger et al. (1990)].

5.3 A numerical method for computing transit light curves of ellipsoidal exoplanets

Our goal in this section is to provide a method for fast, stable calculation of light curves of ellipsoidal exoplanets (so that the projection of the body on the observing plane traces an ellipse[5]). Ideally, our algorithm would be able to produce arbitrary precision light curves in times conducive for use with iterative parameter estimation codes. In particular, we utilize a Markov Chain Monte Carlo (MCMC) algorithm to produce posterior parameter distributions. Proper convergence of these posterior distributions may require on the order of millions of executions of the light curve calculation. Seager & Hui (2002) and Barnes & Fortney (2003) both provide details for oblate light curve calculation, however, the time required for computation was likely of little concern.

Figure 5-1 illustrates the basic geometry of the calculation. Here an ellipse (the projected ellipsoidal exoplanet), with minor-axis length b and major-axis length a, obscures some portion of a circular disk (the stellar disk). The projected-oblateness, $\hat{f}$, is given by

$$\hat{f} = (a - b)/a. \tag{5.14}$$

It is the projected-oblateness and the projected-obliquity, β, that we will be able to constrain from transit light curve data. Distances are normalized to the stellar radius, $R_\star$, such that the circular disk has unity radius. The semi-major axis of the ellipse is inclined by α from the line connecting the centers of the two ellipses[6]. The centers

---

[5] It can be shown that the projection of an ellipsoid onto any projective plane is an bounded by an ellipse.
See `http://www.geometrictools.com/Documentation/PerspectiveProjectionEllipsoid.pdf`

[6] The angle α is equivalent to the projected-obliquity β only for a central transit (i.e. the two angles are equal if the transit chord is along a diameter of the stellar disk).

are separated by a distance x. We assume the stellar disk has some brightness profile $I_\star(r, \theta)$ including, for example, a radial limb-darkening profile. The obscuring ellipse is assumed to be dark and opaque. The fractional flux deficit, $F(x; a, b, \alpha, I_\star)$, due to the obscuring ellipse is given by

$$F(x; a, b, \alpha, I_\star) = \frac{1}{F_0} \int_{\mathcal{E} \cap \mathcal{C}} I_\star(r, \theta)\ r\ dr\ d\theta \tag{5.15}$$

where F_0 is the total, unobscured flux and the integral is performed over the region bounded by the intersection of the ellipse and circle, denoted as $\mathcal{E} \cap \mathcal{C}$. Specializing, we consider radial brightness profiles describing stellar limb-darkening given by

$$\begin{aligned} I_\star(r, \theta) &= I_\star(r; u_1, u_2) \\ &= I_\star(1) \left[1 - u_1 \left(1 - \sqrt{1 - r^2}\right) - u_2 \left(1 - \sqrt{1 - r^2}\right)^2\right] \end{aligned} \tag{5.16}$$

for two "quadratic" limb-darkening parameters u_1 and u_2 (Claret 2000) and $F_0 / I_\star(1) = \pi(1 - 1/3u_1 - 1/6u_2)$. We denote the quadratic-profile fractional flux deficit by the form $F(x; a, b, \alpha, u_1, u_2)$. For $a = b = R_p/R_\star$, $F(x; a, u_1, u_2) \equiv F(x; a, a, \alpha, u_1, u_2)$ has a closed-form analytic solution in terms of elliptic integrals (Mandel & Agol 2002). For general a, b, u_1, u_2, the problem is more complicated since the intersection region $\mathcal{E} \cap \mathcal{C}$ may not be written in a closed form. Making matters worse, the integral over the brightness profile is also non-analytic.

If the disk has uniform brightness ($u_1 = u_2 = 0$ for the quadratic limb-darkened profile) the integral in Eqn. (5.15) may be calculated analytically by using the formula for the area of an elliptical chord $A(\theta_1, \theta_2, a, b)$:

$$A(\theta_1, \theta_2, a, b) = ab\left[(\theta_1 - \theta_2) - \sin(\theta_1 - \theta_2)\right]/2 \tag{5.17}$$

where the angles $\theta_{1,2}$, specifying the location of the ellipse-circle intersections (in Fig. 5-1, the points $\vec{x}_1$ and $\vec{x}_2$), are measured relative to the semi-major axis and from the ellipse center. To solve for $F(x; a, b, \alpha, 0, 0)$ we add (1) the area of the elliptical chord defined by the line connecting the points of intersection[7] and the curve bounding the ellipse that is internal to the stellar disk to (2) the area of the circular chord defined by the complement of this elliptical chord and $\mathcal{E} \cap \mathcal{C}$.

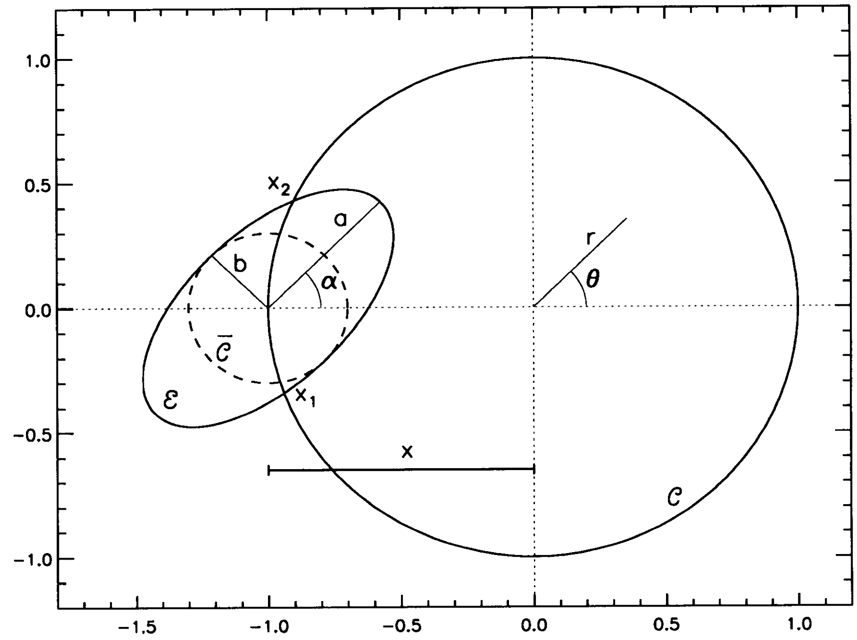

Figure 5-1 Geometrical configuration for the transit of an ellipsoidal planet across a spherical star. The deficit in flux due to the obscuring ellipse with semi-major axis a, semi-minor axis b and orientation α a distance x from the center of the star may be found by integrating the brightness profile of the star over the intersection region of the ellipse $\mathcal{E}$ and circle $\mathcal{C}$ (with intersection points $\vec{x}_{1,2}$). An analytic closed form solution for the flux deficit exists for the inscribed circle $\bar{\mathcal{C}}$ of radius b; only the intersection region less this circle, $\mathcal{E} \cap \mathcal{C}/\bar{\mathcal{C}}$ need be integrated numerically.

Routines exist (see, e.g., Hill 1994) to quickly compute the points of intersection of two ellipses. We find, however, that these algorithms are numerically stable only for a portion of the configuration space. In some cases no intersections are reported when using these routines for intersecting configurations. We choose, therefore, to sacrifice runtime (by a small factor) to ensure stability by finding the intersections of a "poly-line" representation of the obscuring ellipse and the disk as described as follows. We find the coordinates in the $x-y$ plane of points on the boundary of the

---

[7] In general, a circle and an ellipse may intersect at up to four distinct locations; for an exoplanet with small oblateness and $R_{\rm mean} \approx R_{\rm Jupiter}$ orbiting a star with $R_\star \approx R_\odot$, only two intersections (at most) are expected.

ellipse at N uniformly selected angles (from 0 to 2π) that are measured relative to the semi-major axis and from the ellipse center. These points in the plane are then connected by line segments. The routine accumulates approximate intersections of the ellipse and circle by (stably) determining the intersections of the circle and each line segment. We choose $N \approx 200$ to ensure adequate precision.

Having determined the points of intersection, we must now tackle the integral given by Eqn. (5.15) for non-zero u_1 and u_2. We could perform the two-dimensional integral over $\mathcal{E} \cap \mathcal{C}$ [as was done by Seager & Hui (2002)] or integrate in the radial direction, calculating the intersections of circle and ellipse at each integration step [as was done by Barnes & Fortney (2003)]. For our application we preferred the former. We found using standard integration techniques [see, e.g., Press et al. (2007) Chap. 4] over the integration region to be too slow for our application.

Instead, we choose to use Monte Carlo integration [see, e.g., Press et al. (2007)]. In Monte Carol integration, the integral $\int_{\mathcal{A}} f(x,y)\, dx\, dy$ over the region $\mathcal{A}$ is calculated by sampling the function $f(x,y)$ at N uniformly distributed random points in a region $\mathcal{R}$ that covers $\mathcal{A}$ [for sample points (x,y) in $\mathcal{R}$ but not in $\mathcal{A}$ select $f(x,y) = 0$]. The fundamental theorem of Monte Carlo integration is then expressed as

$$\int_{\mathcal{A}} f(x,y)\, dx\, dy \quad \approx \quad \mathrm{Area}[\mathcal{R}] \left[\langle f \rangle \pm \sqrt{\frac{\mathrm{Var}(f)}{N}} \right] \tag{5.18}$$

where means ($\langle\cdot\rangle$) are calculated over the N randomly sampled points (Press et al. 2007). In general, Monte Carlo integration is reserved for integrals of high dimension due to its slow convergence ($1/\sqrt{N}$) compared to more traditional two-dimensional integration methods. For our purposes, Monte Carlo integration reduces the difficulty associated with bounding the intersection region $\mathcal{E} \cap \mathcal{C}$. Some improvement in convergence ($\sim 1/N$) may be obtained by sampling $\mathcal{R}$ "quasi-randomly" as opposed to randomly, using a low-discrepancy random sequence such as the Sobol' sequence (Sobol' & Shukhman 1995, Press et al. 2007). The Sobol' sequence is a sequence of uniformly distributed values, s_i for $i \geq 0$, on the unit interval such that a given sample at index I is *maximally* distant from all samples $i < I$. As we shall show, by carefully tailoring our problem, *quasi*-Monte Carlo integration yields oblate light curves with speed and efficiency.

To this end, we first simplify the computational effort involved in the integral in Eqn. (5.15) when using a quadratic limb-darkened profile [Eqn. (5.16)]:

$$\begin{aligned} F(x;a,b,\alpha,u_1,u_2) & \times \pi\left(1-\frac{u_1}{3}-\frac{u_2}{6}\right) \\ &= \int_{\mathcal{E}\cap\mathcal{C}} 1-u_1\left(1-\sqrt{1-r^2}\right)-u_2\left(1-\sqrt{1-r^2}\right)^2 \; r\,dr\,d\theta \\ &= (1-u_1-u_2)\left(\int_{\mathcal{E}\cap\mathcal{C}} r\,dr\,d\theta\right) \\ &\qquad + \int_{\mathcal{E}\cap\mathcal{C}} \left[(u_1+2u_2)\sqrt{1-r^2}-u_2\left(1-r^2\right)\right] \; r\,dr\,d\theta \\ &= (1-u_1-u_2)\times F(x;a,b,\alpha,0,0) \\ &\qquad + \int_{\mathcal{E}\cap\mathcal{C}} \hat{I}_\star(r;u_1,u_2)\; r\,dr\,d\theta \end{aligned} \tag{5.19}$$

where $F(x;a,b,\alpha,0,0)$ is the uniform-brightness solution (calculated as described above) and $\hat{I}_\star(r;u_1,u_2) = (u_1+2u_2)\sqrt{1-r^2}-u_2\left(1-r^2\right)$ describes the non-uniform component of the brightening profile, with $\hat{I}_\star(r=1;u_1,u_2)=0$ at the stellar "limb". As a result of this reorganization, the only non-trivial integral is of a function $[\hat{I}_\star(r;u_1,u_2)]$ that has low variance over any bounding region $\mathcal{R}$ covering the limb, making the absolute error in the Monte Carlo integration smaller for a fixed N as compared to integrating $I_\star$ in full.

Second, we recognize that we only need to integrate over the intersection region excluding the circle $\bar{\mathcal{C}}$ of radius b inscribed in $\mathcal{E}$ [see Fig. (5-1)]. We denote this smaller region as $\mathcal{E}\cap\mathcal{C}/\bar{\mathcal{C}}$. The integral over $\bar{\mathcal{C}}$ may be computed instead by utilizing the analytic solutions, $F(x;b,u_1,u_2)$, by Mandel & Agol (2002). Then Eqn. (5.19) may be reduced to

$$\begin{aligned} F(x;a,b,\alpha,u_1,u_2) & \times \pi\left(1-\frac{u_1}{3}-\frac{u_2}{6}\right) \\ &= (1-u_1-u_2)\times\left[F(x;a,b,\alpha,0,0)-F(x;b,0,0)\right] \\ &\qquad +F(x;b,u_1,u_2) \end{aligned}$$

$$+\int_{\mathcal{E}\cap\mathcal{C}/\bar{\mathcal{C}}} \hat{I}_\star(r;u_1,u_2)\; r\; dr\; d\theta. \tag{5.20}$$

The only non-trivial component left to compute for our total solution, for which we shall use quasi-Monte Carlo integration to solve, is the final integral in the sum, $\mathcal{I} \equiv \int_{\mathcal{E}\cap\mathcal{C}/\bar{\mathcal{C}}} \hat{I}_\star(r;u_1,u_2)\; r\; dr\; d\theta$. For slightly oblate exoplanets or weakly limb-darkened brightness profiles, the contribution of this integral in the total flux deficit is small compared to the remaining terms in Eqn. (5.20). To be exact, it is easy to show the absolute contribution of this term to the flux deficit is bounded by

$$\begin{aligned} \frac{1}{\pi}\left(1-\frac{u_1}{3}-\frac{u_2}{6}\right)^{-1}\mathcal{I} &\leq \frac{u_1+2u_2}{1-1/3u_1-1/6u_2}\times\left(ab-b^2\right) \\ &= ab\hat{f}\frac{u_1+2u_2}{1-1/3u_1-1/6u_2}. \end{aligned} \tag{5.21}$$

For comparison, the depth of the flux deficit in the absence of limb-darkening is ab.

Now we may focus on calculating $\mathcal{I}$ via quasi-Monte Carlo integration. We must select a covering region, $\mathcal{R} \supset \mathcal{E}\cap\mathcal{C}/\bar{\mathcal{C}}$, that can be sampled easily in the Monte Carlo technique ($\mathcal{E}\cap\mathcal{C}/\bar{\mathcal{C}}$ cannot). Here, easily sampled means that we may take a two dimensional Sobol' sequence, uniform in $[0,1]\times[0,1]$, and analytically transform it such that the transformed points uniformly sample our chosen covering region without omitting any sequence points. Rectangular regions are easily sampled via a Sobol' sequence, however, are poorly shaped to efficiently bound $\mathcal{E}\cap\mathcal{C}/\bar{\mathcal{C}}$, especially for the crucial phases of ingress or egress (where the boundaries of ellipse and circle have greater than one intersection point). A more useful sampling region may be found via equal area coordinate transformations. A uniform sampling on the unit square ($[0,1]\times[0,1]$) may be mapped to a uniform sampling over a region bounded by an elliptical annular sector (see Appendix A for details). An elliptical annular sector is the region bounded between two "concentric" ellipses (having equal axis ratio, center and orientation with the obscuring ellipse) with semi-major axes $a_1 < a_2$ and by the rays (emanating from the common center) at angles θ_1, θ_2 relative to the semi-major axis. We now only need to provide formula for the specific values of θ_1, θ_2, a_1 and a_2 for use in the integration routine as functions of the parameters x, a, b, and α.

We specify $\theta_{1,2}$ and $a_{1,2}$ with respect to the following cases [refer to Figure (5-2) for illustrations of these phases]:

(I) One or fewer intersections between circle and ellipse.

(a) $x > 1$; The obscuring ellipse is external to the circle, $\mathcal{I} = 0$.

(b) $x < 1$; The obscuring ellipse is properly contained in the circle;
$\theta_1 = 0$, $\theta_2 = 2\pi$, $a_1 = b$, $a_2 = a$.

(II) Two points of intersection between circle and ellipse.

The circle and ellipse intersect at x-coordinates $x_{1,2}$ at angles $\theta'_{1,2}$ from the x-axis.

(a) $x > 1$; For each intersection x-coordinate $x' = x_i$, angle $\theta' = \theta'_i$ and final, solution angle $\theta = \theta_i$:

(1) $|x| > |1/x'|$; The line connecting the point of intersection and the ellipse center [at (x,0)] intersects the circle exactly once: $\theta = \theta' - \alpha$.

(2) $|x| < |1/x'|$; The line connecting the point of intersection and the ellipse center [at (x,0)] intersects the circle twice: use $\theta = \hat{\theta} - \alpha$ where the angle $\hat{\theta} = \tan^{-1} 1/\sqrt{1 - x^2}$ defines the line segment that connects the ellipse center and the point of tangency on the circle.

If a' is the semi-major axis of the concentric ellipse "kissing" the circle[8] then $a_1 = \max(b, a')$, $a_2 = a$.

(b) $x < 1$; Use the angle θ_b defined by the intersection points of the inscribed circle with radius b ($\bar{\mathcal{C}}$) and the stellar disk ($\mathcal{C}$): $\theta_{1,2} = \pm\theta_b - \alpha$, $a_1 = b$, $a_2 = a$.

The integration region bounded by the elliptical annular sector as defined above ensures that approximately 50% of its area covers the desired integration region $(\mathcal{E} \cap \mathcal{C}/\bar{\mathcal{C}})$ for all possible parameters x, a, b, and α. As a result of our reorganization and bounding, the time to compute $F(x; a, b, \alpha, u_1, u_2)$ to one ppm precision during ingress for a typical example takes on average 0.5 ms[9]. This completes the description of our numerical algorithm. Refer to Seager & Hui (2002) or Barnes &

---

[8] Ellipses that are "kissing" intersect at exactly one point. The "kissing" ellipse semi-major axis a' is found via a (quick) linear search.

[9] In this test: $x = 1$, $a = 0.155$, $b = 0.148$, $u_1 = 0.2$, $u_2 = 0.3$, $\alpha = 0.5$. These parameters correspond to a strongly oblate ($\hat{f} = 0.05$) HD 189733b in transit. Coded in C++ and executed on a 2.6 GHz Intel Core 2 Duo MacBook Pro.

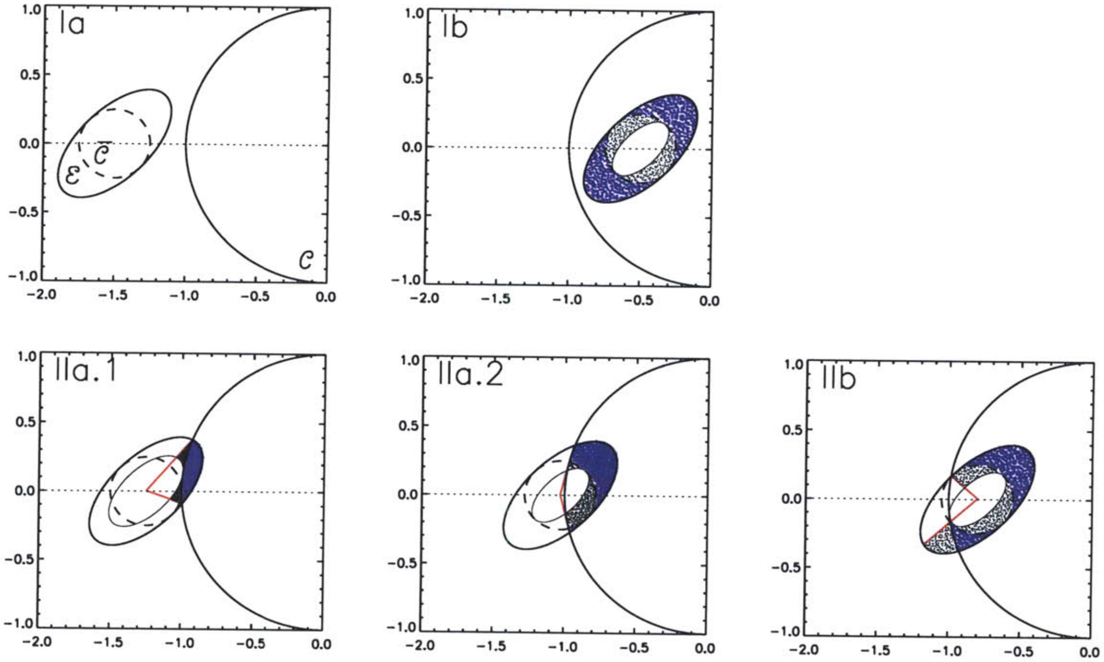

Figure 5-2 Quasi-Monte Carlo integration of the non-trivial component of the total flux deficit for the stellar transit of an oblate planet. These figures demonstrate graphically the transit phases and quasi-Monte Carlo integration regions as described in the § 5.3 to evaluate the non-trivial integral $\mathcal{I}$ [see Eqn. (5.20)]. The labels in the upper left hand corners of each figure correspond to those in the text. In each figure, the inner ellipse with semi-major axis a_1 (thin black line) gives the inner boundary of the elliptical annular sector. The red rays indicate the angular extent of the sector (with angles $\theta_{1,2}$ relative to the semi-major axis). The black and blue points are 1000 uniformly distributed Sobol' points in the elliptical annular sector. The blue points are those which fall in the integration region $\mathcal{E} \cap \mathcal{C}/\bar{\mathcal{C}}$.

Fortney (2003) for qualitative descriptions of transit light curves for oblate exoplanets. We verified our computed light curves exactly match those computed for their respective papers.

5.4 Spitzer transits of HD 189733b: An oblate analysis

In this section, we present the results of an analysis of seven high precision transit light curves of HD 189733b, for the first time constraining parameters relating to oblateness ($\hat{f}$ and β). All seven light curves were observed in a common mode using the 8 micron channel of the InfraRed Array Camera (IRAC) instrument (Fazio et al 2004) onboard *Spitzer*. One observation is the subject of the work by Knutson et al. (2007a) whereby the first light curve indicating a changing "phase" of an exoplanet was collected. The remaining six observations are part of an observing program by E. Agol and others [see Agol et al. (2007)] aiming to detect additional planets by analyzing times of mid-transit (Holman & Murray 2005, Agol et al. 2005). We find each individual observation yields a light curve with a precision of roughly 500 ppm, making these light curves some of the most precise currently available [with mid-transit times precise to ~ 3 seconds (Agol et al. 2009, and this work)]. While other space-based observations with *HST* can offer greater precision (see, for example, Brown et al. 2001, Knutson et al. 2007b, and Pont et al. 2008), light curves observed in optical wavelengths are subject to increased stellar variability, star spots (see § 5.2.3) and strong limb-darkening whose combined effects may suppress our ability to detect oblateness. For this reason, we consider the seven *Spitzer* observations as the most capable for constraining oblateness for HD 189733b (or any other transiting exoplanet).

We do not expect to actually *measure* the parameters $\hat{f}$ or β for HD 189733b with this data, for the following reason. If we make the reasonable assumption that HD 189733b is synchronously rotating (see § 5.2.1) with a rotational period of $P_{\text{rot}} = 2.2$

days, the expected *true* oblateness, in the absence of any tidal effects, is $f \approx 6 \times 10^{-3}$. In Figure (5-3), we plot the excess signal after fitting and subtracting a spherical model ($\hat{f} = 0$) to the transit light curve of a $\hat{f} = 6 \times 10^{-3}$, zero projected-obliquity transit of a hypothetical HD 189733b [taking other relevant transit parameters from Torres et al. (2008)]. The amplitude of the excess signal is roughly 2×10^{-6} in relative flux. This can be compared to the $\sim 2 \times 10^{-4}$ precision of the combined data comprised of the seven *Spitzer* light curves binned to a cadence of 30 seconds (see below). Nevertheless, at this precision we may make physically relevant constraints on the projected-oblateness $\hat{f}$ for HD 189733b. Namely, we show (below) that projected-oblatenesses consistent with Jupiter ($\hat{f} = 0.06487$) or Saturn ($\hat{f} = 0.09796$) are effectively ruled out by this data set. We have plotted, in Fig. (5-3), the excess signal after fitting a spherical model to the transit light curve of a hypothetical HD 189733b with an oblateness comparable to that measured for Saturn at an obliquity of 45°. We also plot the expected noise level at 60s binning for our seven light curves with this excess, visually demonstrating that such an oblateness signature should be easily excluded by the data.

5.4.1 Observations and data reduction

We reanalyzed all the data from the seven *Spitzer* observations from the raw images. (all publicly available for download by using the *Spitzer* data request software "Leopard[10]") All observations were performed with the 8 micron channel of IRAC instrument in "subarray" mode. Here, only a 32×32 pixel portion of the detector is recorded. Images are taken in quick succession at a cadence of 0.4 s (0.32 s total integration time). The data are packaged into post-calibration[11] FITS files of 64 images each ($\sim$ 26s total observation time for each 64). Roughly 500 files per observation cover the transit of HD 189733b [$T = 1.41$ hr, $\tau = 0.41$ hr in the notation of Carter et al. (2008)] for a total of $\sim$ 32000 images collected for each transit event. In addition to excellent time resolution, the subarray mode avoids issues of saturation with the

---

[10] `http://ssc.spitzer.caltech.edu/propkit/spot/`

[11] As performed by the pipeline at the *Spitzer* Science Center.

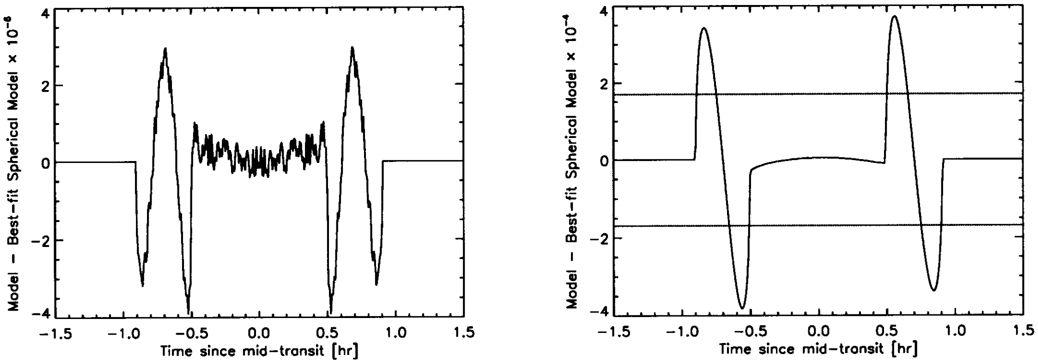

Figure 5-3 Signals of oblateness for hypothetical transit light curve models of HD 189733b. The left panel plots the excess signal after fitting a spherical model transit light curve ($\hat{f} = 0$) to the transit light curve expected for a hypothetical synchronously rotating, tidally uninfluenced HD 189733b. Here, theory predicts oblateness $f \sim 0.006$ and zero obliquity (see § 5.2). The right panel gives the analogous excess signal for a Saturn-like oblateness. The dotted lines mark the level of noise expected for the seven *Spitzer* observations when binned to 60 s (the spacing between dots is 60 s). The high frequency "noise" seen in the light curve in the left panel is an artifact of the random noise induced via the Monte Carlo integration technique (with absolute precision set to 10^{-6}; see § 5.3).

brightness of HD 189733 ($V = 7.7$, Høg et al. 2000).

The sub-second time sampling is overkill for the analysis of an effect that is evident on $\sim \tau$ timescales (Seager & Hui 2002, Barnes & Fortney 2003). As such, to reduce the data volume to a manageable level and, as an added bonus, correct for transient pixel effects (for example, cosmic rays) we generate mean images of each set of 64 as described as follows. First, we "stack" the 64 images and take the median value for each pixel coordinate. Second, we flag values 3.5-sigma from the median value for each image and each pixel in the 32×32 array. A mean image is then constructed for each 64 image data cube, excluding flagged pixels in the average.

We perform aperture photometry on the resulting mean images, with an aperture size 3.5-4.5 pixels in radius around HD 189733 (depending on observation). Several rectangular apertures are placed in regions away from HD 189733 and a dim companion to produce a median estimate of the sky background, which is subtracted from the circular aperture sum for each mean image. At this point, our reductions result in a sequence of flux measurements indexed by the time stamps giving in the image headers. These time-series are reproduced, without systematic correction, in Figure (5-4).

The most evident systematic is the "ramp" effect, a rise in the observed flux with time, likely due to charge-trapping in the detector. This effect has been seen in other similar campaigns (Knutson et al. 2007a, Gillon et al. 2007, Nutzman et al. 2008) and may be modeled by a multiplicative correction that is a function of time. We choose to use a quadratic function of time as described below, although more complicated models have been utilized (see, e.g., Nutzman et al. 2008).

The seven observations span 268 transits of HD 189733b, which occur every 2.218 days (Torres et al. 2008). In particular, if epoch zero is arbitrarily chosen to coincide with the transit on June, 30 2007 (UT), then the observations occurred at epochs -110, 0, 1, 51, 62, 157, and 158. In each light curve, we clip the most strongly varying portion of the "ramp" systematic and additionally the majority of the Knutson et al. (2007a) observation that was collected after transit [which follows HD 189733b to the secondary transit (occultation)]. The final time-series [shown after corrections in Fig (5-5)] have 410, 491, 407, 386, 451, 402, 451 data points for the epochs listed above, respectively.

5.4.2 The combined transit light curve

The oblateness analysis is performed on the combined light curve comprised of all seven light curves (as described below). In combining these time-series we have implicitly assumed that the parameters $\hat{f}$ and β are *constant* across all transit epochs in our sample. In § 5.5 we consider analyses where this restriction is not applied.

Prior to combining the seven light curves, we use a global *oblate* model fit to the light curves to determine the times of mid-transit, the coefficients of the ramp-correction and the out-of-transit flux level for each curve. We model the multiplicative ramp-correction as

$$C_{\mathrm{ramp}}(t; c_0, c_1; t_0) = 1 + c_0(t - t_0) + c_1(t - t_0)^2 \tag{5.22}$$

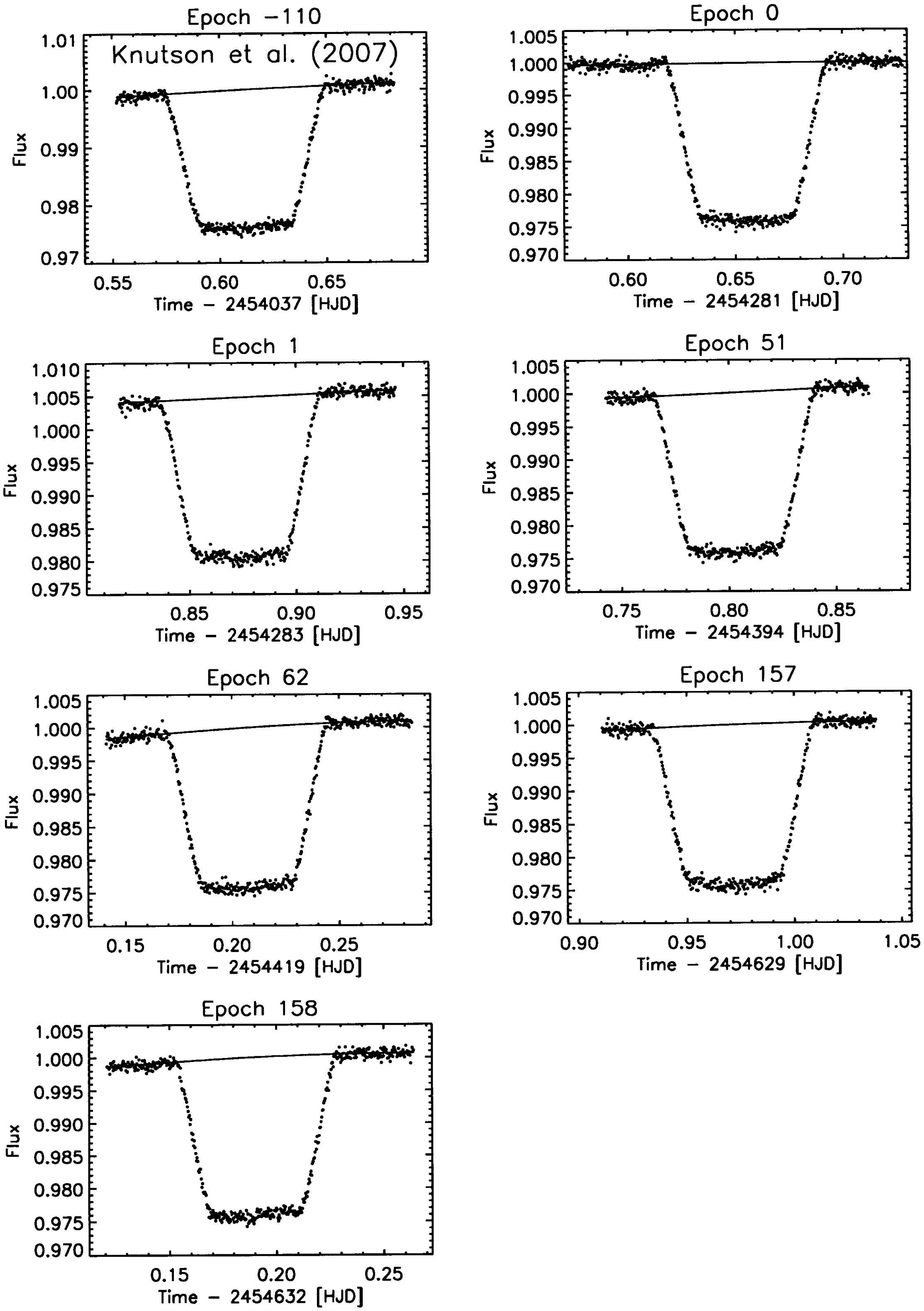

Figure 5-4 Light curves from seven *Spitzer* observations of HD 189733b. Each figure gives the aperture summed flux from a continuous observation of a single transit of HD 189733 by HD 189733b. The blue lines give the best-fit correction [see Eqn. (5.22)] for the "ramp" systematic for each individual light curve. See § 5.4 for reduction details.

where c_0 and c_1 are the variable model parameters, and t_0 is fixed to some time near mid-transit for each light curve. In addition to the ramp-correction, each light curve is described by an out-of-transit flux level, F_0 and a mid-transit time T_0. All seven transit light curves are fitted to a common transit model parameterized by the mean radius ratio $R_{\rm mean}/R_\star$, inclination i, normalized semi-major orbital radius $a/R_\star$ (not to be confused with the projected semi-major axis of the exoplanet ellipsoid), projected-oblateness $\hat{f}$, projected-obliquity β, and a single limb-darkening coefficient u_1[12].

In fitting, we use the standard χ^2 statistic in which we have assumed the data are affected in an additive sense by Gaussian, white noise. In other words, we assess χ^2 defined as

$$\chi^2 = \sum_{o=1}^{7} \sum_{i=0}^{N_o} \left[\frac{F_i^o(\mathrm{obs}) - F_i^o(\mathrm{calc})}{\sigma^o} \right]^2 \tag{5.23}$$

where o indexes the observations, N_o is the number of data points in observation o, $F_i^o(\mathrm{obs})$ is a measurement of the relative flux of HD 189733 during observation o, $F_i^o(\mathrm{calc})$ is the calculated model describing this flux measurement, and σ^o is the uncertainty in flux measurement during observation o. We estimate uncertainties by calculating the root-mean-square flux out-of-transit after the transit (where the ramp effect is less-pronounced) for each transit light curve. Epochs -110, 0, 1, 51, 62, 157, and 158 have uncertainties in relative flux of 597 ppm, 567 ppm, 576 ppm, 536 ppm, 541 ppm, 520 ppm, and 540 ppm, respectively. Photon noise is the dominate noise type for these observations, with an expected photon-limited precision of $\approx$460 ppm.

We minimize χ^2 using the AMOEBA routine (Press et al. 2007) so as to solve for the model parameters. We find a global limb darkening parameter $u_1 = 0.113 \pm 0.011$ [consistent with that estimated by Agol et al. (2009)] and mid-transit times precise to $\sim$ 3s. We find no evidence for transit-timing variations greater than $\sim$ 6s in amplitude for the seven mid-transit times [in agreement with a preliminary analysis of five transits by Agol et al. (2009)].

---

[12] Here, we fix $u_2 = 0$ for simplicity. Limb-darkening is expected and is found to be small at 8 micron (Agol et al. 2009).

After finding the best-fit solution, we correct each light curve by dividing each by its estimated ramp correction [Eqn. (5.22)] and out-of-transit flux level. We then construct the combined light by shifting the time indices in each time-series by the respective midtransit time solutions such that midtransit occurs at $t = 0$ and, finally, merge the individual time-series into one. The resulting light curve, containing 2999 points, is binned by a factor of six to average out any "red" or time-correlated noise and to reduce data volume. The final combined light curve contains 512 data points. Each data point corresponds to an integration time of 30s, such that we have $\approx$50 samples during the phases of ingress or egress. The combined light curve has a precision of 240 ppm or 1.28 times the photon-limited noise expectation. The light curves, corrected for the ramp systematics, are shown in Fig. (5-5). The final combined light curve is shown in Fig. (5-6). Two unexpected features, occurring before ingress and near mid-transit, are evident in the combined light curve.

5.4.3 Oblateness constraints

In this section we present the methods and results from an analysis of the combined light curve of the seven observations. We employ a Markov Chain Monte Carlo (MCMC) algorithm[13] for the purpose of determining posterior parameter distributions of the following parameters: $R_{\rm mean}/R_\star$, inclination (i), normalized semi-major orbital axis ($a/R_\star$), out-of-transit flux level (F_0), the time of mid-transit (T_0), the projected oblateness ($\hat{f}$), the projected obliquity (β) and a single limb-darkening parameter (u_1). We choose $R_{\rm mean}/R_\star$ as the parameter describing the transit depth as opposed to, for example, $R_{\rm equatorial}$ as the former is nearly uncorrelated with $\hat{f}$ whereas the latter is strongly correlated. The level of limb-darkening at 8 micron only weakly suppresses that oblateness signal, however, we include limb-darkening in our model for completeness. The orbital parameters i and $a/R_\star$ are expected to be correlated with $\hat{f}$ (Barnes & Fortney 2003).

In the MCMC, the jump-transition probability is given by the likelihood propor-

---

[13] For background on the MCMC method, see Gregory (2005), and for example applications to transit light curves, see e.g. Holman et al. (2006), Winn et al. (2007), or Burke et al. (2007).

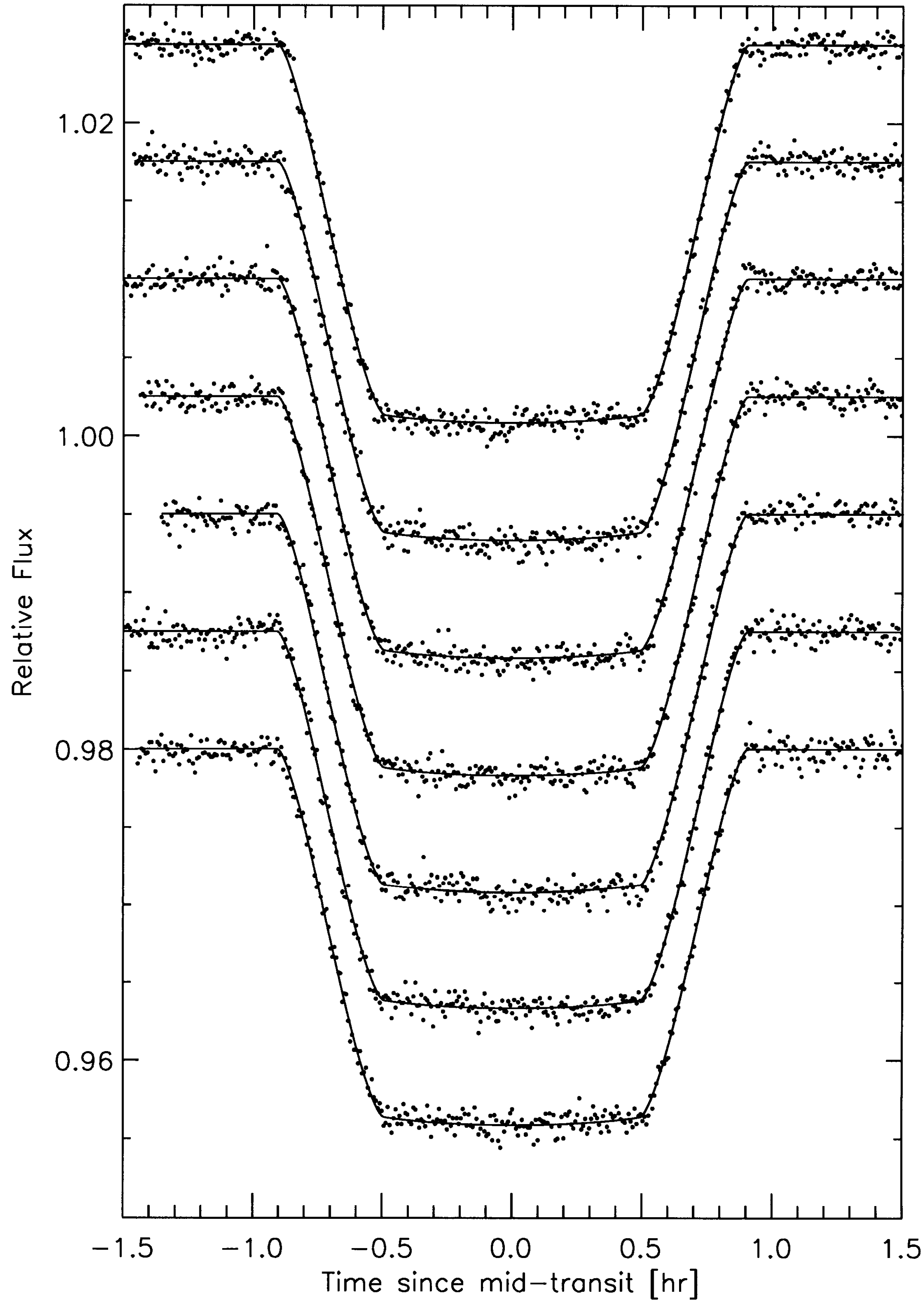

Figure 5-5 Systematic corrected transit light curves of seven *Spitzer* observations of HD 189733b. The "ramp" corrected and mid-transit aligned light curves are offset in flux for comparison. The blue line gives the best fit oblate model. See § 5.4.2 for details.

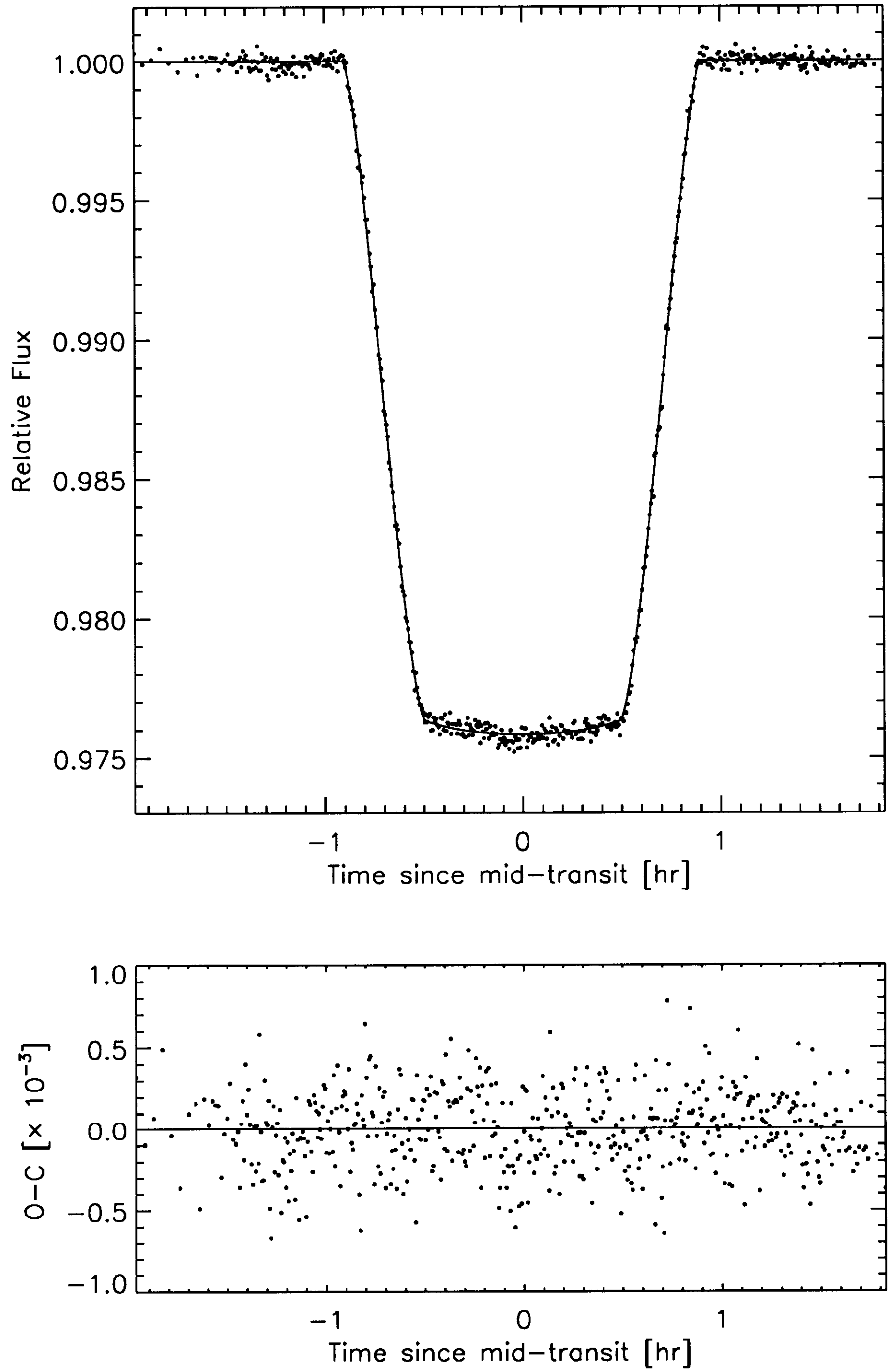

Figure 5-6 Combined transit light curve and residuals of seven *Spitzer* observations of HD 189733. The seven transits [as seen post-systematic corrections in Fig. (5-5)] are merged into one time-series and then binned (such that the final cadence is 30 s) to produce the light curve in the top panel. The bottom panel shows the residuals after subtracting the best-fit oblate model (blue line in top panel) from the combined light curve. The residuals have a root-mean-square of 241 parts-per-million.

tional to $\exp(-\chi^2/2)$ with

$$\chi^2 = \sum_{i=1}^{512} \left[\frac{F_i(\mathrm{obs}) - F_i(\mathrm{calc})}{\sigma} \right]^2 \tag{5.24}$$

in analogy with Eqn. (5.23) for the combined light curve (where now $\sigma = 2.41 \times 10^{-4}$ is the flux uncertainty for the combined, binned light curve). We use Gibbs sampling in the construction of a chain of one million links. For a Gibbs sampler, at each jump-attempt we perturb a randomly selected parameter. We select individual parameter jump sizes so that the fraction of jumps accepted by a Metropolis-Hasting condition is approximately 40% for each parameter. We trim the first 10% of the resulting chain to accommodate a period of "burn in," resulting in a chain comprised of 9×10^5 links.

We report for each parameter the median value and the 15.85% and 84.15% levels of the cumulative distribution. Table (5.2) collects these results. Curiously, our estimate of the normalized mean radius is consistent with the normalized planetary radius listed in Table 2 of the paper by Torres et al. (2009) [completed on an analysis of the single Knutson et al. (2007) light curve], but different by nearly 7 of our sigma from the result quoted by Agol et al. (2009) after a preliminary global analysis of five of the *Spitzer* light curves. Our limb-darkening coefficient is consistent with that found by Agol et al. (2009). Some of this discrepancy may be due to the additional freedom in determining the oblateness parameters (and as a result of their effect on the normalized impact parameter). Agol et al. (2009) did not provide estimates of other transit parameters that could be used to help elucidate this discrepancy.

Figure (5-7) presents constraints on the oblateness parameters via confidence curves in the $\hat{f}$-β plane. The projected-oblateness $\hat{f}$ is more tightly constrained, relative to zero projected-oblateness, for projected obliquities $\beta \sim \pi/4$ becoming less constrained in a symmetric fashion for $|\beta - \pi/4| > 0$. We determine the 68% and 95% confidence regions by the following technique. We divide the projected obliquity range, $[0, \pi/2]$, into M uniform bins. For chain links with β-coordinate falling in a specific projected-obliquity bin, we sort the associated $\hat{f}$-coordinates and determine $\hat{f}'$ such that 68% (95%) of the $\hat{f}$ in the bin are less than $\hat{f}'$. In Fig. (5-7), we plot,

Table 5.2. Parameters for HD 189733b and the combined *Spitzer* transit light curve

Parameter	Median	Uncertainty
Global transit parameters:		
R_{mean}	0.15498	± 0.00009
Orbital Inclination, i [degree]	85.75	± 0.03
$a/R_\star$	8.929	± 0.022
Projected Oblateness $\hat{f}$	0	< 0.034 (95% conf.)
Projected Obliquity β	unconstrained	
Light curve parameters:		
Limb-darkening parameter u_1	0.113	$\pm$ 0.011
Mid-transit time T_0 [s]	0.0	± 1.3
Out-of-transit level F_0	0.99999	± 0.00002
Derived parameters:		
Rotational Period [day]^a	2.37^b	> 0.88 (95% conf.)
$J_2{}^c$	0	$< 2.1 \times 10^{-2}$ (95% conf.)

Note. — For the light curve shown in Fig. (5-6). (a) Calculated using Eqn. (5.10) with $R_\star = 0.756\ R_\odot$, $M_p = 1.144\ M_{\mathrm{Jupiter}}$ (Torres et al. 2008). Assuming HD 189733b is unaffected by tides. Assuming HD 189733b has a uniform mass distribution. (b) The median estimate of P_{rot} is strongly affected by our *a priori* assumptions on $\hat{f}$. Here we have assumed a uniform prior on $\hat{f}$. (c) Calculated using Eqn. (5.7) assuming $P_{\mathrm{rot}} = 2.218573$ day (Torres et al. 2008).

in blue, confidence curves at 68% and 95% confidence levels. Additionally, we plot as black points the first 5000 points $(\hat{f}, \beta)$ from a random reindexing (shuffle) of the Markov chain.

The oblateness parameters are consistent with the theoretical expectation for a synchronously rotating, tidally unmodified HD 189733b (where $\hat{f} \approx f \approx 6 \times 10^{-3}$ and $\beta \approx 0$). The red, dashed line marks $\hat{f} = 6 \times 10^{-3}$ in Fig. (5-7). We also mark the oblatenesses as measured for the Solar System planets Saturn, Jupiter and Uranus. The data have ruled out a projected-oblateness comparable to that found for Saturn at greater than 95% at all projected obliquities. Projected-oblateness similar to that seen with Jupiter is ruled out at 95% confidence for all projected obliquities $\beta \gtrsim \pi/30$. Projected-obliquities near $\pi/4$ and $\hat{f}$ similar to that measured for Uranus are also ruled out at greater than 95% confidence. See Table (5.1) for the specific values of f used in Fig. (5-7).

We use Eqn. (5.10) and the results from the MCMC analysis to derive the rotational period, $P_{\rm rot}$, of HD 189733b under the following assumptions. First, we assumed the projected-oblateness is equivalent to the true oblateness (i.e., $\hat{f} = f$, $\beta = \theta$). Second, we assumed the measured oblateness is as a result of rotation alone (i.e., tides are negligible). Third, we have assumed that HD 189733b has a uniform mass distribution. We used in our analysis the values for the stellar radius of HD 189733 and planetary mass of HD 189733b as provided by the tables in the work by Torres et al. (2008). The median value, $P_{\rm rot} = 2.37$ days, is similar to the expected synchronous rotation period of 2.12 days, however, the posterior distribution in $P_{\rm rot}$ is significantly affected by the assumption that $\hat{f}$ is *a priori* uniformly distributed. We can, nevertheless, constrain the planet to be rotating slower than once every 0.88 days ($\sim$ 20 hours) at 95% confidence. It is likely that HD 189733b is more centrally condensed than a uniform model would allow. The derived rotational period, at a fixed oblateness, reduces as more mass is relocated to the core in our model (Murray & Dermott 2000).

If, conversely, we assume HD 189733b is synchronously rotating with $P_{\rm rot} = 2.218573$ days, we may place constraints on J_2 via Eqn. (5.7). In particular, we find,

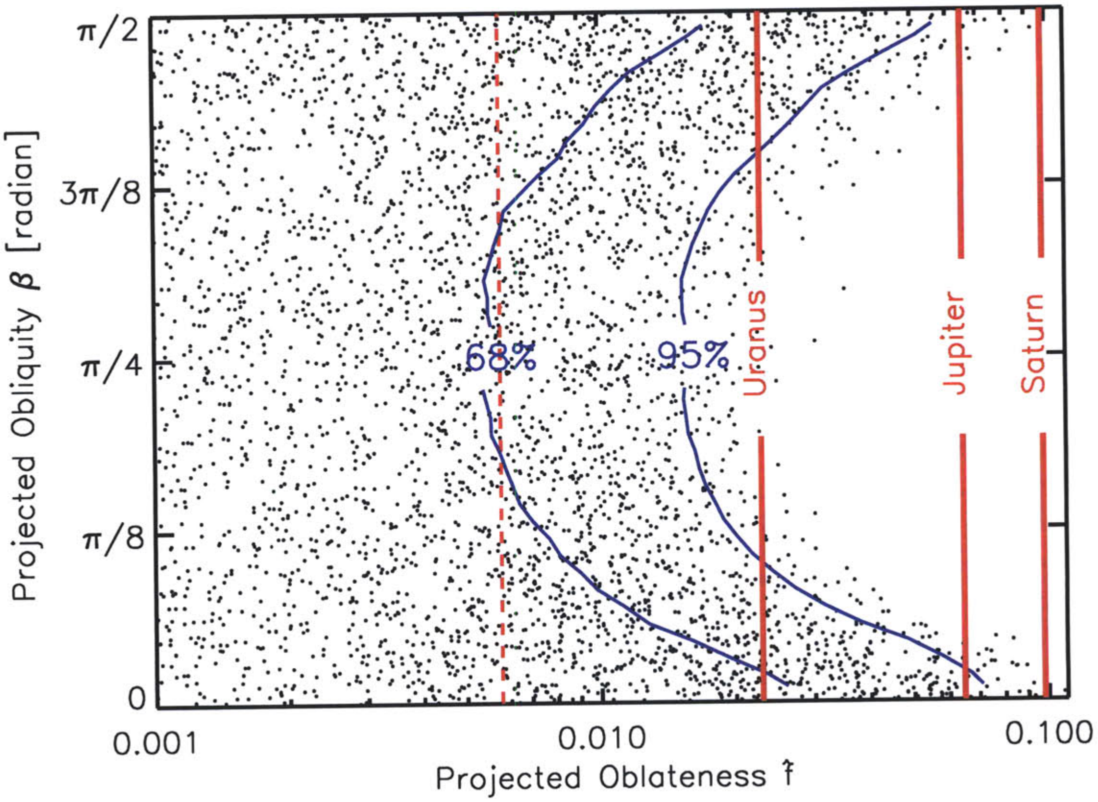

Figure 5-7 Oblateness constraints for HD 189733b based upon seven *Spitzer* transit observations. The solid blue curves bound the regions containing 68% or 95% of the Markov Chain Monte Carlo samples in the projected oblateness-obliquity ($\hat{f} - \beta$) plane. The black points represent a shuffled and "sifted" sample of the full Markov Chain of 9×10^5 links. The red dashed line marks the oblateness of a theoretical synchronously rotating, tidally uninfluenced HD 189733b; $f = 0.006$ (see § 5.2 for details). The solid red lines mark the oblatenesses measured for Jupiter, Saturn and Uranus.

with this assumption, that J_2 is smaller than 2.1×10^{-2} at 95% confidence [Jupiter has $J_2 = 1.47 \times 10^{-2}$ for reference; Hubbard et al. (1984)]. In Figure (5-8) we plot the posterior distributions of $P_{\rm rot}$ and J_2 as calculated with the above assumptions. We have indicated the location of the median and 68% confidence interval with vertical lines.

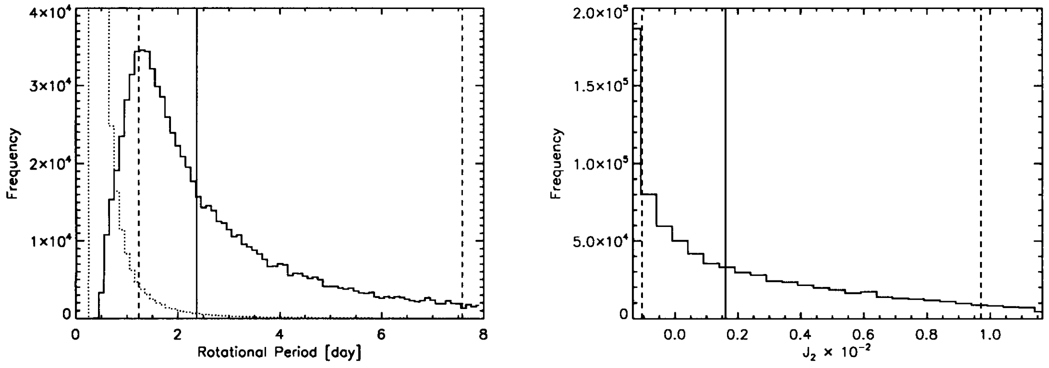

Figure 5-8 Posterior distributions for the rotational period, $P_{\rm rot}$, and the second spherical moment of the mass distribution, J_2, of HD 189733b based upon seven *Spitzer* transit observations. The solid black lines mark the median value of the respective parameter, while the dashed lines mark the 15.85% or 84.15% confidence intervals. The dotted curve over-plotted in the left panel marks the expectation in posterior distribution for $P_{\rm rot}$ for a uniformly distributed posterior distribution in $\hat{f}$ for $0 < \hat{f} < f_{\rm breakup}$ where $f_{\rm breakup}$ is the rotational breakup limit defined in § 5.2. See § 5.4.3 for more details.

5.5 Discussion

In this chapter, we have made the first attempt to constrain the projected shape of a transiting exoplanet: we have placed interesting bounds on the shape of HD 189733b. To this end, we developed an algorithm capable of producing high precision light curves of ellipsoidal exoplanets transiting their stellar host. The algorithm, involving analytical and semi-analytical components along with quasi-Monte Carlo integration, can yield part-per-million light curves in fractions of seconds (see § 5.3). Such speed allowed for proper determination of parameter estimates and uncertainties via a Markov Chain Monte Carlo (MCMC) method.

We applied MCMC to determine both traditional transit parameters and two parameters describing the elliptical perturbation to the projected shape of HD 189733b. In particular, we were able to constrain both the projected-oblateness $\hat{f}$ which is the sky-plane projection of the true oblateness, f, as defined in terms of the exoplanetary equatorial and polar radii as

$$f = \frac{R_{\text{eq}} - R_{\text{pol}}}{R_{\text{eq}}}, \tag{5.25}$$

and the projected-obliquity β, which is the sky-plane projection of the angle θ between the minor axis of the oblate spheroid shaped exoplanet and the orbital spin axis. In the absence of tidal effects, the oblateness f may be directly related to the rotational period, P_{rot} of the planet where, assuming a uniform planetary density,

$$P_{\text{rot}} = 2\pi\sqrt{\frac{R_{\text{mean}}^3}{\frac{4}{5}GM_p f\left(1-f\right)^{3/2}}} \tag{5.26}$$

with $R_{\text{mean}} = \sqrt{R_{\text{eq}}R_{\text{pol}}}$ and M_p is the planetary mass (see § 5.2 for details).

Our constraints on $\hat{f}$ and β are found from an analysis of seven high-precision *Spitzer* light curves, identical in observation technique and nearly identical in precision. The seven light curves all have a precision of approximately 550 parts-per-million at $\sim$ 30 s cadence. Figure (5-5) shows the individual observation light curves while Fig. (5-6) gives the light curve after all seven light curves are appropriately merged and the final time-series is resampled to a 30 s cadence.

While we cannot measure the oblateness expected for a theoretical synchronously rotating, tidally uninfluenced HD 189733b (see § 5.2.1 for relevant timescales), we can set interesting limits on parameters describing its shape and, with some additional assumptions, the rotational period. Namely, we exclude Saturn-like projected-oblateness at all obliquities with greater than 95% confidence and for Jupiter-like projected-oblateness at all but obliquities near zero. Table (5.2) collects estimates and uncertainties of the oblateness parameters ($\hat{f}$, β), transit parameters and the derived parameters P_{rot} and J_2. The rotation period is found to be $P_{\text{rot}} = 2.37^{+5.11}_{-1.11}$ days, where we have given the 68% confidence interval in this parameter. By chance, the median value of 2.37 days is close to the theoretical expectation for a synchronously rotating HD 189733b where $P_{\rm rot} = P_{\rm orbital} = 2.22$ days. However, this median estimate is significantly affected by our prior assumption that $\hat{f}$ is uniformly distributed from $0 < f < f_{\rm breakup}$ with $f_{\rm breakup}$ being the oblateness at rotational breakup.

We also constrained the second sperical mass distribution moment $J_2 < 2.1 \times 10^{-2}$, using Eqn. (5.7), by assuming HD 189733b has rotation period $P_{\rm rot} = 2.218573$ day. This upper-bound value of J_2 is larger than J_2 for any Solar System planet (see Table 5.1). While our data constraint on J_2 tells us little about the internal structure of HD 189733b, the technique applied to more precise transit data can yield useful information about how centrally condensed an exoplanet is (as compared to Saturn and Jupiter, for example). In particular, the ratio J_2/f depends only on the moment of inertia of the exoplanet via, for example, the Darwin-Radau relation (Dermott 1984). Of course, in order to independently constrain J_2/f via Eqn. (5.7), we must have some fixed assumption about the value of the rotational period of the planet, $P_{\rm rot}$. For Hot Jupiters, for which the stellar age is several times the synchronization timescale [Eqn. (5.2)], $P_{\rm rot} = P_{\rm orbital}$ is an expected theoretical equivalence.

In merging our data into a single light curve, we have implicitly assumed for our analysis that HD 189733b has both constant projected-oblateness and projected-obliquity over the 1.62 year span encompassing the seven *Spitzer* observations. However, it is the case that both projected quantities should be time-variable even if the true oblateness f and obliquity θ are constant if the exoplanet's spin axis precesses. In particular, the spin axis should precess in response to torques exerted by the star on a non-uniform mass distribution for the exoplanet. If the orbital angular momentum vector is constant, then the spin axis will precess about it with precession period

$$P_{\rm prec} = \frac{2}{3}\frac{P_{\rm orb}^2}{P_{\rm rot}}\frac{\mathbb{C}}{J_2}\frac{1}{\cos\theta} \tag{5.27}$$

where $P_{\rm orb}$ is the orbital period and $2\pi \mathbb{C} M_p R_{\rm eq}^2 / P_{\rm rot}$ is the planet's spin angular momentum (see, for example, Tremaine 1991). Spin precession for planets in the Solar

System occurs on timescales much longer than one year. For example, Saturn completes one precession cycle in $\approx 10^6$ year. However, $P_{\rm prec}$ is a strong function of semi-major orbital distance with $P_{\rm prec} \propto a^3$. For the case of a hypothetical HD 189733b with oblateness similar to that measured for Saturn with a spin-orbit obliquity angle of 45°, the precession period is, in fact, of order one year. In this calculation we have assumed $P_{\rm rot}$ is related to f and J_2 through Eqn. (5.8) and f and J_2 are related to $\mathbb{C}$ through the Darwin-Radau relation (Murray & Dermott 2000)

$$\frac{J_2}{f} \approx -\frac{3}{10} + \frac{5}{2}\mathbb{C} - \frac{15}{8}\mathbb{C}^2. \tag{5.28}$$

We have fixed $\mathbb{C} = 0.225$ which is a reasonable value for gas giant planets (Murray & Dermott 2000). In Fig. (5-9) we have reproduced the content of Fig. (5-7) and have additionally specified approximate spin precession periods in the $\hat{f} - \beta$ plane as solid black contour lines.

Our assumption of constant projected-oblateness and obliquity over the timeframe of the observations appears to be invalid for arbitrary f and θ. At first glance, it would seem that a more appropriate reanalysis assuming unique projected-oblateness parameters for each of the seven measured transit light curves allowing for arbitrary precession periods would result in much weaker constraints than have been presently determined. However, our current constraints, while not fully self-consistent, are not entirely useless. In fact, it may be argued that there is already ample evidence with our data that HD 189733b cannot be precessing at periods shorter than the timescale of our observations for even large values of f. Our evidence is as follows. Should an oblate planet execute spin precession, the area of the sky-projection of the exoplanet should be time variable. The depth, δ, of the transit light curve, in the absence of stellar limb-darkening, is a direct, precise measurement of this area divided by the area of the sky-projection of the star (see, for example, Chapter 2). In particular, it can be shown that the limb-darkening free transit depth measured for an oblate spheroidal exoplanet varies with precession phase, ψ, as

$$\delta(\psi) = \left(\frac{R_{\text{eq}}}{R_\star}\right)^2 \sqrt{(1-f)^{-2}\sin^2\left[\theta\sin\left(\psi-\psi_0\right)\right] + \cos^2\left[\theta\sin\left(\psi-\psi_0\right)\right]} \qquad (5.29)$$

for some initial phase ψ_0 where $\delta(\psi_0) = (R_{\text{eq}}/R_\star)^2$ [we have utilized the formalism by Fillmore (1986) in this derivation].

For strongly oblate planets ($f = 0.1$) the amplitude of the fractional change in the transit depth, $\delta(\psi)/\delta(\psi_0)$, can be several percent. In comparison, we can measure the depth of a single *Spitzer* transit light curve for HD 189733b to better than 0.3% (this work, Torres et al. 2008). The transit depths of each of the seven transit light curves analyzed in this chapter agree with one another at this precision. Therefore, the data exclude transit depth variations whose fractional amplitudes are in excess of this fractional uncertainty in transit depth. In Fig. (5-9) we plot, as dashed contour lines, the peak-to-peak amplitude of $\delta(\psi)/\delta(\psi_0)$ in units of fractional transit depth uncertainty (i.e., the contours mark the number of "sigma" the function $\delta(\psi)/\delta(\psi_0)$ spans). It is clear from this figure that we may constrain a significant portion of the $f - \theta$ plane even when using an analysis including a precession model that allows arbitrary spin precession periods. Future work includes such a reanalysis with this data set. It may even be possible to constrain parameters relating to oblateness, obliquity and precession period by monitoring transit light curve depths over many transit epochs. Such long term monitoring may be possible using transit data collected from the *Kepler* mission, for example.

The recently-found-to-transit exoplanet HD 80606b would be an ideal target for an oblate analysis with a single, high precision, space-based transit light curve for the following reasons. First, the total duration of the transit, at around 12 hr (Winn et al. 2009), would make even a single epoch's transit light curve viable as a high precision measurement, owing to the superb time sampling. Second, the ultra-high eccentricity ($e = 0.93279$) and possibly unique dynamical status of HD 80606b (see § 5.2, Winn et al. 2009 and references therein) will make any determination of oblateness of great scientific interest.

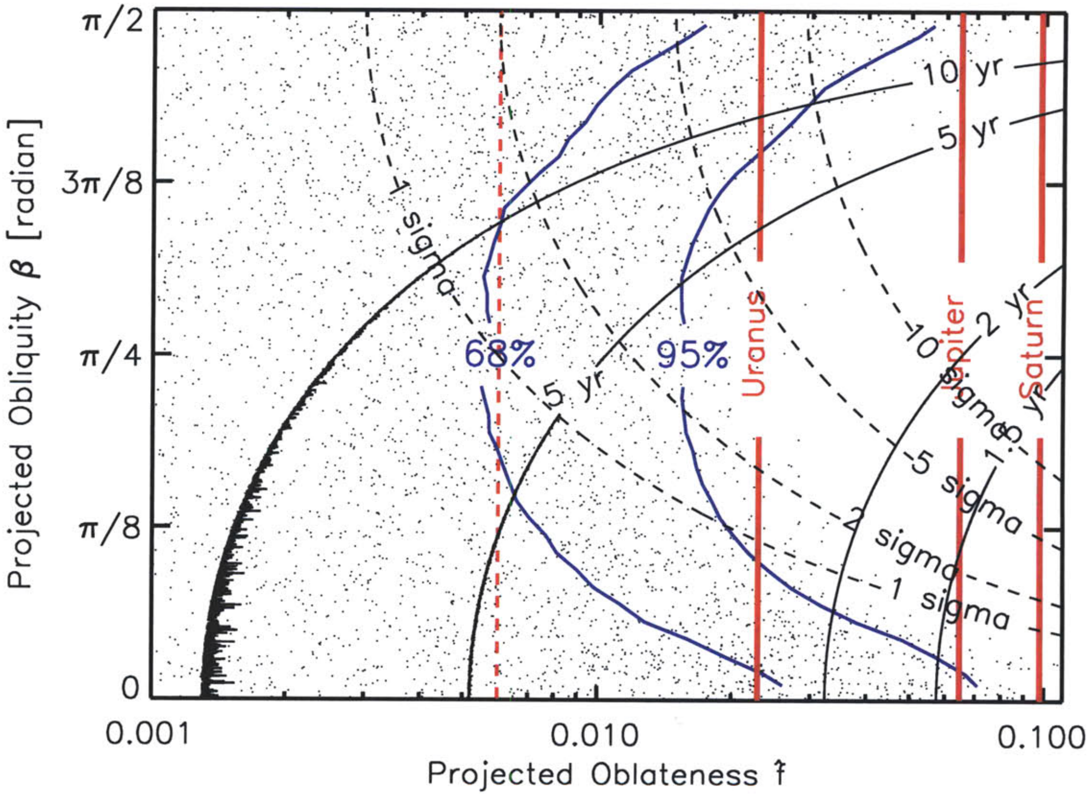

Figure 5-9 Theoretical spin precession periods and transit depth variations for HD 189733b. Refer to the text in § 5.5 for details.

To investigate what what may be learned, we constructed a hypothetical transit light curve for HD 80606b using "standard" transit parameters from Winn et al. (2009) and we additionally assumed the planet has an oblateness $f = 0.097$ (that measured for Saturn) at a 45° obliquity. In Figure (5-10), we plot this transit light curve (with Gaussian white noise added to the light curve as described below) and we also plot the asymmetry of the transit light curve about the time of midtransit. The large asymmetry curve is as a result of the combined effects of oblateness, obliquity and the large eccentricity of the orbit. A single space-based transit observation should uncover this asymmetry with *visual* significance [the asymmetry due to eccentricity alone is at the $\sim 10^{-4}$ level, see Fig. (5-10)].

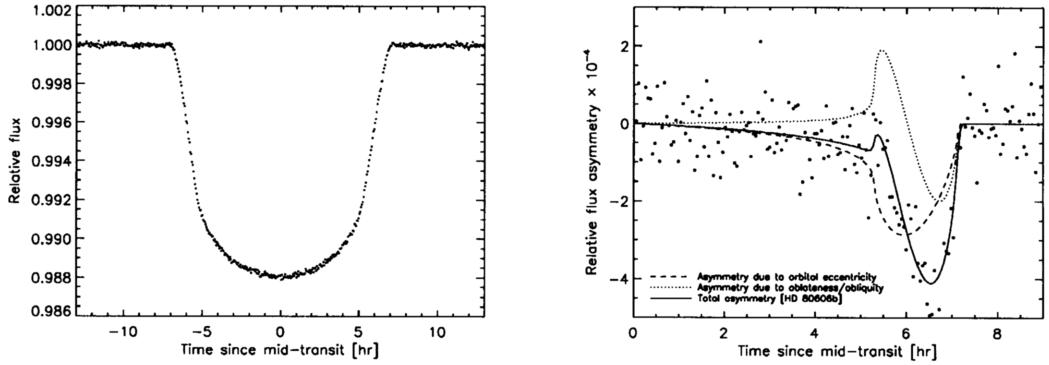

Figure 5-10 Simulated transit light curves for an oblate HD 80606b. The left panel, shows a simulated transit light curve as would be measured with the STIS spectrograph on the *Hubble Space Telescope* for an oblate HD 80606b. The right panel shows the asymmetry (about mid-transit) and simulated noise in a transit light curve for HD 80606b having a projected oblateness equal to that of Saturn and at a 45° obliquity. The asymmetry is as a result of both oblateness/obliquity and the high orbital eccentricity ($e = 0.93$) of HD 189733b.

Going further, we sample this hypothetical transit light curve model for HD 80606b at a 186 s cadence and assume Gaussian, white noise having characteristic width $\sigma = 7.2 \times 10^{-5}$ (in relative flux) is added to each flux measurement. This noise is consistent with that found using data from the STIS spectrograph on the *Hubble Space Telescope* with observing parameters described by Brown et al. (2001) for an observation of HD 209458 (which is 10× brighter than HD 80606; we assume our observation would be photon-limited). The time sampling ensures that ~ 50 points are sampled during the phases of ingress of egress (with duration $\tau \sim 2.6$ hr). We complete an analogous analysis for our hypothetical transit light curve for HD 80606b as was performed on the seven HD 189733b light curves (as was described in § 5.4.3). Figure (5-11) shows the results of this analysis in terms of confidence regions in the $\hat{f} - \beta$ parameter plane. In this case, we would have in fact *measured* the projected-oblateness and projected-obliquity. Such a result would offer priceless constraints on possible modes of formation and evolution of this intriguing planet and system!

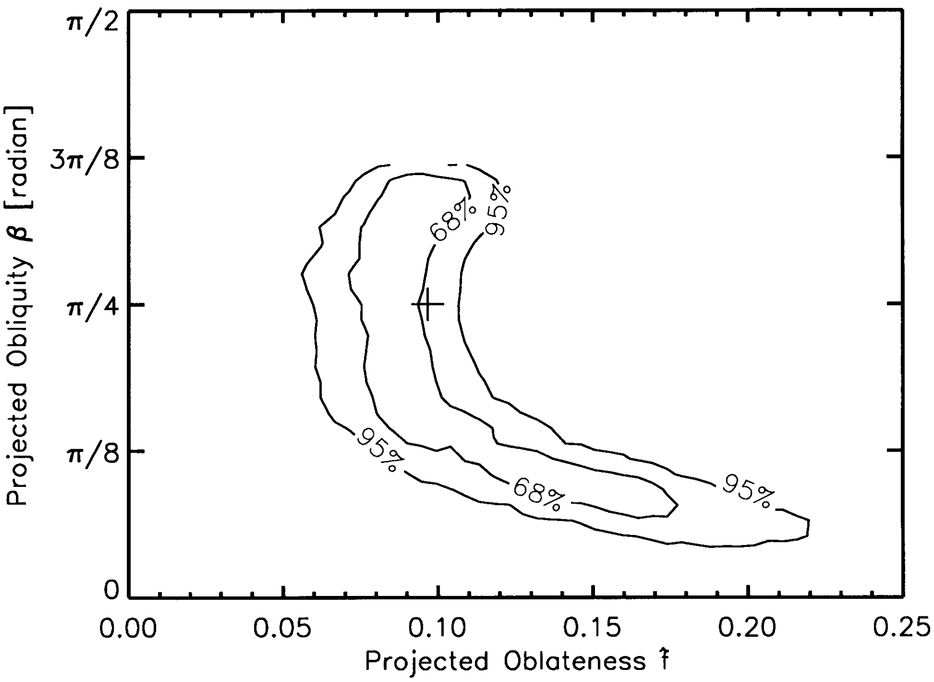

Figure 5-11 Measuring oblateness in a simulated transit light curve for HD 80606b. This figure shows confidence regions in the projected oblateness-obliquity ($\hat{f} - \beta$) plane based upon a Markov Chain Monte Carlo analysis of the simulated transit light curve shown in Fig. (5-10). The cross marks the simulated oblateness and obliquity (equal to that of Saturn and at a 45° obliquity).

As the number of transiting exoplanets grows (with the help of the recently launched *Kepler* mission, for example), we will without doubt encounter many more systems having and exceeding the dynamical complexity of HD 80606b and certainly HD 189733b. A transit light curve model including exoplanet oblateness will likely be *required* for proper (and likely exciting) interpretation. For transiting exoplanets not tidally influenced by their host and possibly residing in their respective habitable zone, the detection of oblateness would be a key measurement to precisely constrain the length of the "exoday."

Thanks to D. Fabrycky for his useful comments concerning exoplanetary spin precession.

Bibliography

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567

Agol, E., Charbonneau, D., Cowan, N., Deming, D., Knutson, H., & Steffen, J. 2007, Spitzer Proposal ID #40238, 40238

Agol, E., Cowan, N. B., Bushong, J., Knutson, H., Charbonneau, D., Deming, D., & Steffen, J. H. 2009, IAU Symposium, 253, 209

Alonso, R., et al. 2004, ApJ, 613, L153

Barbieri, M., et al. 2007, A&A, 476, L13

Barnes, J. W., & Fortney, J. J. 2003, ApJ, 588, 545

Barnes, J. W., & Fortney, J. J. 2004, ApJ, 616, 1193

Borucki, W., et al. 2009, IAU Symposium, 253, 289

Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W., & Burrows, A. 2001, ApJ, 552, 699

Burke, C. J., et al. 2007, ApJ, 671, 2115

Carter, J. A., Yee, J. C., Eastman, J., Gaudi, B. S., & Winn, J. N. 2008, ApJ, 689, 499

Claret, A. 2000, A&A, 363, 1081

Czesla, S., Huber, K. F., Wolter, U., Schröter, S., & Schmitt, J. H. M. M. 2009, arXiv:0906.3604

Dermott, S. F. 1984, Royal Society of London Philosophical Transactions Series A, 313, 123

Eddington, A. S. 1926, The Internal Constitution of the Stars, Cambridge: Cambridge University Press, 1926,

Fazio, G. G., et al. 2004, ApJS, 154, 10

Fillmore, J. A. 1986, AJ, 91, 1096

Gillon, M., et al. 2007, A&A, 471, L51

Goldreich, P., & Soter, S. 1966, Icarus, 5, 375

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with *Mathematica* Support (Cambridge Univ. Press, UK).

Guillot, T., Burrows, A., Hubbard, W. B., Lunine, J. I., & Saumon, D. 1996, ApJ, 459, L35

Hill, Kenneth J., Graphics Gems V, ed A.W. Paeth, Orlando, Academic Press, Inc., 1994, p. 72-77

Høg, E., et al. 2000, A&A, 355, L27

Holman, M. J., & Murray, N. W. 2005, Science, 307, 1288

Holman, M. J., et al. 2006, ApJ, 652, 1715

Hubbard, W. B. 1984, Planetary Interiors, New York, Van Nostrand Reinhold Co., 1984, 343 p.,

Hubbard, W. B., Fortney, J. J., Lunine, J. I., Burrows, A., Sudarsky, D., & Pinto, P. 2001, ApJ, 560, 413

Hui, L., & Seager, S. 2002, ApJ, 572, 540

Knutson, H. A., et al. 2007a, Nature, 447, 183

Knutson, H. A., Charbonneau, D., Noyes, R. W., Brown, T. M., & Gilliland, R. L. 2007b, ApJ, 655, 564

Moutou, C., et al. 2009, A&A, 498, L5

Murray, C. D., & Dermott, S. F. 2000, Solar System Dynamics, by C.D. Murray and S.F. Dermott. Cambridge, UK: Cambridge University Press, 2000.,

Nutzman, P., Charbonneau, D., Winn, J. N., Knutson, H. A., Fortney, J. J., Holman, M. J., & Agol, E. 2009, ApJ, 692, 229

Naef, D., et al. 2001, A&A, 375, L27

Ohta, Y., Taruya, A., & Suto, Y. 2009, ApJ, 690, 1

Peale, S. J. 1999, ARA&A, 37, 533

Pont, F., et al. 2007, A&A, 476, 1347

Press, William H., Teukolsky, Saul A., Vetterling, William T., & Flannery, Brian P. 2007, Numerical Recipes: The Art of Scientific Computing, 3rd ed. (New York: Cambridge University Press)

Rabus, M., et al. 2009, A&A, 494, 391

Seager, S., & Hui, L. 2002, ApJ, 574, 1004

Seager, S., & Sasselov, D. D. 2000, ApJ, 537, 916

Sengupta, S., & Maiti, M. 2006, ApJ, 639, 1147

Silva, A. V. R. 2003, ApJ, 585, L147

Sobol', I. M., & Shukhman, B. V. 1995, International Journal of Modern Physics C, 6, 263

Steinegger, M., Brandt, P. N., Schmidt, W., & Pap, J. 1990, Ap&SS, 170, 127

Torres, G., Winn, J. N., & Holman, M. J. 2008, ApJ, 677, 1324

Tremaine, S. 1991, Icarus, 89, 85

Ward, W. R. 1975, Science, 189, 377

Winn, J. N., Holman, M. J., & Roussanova, A. 2007, ApJ, 657, 1098

Winn, J. N., et al. 2009a, ApJ, 693, 794

Winn, J. N. 2009, IAU Symposium, 253, 99

Winn, J. N., & Holman, M. J. 2005, ApJ, 628, L159

Appendix A

Uniform sampling of an elliptical annular sector

A.1 Elliptical annular sector

A point $(x, y) \in R$ is inside the elliptical annular sector centered at $(0, 0)$ with semi-major axis in the x-direction with axis ratio $\epsilon = A/B$, inner radius a_1, outer radius a_2, and sector angles $\theta_{1,2}$ if the following conditions are all satisfied:

$$(1) \qquad \frac{x^2}{a_1^2} + \frac{y^2}{(\epsilon a_1)^2} > 1 \tag{A.1}$$

$$(2) \qquad \frac{x^2}{a_2^2} + \frac{y^2}{(\epsilon a_2)^2} < 1 \tag{A.2}$$

(3) The line from $(0, 0)$ to (x, y) is at an angle θ such that $\theta_2 > \theta > \theta_1$.(A.3)

as illustrated in Fig. (A-1).

A.2 Uniform sampling

Let (u, v) be a uniform sample of the unit square $[0, 1] \times [0, 1]$. Then (u', v') is uniform sample of the elliptical annular sector centered at $(0, 0)$ with semi-major axis in the

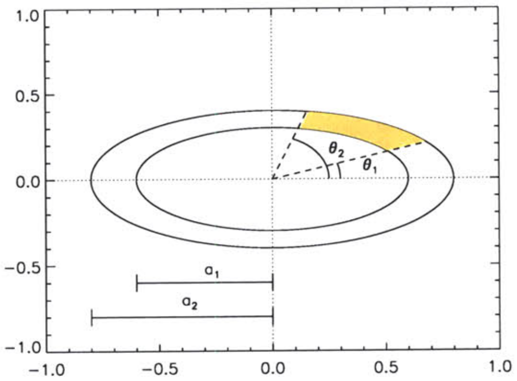

Figure A-1 **An elliptical annular sector.**

x-direction and described by parameters ϵ, $a_{1,2}$, and $\theta_{1,2}$ where

$$u' = a_2 r \cos(\theta) \tag{A.4}$$

$$v' = \frac{a_2 r}{\epsilon} \sin(\theta) \tag{A.5}$$

where

$$r = \sqrt{(1-u)\ a_1^2 + u} \tag{A.6}$$

$$\theta = (1-v)\tan^{-1}(\epsilon \tan\theta_1) + v\tan^{-1}(\epsilon\tan\theta_2). \tag{A.7}$$