On Constraining Nontrivial Properties of Exoplanets and Other Topics in Astrophysics

David Solomon Spiegel

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2008

ABSTRACT

On Constraining Nontrivial Properties of Exoplanets and Other Topics in Astrophysics

David Solomon Spiegel

Part I

At the present epoch, a significant fraction of the baryonic matter in the universe is in a hot, X–ray emitting intragroup medium of small galaxy groups; but at higher redshift the nature of the intergalactic medium of such groups is unconstrained. I examine the X–ray luminosity of 21 galaxy groups in the *CNOC2* Field Galaxy Redshift Survey, at redshifts $0.1 < z < 0.6$, and find that not one was visible in over 100 ksec of observation with XMM-*Newton*, therefore providing tentative evidence that groups at intermediate redshift are underluminous relative to their local cousins.

Part II

As the emerging field of astrobiology pursues the quest to learn if there is life on planets around other nearby stars, there are various properties of planets that it will be important to constrain yet difficult to measure. I address how to constrain three nontrivial properties of extrasolar planets: (*i*) the composition, and (*ii*) the

rotation rate and atmospheric dynamics of hot Jupiters; and (*iii*) the habitability of terrestrial planets.

(*i*) When a giant planet in the source plane of a gravitational microlensing event crosses the fold-caustic of a foreground binary lens, the planet-star brightness ratio is greatly enhanced, which in a favorable situation could allow for a crude optical spectrum of the planet.

(*ii*) As a giant, tidally locked planet transits across the face of its star, the rotation of the planet's atmosphere will impose a Doppler shift on absorption features arising from the planet's atmosphere relative to where the features would be on a non-rotating planet. With future telescopes that have larger collecting areas, we will be able to place constraints on the rotation rate of nearby transiting hot Jupiters.

(*iii*) Finally, I revisit the habitability of terrestrial planets around sun-like stars: Using a 1-dimensional energy balance model, I investigate how the temperature distribution is likely to depend on many factors, including the covering fraction of ocean, the rotation rate, the obliquity, and more.

Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

It is with great pleasure that I take the time to consider, and to thank, the many people who have contributed to my life over the past 7 years and to whom I, and this dissertation, are indebted:

First, family. I thank my parents Stanley and Diana, my sisters Sarah and Stephanie, my girlfriend Giao, and my brother-in-law Aaron. Your enduring love and support, through the difficulties of life and of graduate school, mean everything to me.

To my thesis committee, without whom I could not have even approached completing this thesis –

Frits: you have given me the freedom to find my own winding, branching road through graduate school. Always up for discussing whatever crazy ideas I have had, you have also provided timely and accurate advice, and been as supportive during rough times as I could possibly have hoped for. Thank you.

Caleb, Zoltán, and Kristen: I have enjoyed our collaborations very much; you have each taught me a great deal about how to be a research astronomer. To all of you, and to director of graduate studies Jacqueline van Gorkom: thank you for providing guidance, encouragement, inspiration, and the occasional nudging.

I am endlessly grateful to Ed Spiegel – your wisdom and humor in exchange for my help with your Powerbooks is a trade I'll take any day of the week, and twice on Sundays. Thank you, too, Joe Patterson and David Helfand, without

whom I wouldn't have ended up at Columbia in the first place. Finally, thank you Norm Baker. I am honored to have been a member of the last class you taught.

I thank the graduate students, as colleagues and as friends, with whom have spent my time here. In particular, I appreciate the many scientific conversations and the friendship I have shared with Maurice Leutenegger, Mark Dijkstra, Michel Zamojski, Tony Mroczkowski, Ben Sugerman, Aeree Chung, Suvi Gezari, Cameron Hummels, Neil Zimmerman, Nestor Mirabal, Emily Rauscher, Roban Kramer, Jake Noel-Storr, Jen Donovan, Alan Gersch.

Thank you, too, non-Columbia friends throughout the years: Matt Butler, Zimri Yaseen, Mark Macleod, Jeremy Sosenko, Fred Blumberg, Chris Ingraham, Molly Shearer, Sam Talbot, Joshua Friedman, Jill Schiff, Eve Armstrong, Karen Schrier.

Millie Garcia: thank you so much for your tremendous, nonjudgmental help throughout the years with myriad administrative affairs. You and Ayoune have made it possible for me to focus almost solely on the job of being a graduate student.

Other people who have contributed useful ideas, criticism, and advice to this document include, most significantly, Scott Gaudi, and additionally Michael Allison, Tony Del Genio, Sara Seager, Kevin Roy Briggs, George Rybicki, James Cho, David Graff, and several anonymous referees.

To everyone else not named above, thank you, too.

Chapter 1

Introduction

This thesis comprises observational work searching for the intergalactic medium and theoretical work on extrasolar planets.

1.1   Part I: Seeking Hidden Baryons

1.1.1   Background and Motivation

A fundamental goal of cosmology is to account for the matter in the universe (since Einstein, the matter/energy-density). This endeavor is difficult because nature has ways of hiding matter and energy that are clever enough that cosmologists have remained in the dark for many years. A major complication was discovered by Fritz Zwicky in the 1930s, when he found that the luminous content of galaxy clusters seems to be insufficient to bind galaxies together, given their velocity dispersions (Zwicky 1937, 1938). Many different lines of astronomical inquiry

during the last 70 years have begun to provide a framework for understanding Zwicky's observations: The available evidence now strongly points to the daunting proposition that not only is 73% of the energy-density of the universe dark energy, but of the remaining 27% that consists of matter only a small fraction is the sort that we experience in everyday life – baryonic matter.

The energy budget of the universe, then, is apparently dominated by forms of energy with which we are entirely unfamiliar. Moreover, despite some remarkable success, our current understanding of the evolution and morphology of even the baryons in the universe is strikingly incomplete. Although estimates based on various kinds of observations of the high-redshift universe are converging to form a picture of the universe in which ~ 4.5% of the present-day energy-density consists of baryonic matter (Fukugita & Peebles 2004, hereafter FP04), only ~10% of the predicted baryons are detected in the local universe in the form of stars and stellar remnants, cool neutral gas, or hot gas in galaxy clusters, leaving nearly 90% unaccounted (FP04; see Figure 1.1 below). This disparity leads to a profound conclusion that is also suggested by cosmological simulations (Davé et al. 2001): At the present epoch, a large fraction of baryons are in very low-density environments – the warm-hot intergalactic medium (WHIM) and the intragroup medium of small groups of galaxies.

Mulchaey (2000) describes *ROSAT* X-ray observations that indicate that such simulations are correct in predicting that part of the missing baryonic matter is in the intragroup medium of small galaxy groups: In the local universe, a majority of

Present Epoch Baryon Budget
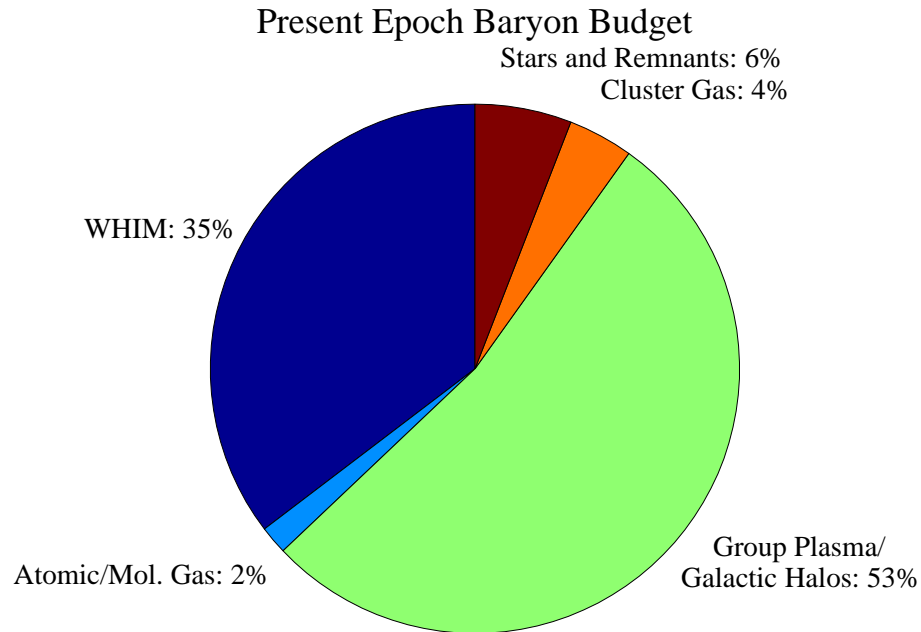
Figure 1.1 A summary of the detailed accounting of the baryon budget of the local universe presented in Fukugita & Peebles (2004) The two largest pie-wedges (Warm-Hot Intergalactic Medim – or WHIM – and Group Plasma/Galactic Halos) are almost entirely undetected and presumably contain nearly 90% of the baryons at the present epoch. The sum of these wedges is known to greater precision than either one is individually. This is because their sum is the difference between unity and the sum of all observed sources of baryonic matter, whereas the breakdown into 35% and 53% is motivated in part by simulations the simulations of Davé et al. (2001).

small galaxy groups exhibit diffuse X-ray emission that reveals significant quantities of baryonic matter that are not visible at optical wavelengths. How the universe evolved to this state, and when the baryons began collecting in the potential wells of small groups, however, is observationally undetermined.

1.1.2 The Baryonic Content of Groups at Moderate Redshift?

Part I describes an attempt to learn about the state of the intergalactic medium of small groups when the universe was only $\sim 2/3$ its current age. It consists of a single chapter (Chapter 2[1]) in which I present results and analysis of a sensitive X-ray observation of galaxy groups at intermediate redshift. The *CNOC2* Field Galaxy Redshift Survey (Carlberg et al. 2001a,b), a deep spectroscopic survey of several fields on the sky, identified a number of group-candidates – collections of three or more galaxies that meet a specified set of requirements that are intended to minimize the number of chance-alignments. The statistical weak-lensing signal from these putative groups (Hoekstra et al. 2003) provides compelling complementary evidence that they are indeed gravitationally bound objects, not physically unconnected galaxies that just happen to lie near one another in redshift space. Of the 21 of these *CNOC2* groups in the field of view of the *EPIC-PN* camera on XMM-*Newton*, not one was visible in over 100 ksec of observation, even though a simple estimate implies that three of the groups have velocity dispersions high enough that they would easily be visible if their luminosities scaled with their velocity

[1]Chapter 2 is nearly identical to the text of Spiegel et al. (2007c).

dispersions in the same way as nearby groups' luminosities scale.

Note, however, that the statistical power of the survey is insufficient to draw definitive conclusions. Although Fang et al. (2007) similarly find evidence that groups at intermediate redshift are less X-ray luminous than dynamically comparable ones in the local universe, Andreon et al. (2008) and Fassnacht et al. (2007) make some arguments suggesting otherwise. Therefore, the question of how the cosmic baryon inventory has evolved from redshift ~ 0.5 to the present remains open.

1.2 Part II: Extrasolar Planets

1.2.1 Background and Motivation

On February 17, 1600, 9 years before Galileo built his first telescope, Giordano Bruno was burned at the stake for having proposed that there are many worlds, including ones more distant than our Sun. Although observing technology remained insufficient to detect worlds around other stars until the last half of the 20th century, astronomers never lost the desire to engage in observations and speculation regarding very distant worlds. Starting in the middle of the 19th century, when Jacob (1855) ascribed what he said was an anomalous motion in the binary system 70 Ophiuchi to the presence of an unseen planet, and continuing for more than 100 years, more than a dozen observers claimed to have detected astrometric wobbling of nearby stars that was caused by planetary companions. However,

such early announcements of planets found beyond our Solar System are now generally considered to have been mistaken (Batten et al. 1984; Heintz 1988). Although the claims based on astrometry were erroneous, Otto Struve prophetically proposed that high precision radial velocity (RV) analysis of nearby stars might reveal massive planets in close orbits (Struve 1952). When Campbell et al. (1988) found spectroscopic evidence of what is now considered to be the first extrasolar planet ever discovered, Bruno's fateful belief was finally confirmed.

In the last 20 years, progress in extrasolar planet detection has accelerated rapidly. Wolszczan & Frail (1992) were the first to obtain essentially indisputed evidence of a planet around another star (although the star was a pulsar). Several years later, Mayor & Queloz (1995) and Marcy & Butler (1996) found the first widely accepted evidence of planets around solar–type stars.[2]

Huang (1959) had considered and rejected the possibility of performing systematic photometric searches for extrasolar planets transiting along the lines-of-sight from Earth to their parent stars. Even though contemporary observers could achieve the ~1% photometric accuracy that would be needed to detect a transiting Jupiter around a nearby Sun-like star, the probability that any particular planetary system would have the appropriate geometrical configuration was considered to be very low. For randomly oriented orbital planes, the probability that a planet – at some point during its orbit – will transit across the face of its star from Earth's vantage is approximately R_*/a, where R_* is the star's radius and a is the orbital semi-major axis. If other solar systems were built like our own, with Jupiter-like

---

[2]The planet of Campbell et al. (1988) was not confirmed until further analysis (Hatzes et al. 2003).

planets at Jupiter-like distances from the star, the probability of seeing a transit for a given solar system would be ~0.1%, and each transit would be seen for only about a day per decade.

On the other hand, since many of the first-discovered planets were of comparable mass to Jupiter and were in very close orbits ($\lesssim 0.1$ AU), it was clear that many extrasolar planetary systems have dramatically different structures from ours. By the late 1990s, therefore, it seemed sensible to perform accurate photometric observations of stars known to have planetary companions – and of other stars as well. Within just a few years, Henry et al. (2000) found that one of the planets that had been discovered through RV analysis – HD209458b – also transits.

The discovery of the first transiting planet sparked tremendous excitement, and led to several dozen surveys dedicated to searching for transits. In a conference proceeding titled, "Status and Prospects of Planetary Transit Searches: Hot Jupiters Galore",[3] Keith Horne predicted that 23 transit surveys operating in the year 2002 would cumulatively detect 191 planets each month (Horne 2003; though this was generally thought to be overly optimistic even at the time). Figure 1.2 summarizes his survey-by-survey estimates. A naive extrapolation (that, in fairness, Horne cautioned against) indicates that, if he had been right, there would now be over ten-thousand transiting exoplanets known. Instead, there are fewer than 300 exoplanets known, fewer than 50 of which transit. Detecting transits from ground-based telescopes, it turns out, is much more difficult than was initially appreciated.

With a pair of space-based transit-detecting observatories set to monitor ~200,000

---

[3]Close-orbiting gas-giant planets are often called "hot Jupiters".

2002 Transit Searches

Figure 1.2 This figure is reproduced from Figure 2 of Horne (2003). That paper predicted that 23 transit surveys operating in the year 2002, using telescopes with diameters ranging from 3.6 cm to 4 m, would cumulatively detect 191 planets each month. Each open circle in this figure indicates the diameter and predicted number of planets per month for one survey telescope.

stars over the next five-to-seven years, astronomers are again expecting that large numbers of new planets will soon be discovered. In its one year of operation, COROT has already found one of the most exotic planets ever discovered (COROT-Exo-1b), orbiting with a period of 1.5 days, with a radius reportedly as great as 1.78 times that of Jupiter.[4] The *Kepler* mission is scheduled to launch in February of 2009, and has been predicted to find anywhere from several hundred to over 10,000 planets, many of which might be smaller, terrestrial ones (Borucki et al. 2007, 2003; Basri et al. 2005).

For most of the planets that have been discovered, all that is known are their orbital elements; since the orientation of the orbital plane is typically not known, the mass cannot be ascertained beyond a $\sin[i]$ degeneracy. For the transiting planets, this degeneracy is broken, and both their mass and radius can be easily determined, which places some constraints on their interior structures. But we would like to know much more about them. We would like to learn about the composition and dynamics of their atmospheres, the detailed structure of their deep interiors, their chemical and dynamical evolution. Ultimately, we would like to know if there are any living organisms on them.

Many planetary properties that we will wish to constrain will be difficult to measure. In Part II, I address how to constrain three nontrivial properties of extrasolar planets: (*i*) the composition of hot Jupiters (§ 1.2.2; Chapter 3); (*ii*) the rotation rate and atmospheric dynamics of hot Jupiters (§ 1.2.3; Chapter 4); and (*iii*) the habitability of terrestrial planets (§ 1.2.4 and § 1.2.5; Chapters 5 and 6).

[4]COROT–Exo–1b is reported at http://www.esa.int/esaCP/SEMCKNU681F_index_0.html.

1.2.2 The Composition of a Microlensed Planet?

In Chapter 3,[5] I revisit the possibility of detecting an extrasolar planet around a background star as it crosses the fold caustic of a foreground binary lens. During such an event, the planet's flux can be magnified by a factor of ~ 100 or more. The detectability of the planet depends strongly on the orientation of its orbit relative to the caustic. If the source star is inside the inter–caustic region, detecting the caustic–crossing planet is difficult against the magnified flux of its parent star. In the more favorable configuration, when the star is outside the inter–caustic region when the planet crosses the caustic, a close–in Jupiter–like planet around a Sun–like star at a distance of 8 kpc is detectable in 8-minute integrations with a 10 m telescope at maximal S/N ~ 15 for phase angle ϕ ~ 10° (assuming an edge-on orbit). Most exoplanets currently known are less than several hundred parsecs from Earth, and the community has considered probing the atmospheric composition of only these relatively nearby exoplanets (e.g., the work of Brown (2001), Charbonneau et al. (2002), Grillmair et al. (2007), etc.). A micro*lensed* planet, in contrast, is likely to be several kiloparsecs away or more. In this chapter, I consider whether, with favorable viewing conditions, it might be possible to probe the composition of such very distant planets. In the example described above, I find that the presence of methane, at its measured abundance in Jupiter, and/or water, sodium and potassium, at the abundances expected in theoretical atmosphere models of close–in Jupiters (Sudarsky et al. 2000, 2003), could be inferred from a

---

[5]Chapter 3 is nearly identical to the text of Spiegel et al. (2005).

non–detection of the planet in strong broad absorption bands at $0.6 - 1.4\mu$m caused by these compounds, accompanied by a $S/N \sim 10$ detection in adjacent bands. I conclude that future generations of large telescopes might be able to probe the composition of the atmospheres of very distant extrasolar planets.

1.2.3 A Transiting Planet's Rotation Rate?

In Chapter 4,[6] I investigate the effect of planetary rotation on the transit spectrum of an extrasolar giant planet, in order to address in detail the observability of some spectral signals originally discussed by Brown (2001). During ingress and egress, absorption features arising from the planet's atmosphere are Doppler shifted by of order the planet's rotational velocity ($\sim 1 - 3$ km s^{-1}) relative to where they would be if the planet were not rotating. I focus in particular on the case of HD209458b, which ought to be at least as good a target as any other known transiting planet. For HD209458b, this shift should give rise to a small net centroid shift of ~ 60 cm s^{-1} on the stellar absorption lines. Using a detailed model of the transmission spectrum due to a rotating star transited by a rotating planet with an isothermal atmosphere, I simulate the effect of the planet's rotation on the shape of the spectral lines, and in particular on the magnitude of their width and centroid shift. I then use this simulation to determine the expected signal–to–noise ratio for distinguishing a rotating from a non–rotating planet, and assess how this S/N scales with various parameters of HD209458b. I find that with a 6 m telescope, an equatorial rotational

---

[6]Chapter 4 is nearly identical to the text of Spiegel et al. (2007a).

velocity of ~ 2 km s^{-1} could be detected with a S/N ~ 5 by accumulating the signal over many transits over the course of several years. With a 30 m telescope, the time required to make such a detection reduces to less than 2 months.

1.2.4 Climatic Habitability

In Chapter 5,[7] I begin to lay the groundwork for interpreting the climatic habitability of the many terrestrial exoplanets that we hope will be found in the next several years and beyond. According to the standard liquid-water definition, the Earth itself is only partially habitable. I reconsider planetary habitability in the framework of energy-balance models (EBMs) – the simplest seasonal models in physical climatology – to assess the spatial and temporal habitability of Earth-like planets. In order to quantify the degree of climatic habitability of these models, I define several metrics of fractional habitability. Previous evaluations of habitable zones may have omitted important climatic conditions by focusing on close Solar System analogies. For example, I find that model pseudo–Earths with different rotation rates or different land–ocean fractions generally have fractional habitabilities that differ significantly from that of the Earth itself. Furthermore, the stability of a planet's climate against albedo-feedback snowball events strongly impacts its habitability. Therefore, issues of climate dynamics may be central in assessing the habitability of discovered terrestrial exoplanets, especially if astronomical forcing conditions generally differ from the moderate Solar System cases.

[7]Chapter 5 is nearly identical to the text of Spiegel et al. (2007b).

1.2.5   Habitability and Planetary Obliquity

In Chapter 6, I expand upon the investigations begun in Chapter 5, adding several new ingredients to the EBM. Without the stabilizing influence of the Moon, the Earth's obliquity might vary significantly. Extrasolar terrestrial planets with the potential to host life may therefore also be subject to large obliquity variations. With this in mind, I revisit the habitability of oblique planets. I further enrich the EBM by allowing variations of land/ocean distribution, and I incorporate an infrared cooling function that explicitly accounts for the large increase in CO_2 concentration expected at orbital distances greater than 1 AU from a 1 $L_\odot$ star. The several new tunable knobs in the EBM together produce a vast new parameter space, one that must be explored in order to understand what climatic trends are likely on less-Earth-like planets. I again pay particular attention to dynamical transitions to ice-covered snowball states that result from ice-albedo feedback. Of course, if one is to trust that the climate model might produce reliable results for conditions that differ from those on Earth, one would hope that it at least predicts Earth-like climates for Earth-like conditions.[8] And indeed, despite the great simplicity of the EBM, it captures rather well the seasonal cycle of global energetic fluxes at Earth's surface. It also performs satisfactorily against a full-physics climate model of a highly oblique Earth, in an unusual regime of circulation characterized by heat transport from the poles to the equator. Having verified that the model produces reasonable results for the conditions for which I can cross-check it, I then use

---

[8]This caveat applies equally to Chapter 5.

it to find several important results. First, I find that oblique terrestrial planets can violate global radiative balance throughout much of their seasonal cycles if they have asymmetrical distributions of continents and ocean. This is likely to limit the usefulness of simple radiative equilibrium arguments. Second, while the climates of high obliquity planets can be severe, with large seasonal variations, I find (as did Williams & Kasting (1997)) that they are not necessarily more prone to snowball transitions than low obliquity planets. In fact, the models indicate that high obliquity planets may be *less* prone to global snowball events, which raises the interesting possibility that moon-less worlds might in general have a higher probability of maintaining a temperate climate. Finally, I find that, depending upon the efficiency of latitudinal heat transport, terrestrial planets with massive CO_2 atmospheres, typically found in the outer regions of habitable zones, also can be subject to dynamical snowball transitions.

1.2.6 Conclusions: Looking Forward

In Chapter 7, I extrapolate beyond the studies described in the first 6 chapters, in anticipation of several challenging investigations that may prove to be important and interesting in coming years. Specifically:

- Various authors have been suggested that a planet may be much more likely to be hospitable to life if it possesses a magnetic field; but can an extrasolar planet's magnetic field be measured? I consider the possibility that, with a space-based array of radio antennas, it will be possible to detect cyclotron

maser radiation from a nearby terrestrial planet.

- Through comparing observed infrared light curves and spectra of hot Jupiters to detailed models of internal structure and composition, various investigators are starting to learn about the formation, evolutionary history, and current structure of these objects. In the coming decades, similar work will improve our understanding of terrestrial planets. I discuss how an enhancement to the sort of simple energy balance models considered here may not only allow us to predict which terrestrial planets may be habitable, but also may be of use in designing, and interpreting the results of, future missions to obtain spectra of terrestrial planets, such as *Terrestrial Planet Finder* missions.

Part I

Seeking Hidden Baryons

Chapter 2

A Possible Dearth of Hot Gas in Galaxy Groups at Intermediate Redshift

2.1 Introduction

According to the paradigm of hierarchical structure formation, as matter falls from low–density environments to high–density environments (i.e., clusters of galaxies) it passes through stages of intermediate density, namely, groups. Rich clusters of galaxies, containing as many as a thousand member–galaxies or more, are visually quite prominent and have therefore attracted research interest for over 70 years. Over the last several decades, however, it has become increasingly apparent that small groups of galaxies, containing fewer than 50 members and often containing

as few as 3–10, constitute by far the most common environment in which galaxies are found in the universe today, and are therefore the dominant stage of structure–evolution at the present epoch. In order to understand the structure and evolution of matter in the universe, then, we must understand the properties of groups of galaxies.

In rich clusters, a large fraction of the baryonic mass ($\gtrsim 80\%$) exists as diffuse, hot, intracluster gas – see Ettori et al. (2003) and the review by Rosati et al. (2002) and references therein – that has been detected with space–based X–ray telescopes for over 30 years (Giacconi et al. 1974; Rowan-Robinson & Fabian 1975; Schwartz 1978). If small galaxy groups have a similar hot–gas component, then a large fraction of the baryonic mass of the universe could be hiding in these groups (Fukugita & Peebles 2004, and references therein). Alternatively, if groups do not contain a diffuse, extended intragroup medium (IGM), this would be troubling news for the idea that structure forms in the hierarchical fashion that is widely assumed.

In the last 15 years, the improved resolution and sensitivity of new X–ray telescopes have allowed us to begin to study the X–ray properties of groups of galaxies. *ROSAT* observations indicate that in the local universe, half to three quarters of groups have a hot, X–ray emitting IGM (Mulchaey 2000). Plionis & Tovmassian (2004, hereafter PT04) examine the relationship between the X–ray luminosity (L_X) and the velocity dispersion (σ_v) of nearby groups in the Mulchaey et al. (2003, hereafter M03) catalog (although there remain significant uncertainties associated with the analysis). At higher redshift, however, a systematic study of this

kind has not yet been performed, although some recent evidence suggests that the relationship at $z \gtrsim 0.5$ is different from the local one: a deep *Chandra* observation failed to reveal X–ray emission from groups discovered by the *DEEP2* Galaxy Redshift Survey, which Fang et al. (2007) find strongly suggests that the groups in their survey (at $z \sim 1$) are less X–ray luminous than the nearby relationship would predict.

A recent deep imaging and spectroscopic study by the Canadian Network for Observational Cosmology Field Galaxy Redshift Survey (*CNOC2*) made it possible to identify a large number of galaxy groups at intermediate redshift ($0.1 < z < 0.6$) (Carlberg et al. 2001a,b). Gravitational lensing analysis of the *CNOC2* fields yields statistical lensing masses of these groups (Hoekstra et al. 2001). This data set, therefore, provides an ideal laboratory for studying the X–ray properties of groups at these redshifts.

Fortunately, the data archive for *XMM–Newton* contains approximately 110 ksec worth of observations mostly overlapping one of the *CNOC2* fields (21 of the groups are in the field of view of the *EPIC–PN* camera; see Table 2.1). If the relation between L_X and σ_v of groups at these redshifts is the same as the relation that holds for nearby groups, then we will argue that a few of the most massive groups ought to have been visible in the aggregate data from these observations. In this analysis, we looked for whether the X–ray photons received are at all correlated to the groups: (*i*) Are there more (or fewer) photons where there are groups than where there are not? (*ii*) Is the spectral energy distribution of the photons where

there are groups different from that of the background?

In § 2.2, we describe the data we used and how we reduced it. In § 2.3 and § 2.4, we describe our analysis of the data – spatial analysis and spectral analysis, respectively. We found no spatial or spectral evidence for the groups; and in § 2.5, we discuss what we had expected the data to look like, and how surprising it is that we failed to detect the groups. In § 2.6, we attempt to draw conclusions, and we speculate as to the reasons why we did not detect the groups. Finally, in Appendix I, we give a brief analysis of several different methods of fitting lines to data; and in Appendix II, we describe how we estimate the probability that groups at intermediate redshift share the same X–ray luminosity function as those in the local universe.

2.2   Observations and Data Reduction

The *CNOC2* survey comprised four fields on the sky. We examined the X–ray properties of the optically identified galaxy groups in the 1447+09 field (Carlberg et al. 2001a,b). For our analysis, we used the *EPIC–PN* data from three publicly available observations of that field in the *XMM–Newton* data archive, of duration 33 ksec, 33 ksec, and 43 ksec. We did not use the corresponding *EPIC–MOS* data in the current analysis because of the lower sensitivity of the *MOS* to low energy photons, and the attendant modest increase in sensitivity in background–limited images: including *MOS* would have increased the number of counts at the locations of the groups in our analysis by a factor of only $\sim 50\%$. The relevant summary data

on the 21 groups in the field of view of *EPIC–PN*, from Carlberg et al. (2001a), is presented in Table 2.1.

For basic data reduction, including generating good–time–intervals (GTI) files to deal with periods of high solar activity, removing known hot–pixels, and generating images and exposure maps for energy bands of interest, we used *XMM–Newton* Science Analysis Software (SAS) release 6.1.0. The light–curves used to generate GTI files consisted of integrated flux in the range 0.3–15.0 keV across the whole detector. Time intervals of particularly high background (`RATE>9`) were flagged as periods of flaring and were removed from the data. For the purpose of the analysis described below, we considered four energy bands: 0.3–0.8 keV, 0.8–1.5 keV, 1.5–4.5 keV, and 0.2–1.0 keV, which we will hereafter refer to as bands A, B, C, and D, respectively. To make images in each band, we selected events in the GTI with `PATTERN<=4` and `FLAG==0`. Images and corresponding exposure maps were made at 2″ per pixel, roughly twice the resolution of the intrinsic pixel size of *EPIC–PN* (4.1″).

Since we were looking for faint emission (from extended sources) that was expected to outshine the background by only a slim margin, it was essential to remove point sources from the images. In addition, it was crucial to characterize the background carefully before further analysis so as to maximize the accuracy of the measurements of both the background flux level and the group luminosities. Two SAS tasks – `eboxdetect` and `emldetect` – were used to find and remove point sources in all observations, in all energy bands; locations of point sources were then

masked in the images and exposure maps. In order to characterize the background, we used two methods that are described in detail in the next section. In short, one was to smooth the point–source excluded image, and the other was to sample the average count–rate over the groups–excluded image.

2.3   Spatial Analysis

We investigated whether the spatial locations of received photons were at all correlated with the groups. This inquiry required a metric of surface brightness that would allow us to compare quantitatively different patches of sky. The idea behind the desired metric is to find the average number of counts per unit sky area per unit time within the patch. A reasonable first guess of the metric would therefore be the sum of the pixel values in the patch, divided by the number of pixels, divided by the total exposure time. In an experiment in which the pixels have nonuniform effective exposure time, however, they have unequal sensitivity. Maximizing the sensitivity of such an experiment requires weighting more heavily those pixels that are more sensitive in the observation. A reasonable second guess of the metric of the surface brightness of a patch of N pixels, then, is

$$\mathcal{R} \equiv k \times \frac{\sum_{i=1}^{N}(p_i/e_i)w_i}{\sum_{i=1}^{N} w_i},$$

where k is the number of square arcseconds per pixel (in our images, $k = 4$ arcsec^2 pixel^{-1}), p_i is the number of counts in the ith pixel of the patch, e_i is the effective exposure

Table 2.1   The 21 *CNOC2* Groups in the Field of View.

Group ID	R.A. (J2000)	Decl. (J2000)	z^a	$N_z{}^b$	σ_v (km s^{-1})^c	Est. M ($h^{-1}M_\odot$)^d
1	14 49 42.453	09 02 44.76	0.165	3	164 ± 126	1.3×10^{13}
2	14 48 55.799	09 08 48.48	0.229	3	162 ± 139	7.2×10^{12}
3	14 48 57.854	08 57 05.58	0.262	4	229 ± 77	1.2×10^{13}
4	14 49 28.406	08 51 54.07	0.270	3	104 ± 92	3.0×10^{12}
5	14 49 13.637	08 50 10.34	0.271	4	112 ± 66	3.0×10^{12}
6	14 49 44.220	08 57 37.84	0.273	3	165 ± 120	8.1×10^{12}
7	14 48 49.049	08 57 50.15	0.306	3	93 ± 65	1.3×10^{12}
8	14 50 20.387	09 06 00.62	0.306	4	199 ± 161	1.6×10^{13}
9	14 50 13.703	08 57 37.28	0.325	4	175 ± 151	9.9×10^{12}
10	14 49 03.618	09 07 02.86	0.359	4	82 ± 66	1.6×10^{12}
11	14 49 40.860	09 02 15.68	0.372	3	126 ± 94	2.8×10^{12}
12	14 50 22.850	09 01 14.74	0.373	4	44 ± 41	2.1×10^{11}
13	14 49 49.201	08 51 23.08	0.374	4	291 ± 193	3.4×10^{13}
14	14 50 00.825	08 49 06.65	0.394	4	308 ± 257	3.6×10^{13}
15	14 49 14.155	09 11 15.38	0.394	3	394 ± 406	4.8×10^{13}
16	14 49 55.433	08 56 11.39	0.394	3	507 ± 469	7.8×10^{13}
17	14 49 32.746	09 03 41.28	0.468	4	488 ± 417	7.5×10^{13}
18	14 49 29.974	09 09 08.15	0.469	4	217 ± 224	8.9×10^{12}
19	14 49 31.428	09 05 05.22	0.472	3	123 ± 99	4.9×10^{12}
20	14 49 23.516	08 58 47.83	0.511	3	565 ± 668	1.2×10^{14}
21	14 49 23.763	08 55 17.84	0.543	3	151 ± 132	8.7×10^{12}

^{a}ID Numbers are assigned to the groups that were used for the analysis in this paper. Groups are listed in order of increasing redshift (z).

^{b}N_z is the number of galaxies with identified redshifts in each group.

^{c}σ_v is line-of-sight velocity dispersion of group galaxies.

^{d}Estimated mass, and all other numbers in this table (except ID Number), are taken from Carlberg et al. (2001a).

time seen by that pixel, and w_i is an arbitrary weighting factor. In order to maximize the signal–to–noise ratio of $\mathcal{R}$, the appropriate weighting factor is $w_i = e_i$, so in the end the metric reduces to

$$\mathcal{R} \equiv k \times \frac{\sum_{i=1}^{N} p_i}{\sum_{i=1}^{N} e_i}. \tag{2.1}$$

In order to determine the average surface brightness of the groups, we had to know which patches of sky to define as the groups. For each group, we defined a circular aperture within which to search for X–ray emission. The best way to assign radii to groups is open to debate, so we tried various ways, including two methods that depend on the groups' properties – r_{500} and r_{200}, calculated with formulae from the literature – and several that do not. The dependence of r_{500} (Osmond & Ponman 2004) and r_{200} (Mahdavi & Geller 2004) on redshift z and line–of–sight velocity dispersion σ_v are given by

$$r_{500} = 130 \text{ kpc} \left(\frac{\sigma_v}{100 \text{ km s}^{-1}} \right) \left(\frac{H(z)}{73 \text{ km s}^{-1} \text{ Mpc}^{-1}} \right)^{-1} \tag{2.2}$$

$$r_{200} = 230 \text{ kpc} \left(\frac{\sigma_v}{100 \text{ km s}^{-1}} \right) \left(\frac{H(z)}{73 \text{ km s}^{-1} \text{ Mpc}^{-1}} \right)^{-1}, \tag{2.3}$$

where $H(z)$ is the value of the Hubble constant at redshift z. All calculations in this paper assume $H_0 = 73 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$. We also tried assigning a fixed physical size to each group (125 kpc, 250 kpc, and 500 kpc). In each case, physical size was converted to angular size using $\alpha = r/D_A(z)$, where r is the group's physical size and $D_A(z)$ is its angular diameter distance. Happily, our main scientific results were independent of how we calculated the size of the

aperture around each group – and this is an important point, on which we will elaborate below. In this paper, we will present results for 250 kpc apertures around each group.

The $\mathcal{R}$–value of a group (according to equation (2.1), within the set of pixels defined by the group's angular radius α) cannot by itself be converted to the group's luminosity, because it includes counts from the background that are unrelated to the groups. The background flux level must therefore be calculated and subtracted. The true measure of the average surface brightness of each group is the difference between the group's $\mathcal{R}$–value and the $\mathcal{R}$–value of the background:

$$SB_g \equiv \mathcal{R}_g - \mathcal{R}_{BG}. \tag{2.4}$$

We calculated the average surface brightness of the background in two different ways, and we also calculated the local background corresponding to each group, which gave us multiple measures of the surface brightness SB_g of each group.

We calculated the average background for the whole image in two ways. First, for each energy band, we determined the $\mathcal{R}$–value of the image for all pixels not contained in any group's aperture. Second, we created a background image for each energy band by smoothing that band's image with a Gaussian smoothing radius of 60 pixels and, as before, determined the $\mathcal{R}$–value of the background image for all pixels not contained in any group's aperture. These two methods of determining the average background of the whole image yielded nearly identical results. According to Scharf (2002), at redshift of ~ 0.5, low mass objects such as

groups should be seen most readily in a low–energy band such as D or A. Since D–band contains more counts than A–band, much of our analysis deals with D–band data. The different measurements of the whole–detector background were similar; in D–band, for example, the background rate was measured to be 3.26 or 3.08×10^{-6} cts arcsec^{-2} s^{-1}, depending on whether the image or the smoothed (background) image was used.

We also measured the local background corresponding to each group by taking an annulus of inner radius 1.05α and outer radius 1.20α around each group, and finding the $\mathcal{R}$–value of the background image in the annulus surrounding the group. These local background measurements ranged from 2.44 to 4.21×10^{-6} cts arcsec^{-2} s^{-1}, and had a median of 2.77×10^{-6} cts arcsec^{-2} s^{-1} and a mean of 2.96×10^{-6} cts arcsec^{-2} s^{-1}.

Finally, to understand the spatial variability of the background, in each energy band we placed an aperture of radius 0.9′ (27 pixels) in 20,000 random locations on the point–source–excluded image, and measured the $\mathcal{R}$–value within each aperture. This procedure gives an empirical measure of the probability distribution function that describes the surface–brightness in apertures of approximately the sizes of the groups. Results for D–band are displayed in the histogram in Figure 2.1. For D–band, the modal $\mathcal{R}$–value of the samples of the background is 2.8×10^{-6} cts arcsec^{-2} s^{-1}; the median $\mathcal{R}$–value, 3.0×10^{-6} cts arcsec^{-2} s^{-1}, is similar; and the mean, 3.5×10^{-6} cts arcsec^{-2} s^{-1}, is somewhat higher owing to a few anomalously high measurements that resulted from apertures with small numbers

of pixels near the edge of the detector or near masked regions.

Surprisingly, the $\mathcal{R}$–values of the groups were frequently less than those of the background. This means both that $\mathcal{R}$–values of individual groups were often less than the corresponding $\mathcal{R}$–values of the local background, and that the stacked $\mathcal{R}$–value for all 21 groups, 2.78×10^{-6} cts arcsec^{-2} s^{-1}, was 10% *less* than the lesser of the two measures of the average background. In fact, the average value of the surface brightness of groups (calculated with equation (2.4) and using the local background $\mathcal{R}$–value for $\mathcal{R}_{BG}$; average was weighted by the product of the number of pixels and the effective exposure time) turned out to be negative: -8.6×10^{-8} cts arcsec^{-2} s^{-1}.

Ultimately, because of the spatial variability of the background, which is evident in Figure 2.1, the most useful description of the background is probably the local one. The top panel of Figure 2.2 shows the number of counts by which each of the 21 groups in our sample is brighter than its estimated local background, plotted against redshift. The bottom panel of that figure shows the surface brightness above the local background SB_g, again plotted against redshift. The 1–σ error bars are computed on the ordinates as the summation in quadrature of the $N^{1/2}$ Poisson noise from the group (σ_g) and from the background (σ_{BG}). No group is seen more than 5σ above the background, which is our minimum criterion for detection. The dashed line in each panel represents the average value of the quantity on the ordinate of that panel, weighted by the product of the number of pixels and the effective exposure time.

In those cases where both the X–ray and the optical centers of low luminosity
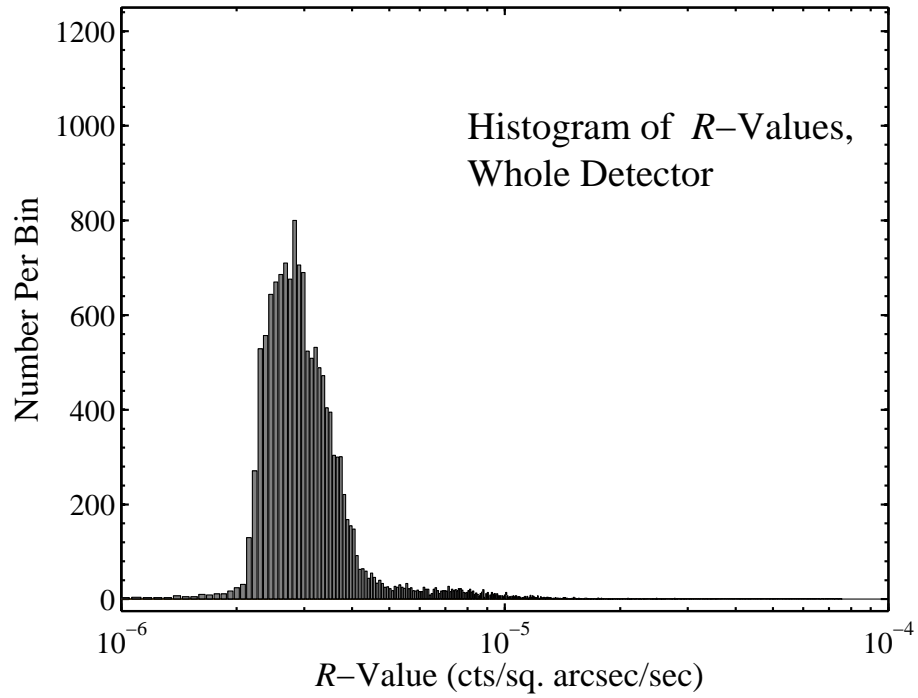
Figure 2.1 Histogram of the result of 20,000 random samples of the spatial variability of the D–band background. The number of samples in each $\mathcal{R}$–bin is plotted. The bin–size is 2×10^{-7} counts per square arcsecond per second. The mode, median, and mean values in this histogram are consistent with other measures of the background.
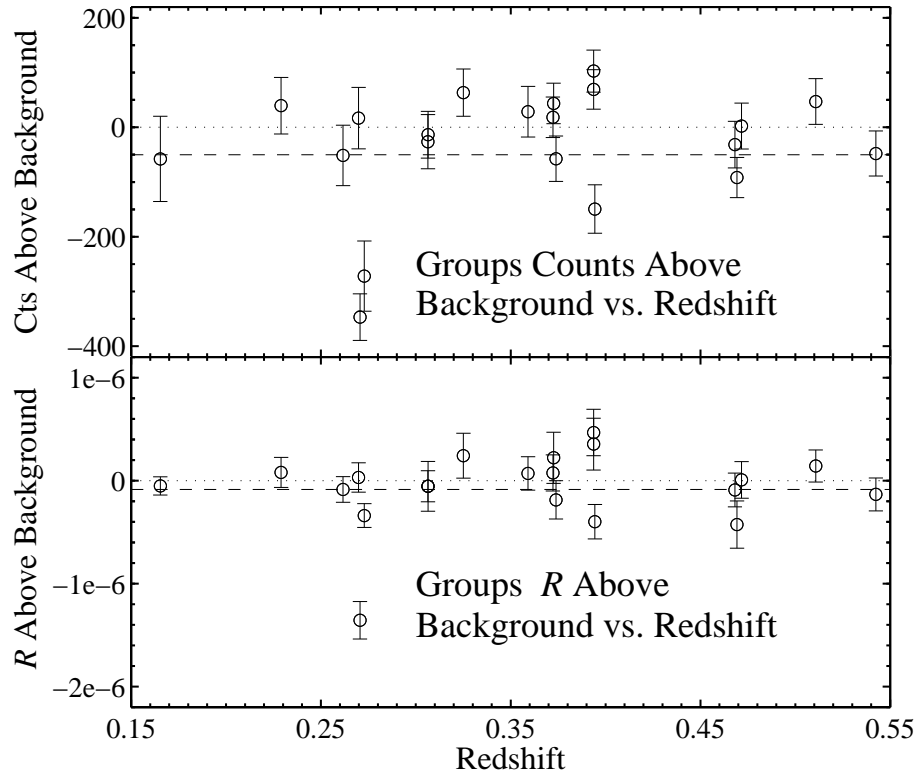
Figure 2.2 Groups intensities above background in D–band (0.2–1.0 keV). *Top Panel*: Counts in groups above estimated local background. *Bottom Pannel*: Surface brightness of groups above background, in units of $\mathcal{R}$–value – i.e., counts per square arcsecond per second. None of the groups is seen more than 5σ above the background, which is our minimum criterion for detection. The dotted line in each panel is the zero line. The dashed line in each panel represents the weighted average value of the quantity on the ordinate of that panel.

groups are known, the correlation between the two shows considerable variance (Mulchaey 2000). Because of the possibility that the X–ray centers might not coincide with the optical centers of the groups in our survey, we paid particular attention to the analysis with an aperture of radius 500 kpc around each group, because an aperture this size should be large enough to capture a large fraction of the X–rays even if there is an offset of $\sim 0.5'$ between the X–ray center and the reported optical center. Inasmuch as no group had an $\mathcal{R}$–value 5σ or more above the local background, the results were substantially the same (no plot shown). Moreover, a greater fraction of groups had negative SB–values than when the 250 kpc radius apertures were used (15/21 vs. 11/21).

We finally note that the groups might potentially contaminate the local background estimator. Using reasonable models for the intensity and the spatial distribution of emission from groups, described in detail in § 2.5.2 below, we estimate that $\geq 30\%$ of flux comes out within our aperture, and $\leq 10\%$ of flux comes out in the annulus from 1.05 to 1.20 times the aperture radius. This portion of group flux that falls within the annulus that we use to define the local background ought to increase the local background measurements, and therefore to depress the surface brightness measurements. The depression, however, is not severe: since the group surface brightness profile is expected to peak toward the center, the surface brightness is still expected to be positive. Moreover, the average $\mathcal{R}$–value of the local background samples was 2.9×10^{-6} cts arcsec^{-2} s^{-1}, which is only slightly more than the modal $\mathcal{R}$–value of the randomly placed apertures, and slightly less

than their median value. It therefore does not appear that the local background estimator was significantly increased by flux from the groups.

As a final test of whether the estimate of the background was contaminated by group–flux, we used an annulus farther from the aperture: the annulus from 1.5 to 2.0 times the aperture radius. This change had no important effect on our results.

2.4   Spectral Analysis

We also explored whether examining the shape of the X–ray spectrum would enhance the contrast to the background in the search for groups. In the simplest possible test, we investigated whether the spectral energy distribution (SED) of photons received from the locations of groups differed from that of photons from elsewhere. The background consists of the particle background, and the combined diffuse glows of our galaxy, the Galactic halo, the local group, and extragalactic point–sources and diffuse sources (McCammon & Sanders 1990; McCammon et al. 2002).

Spectral results from our observations are presented in Figure 2.3. In our data, the spectrum of the groups is indistinguishable from that of the background. This can be seen in several ways. The top panel of Figure 2.3 show, with a bin–size of 10 eV, the stacked spectrum of the groups in blue, the spectrum of the background in green, and the "rest–framed" spectrum of the groups in red (that is, the energy of each photon from the groups is multiplied by $(1 + z_{\mathrm{group}})$ prior to binning). The groups and the background both show broad emission features

centered at ∼ 0.55 keV and ∼ 1.5 keV (top panel of Figure 2.3), that are most likely due to O VII at redshift 0 and neutral Si on the detector, respectively. The difference spectrum, however, is 0 to within the 1–σ error bars (bottom panel), indicating that the SED from regions of groups is identical to that of other regions. Furthermore, the rest–framed groups spectrum shows no signs of any spectral features.

2.5   Discussion

In recent years, research efforts have increasingly been devoted to characterizing the extended X–ray emission from collapsed halos of a variety of masses, from clusters to groups to galaxy–mass halos. For two examples (among many) of studies of gas in higher–mass systems – clusters, Akahori & Masai (2005) investigate the L_X–T_X scaling relation describing intracluster gas, and Arnaud, Pointecouteau, & Pratt (2005) investigate the M–T scaling relation of intracluster gas.

Our study is concerned with the properties of gas in lower–mass systems. Xue & Wu (2000) attempt to bridge the gap between clusters and groups in the local universe by performing an analysis of the L_X–T, L_X–σ_v, and σ_v–T relations of 66 groups and 274 clusters taken from the literature. The *XMM*–Large Scale Structure survey identified groups and clusters out to redshift 0.6 and beyond, and Willis et al. (2005) examine the L_X–T_X relation of the objects in their survey. Interestingly, in contrast to the analysis in this paper, they detect X–ray emission from low–mass objects out to $z = 0.558$. The groups that they found were X–ray selected, so the selection effect surely contributed to their finding groups in their survey while we
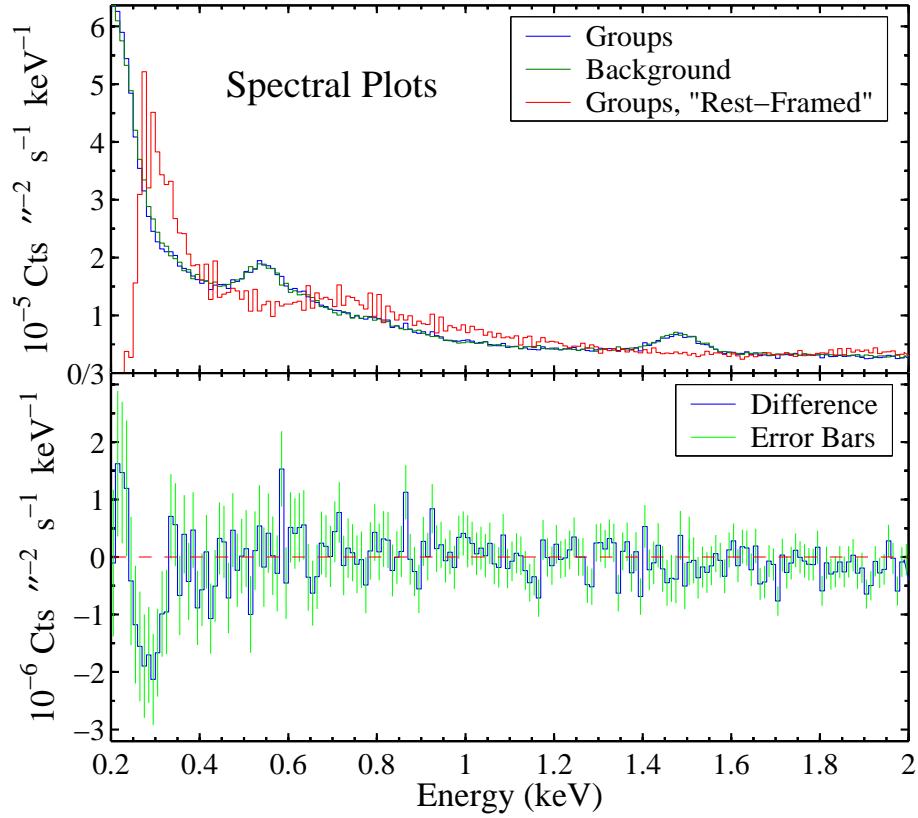
Figure 2.3 Spectra showing $\mathcal{R}$–values per 10 eV bin. The rest–framed groups show no signs of any spectral features *Top Panel*: Groups, background, rest–framed groups spectra, 0.2–2 keV. Spectrum for locations of groups is in blue, for background is in green, and for rest–framed groups is in red. *Bottom Panel*: Difference spectrum, 0.2–2 keV. 1–σ error bars are in green.

did not in ours (although we note that this does not explain the physical difference that is responsible for the discrepancy in X–ray luminosities).

2.5.1 Fits to L_X–σ_v Data

In order to estimate how our data ought to have appeared, we need to know what X–ray luminosity should be expected from a group with a given mass or velocity dispersion. The groups of the M03 catalog are a convenient sample of nearby groups with known velocity dispersion and X–ray luminosity. Since any fit to the L_X–σ_v relation in the M03 groups will show scatter, a fit is not a line or curve but a 2-dimensional region in L_X versus σ_v space.

In fact, properly, one would want to construct a 2D probability distribution function on L_X–σ_v space based on the distribution of the M03 sample. This function would give the probability density for a group drawn from the same population as the reference sample to have a particular velocity dispersion and X–ray luminosity. One could then properly examine the likelihood that the *CNOC2* groups are drawn from the same population as the M03 groups.

In practice, it would be difficult to construct the 2-D distribution function we require, because the M03 sample is not dense enough in L_X–σ_v space. We therefore reverted to a conventional parametric analysis, based on linear fits to $(\log \sigma_v, \log L_X)$ data. We will present several linear fits to the M03 data, and we describe in detail our choice for which linear fit is most appropriate.

PT04 examine the relationship between the X–ray luminosity (L_X) and the ve-

locity dispersion (σ_v) of galaxy groups in the M03 catalog. They perform regression analysis on $\log L_X$ and $\log \sigma_v$ and find that, as with any regression in which there is significant scatter, which line is considered the best–fit depends on which variable is considered to be independent and which dependent. We repeated their analysis of the M03 groups and obtained best–fit lines that are very similar (though not identical) to those that they obtained. In this paper, we present our own fits to the data. If velocity dispersion is taken to be the independent variable, the best–fit regression line is

$$\log\left(L_X/1 \text{ erg s}^{-1}\right) = 36.8 + 2.08\log\left(\sigma_v/1 \text{ km s}^{-1}\right). \tag{2.5}$$

If, on the other hand, luminosity is taken to be the independent variable, the best–fit regression line (what PT04 refer to as the "inverse regression line") is

$$\log\left(L_X/1 \text{ erg s}^{-1}\right) = 30.9 + 4.50\log\left(\sigma_v/1 \text{ km s}^{-1}\right). \tag{2.6}$$

Data from M03 and best–fit lines are displayed in Figure 2.4.

Which regression represents the "true" relation between the X–ray luminosity and the velocity dispersion of groups of galaxies? In Appendix I, we present a more detailed analysis of the relative merits of various procedures for fitting lines to data. The result of this analysis is that neither direct nor inverse regression is ideal, because surely the measurements of both σ_v and L_X were subject to errors, and furthermore there is undoubtedly intrinsic scatter in the L_X–σ_v relation. The better

relation out of these two is the one in which the variable taken to be independent has lower relative uncertainty. In our case, this is probably the inverse regression, because a velocity dispersion determined from only a few galaxies is likely to be quite uncertain. Furthermore, the steeper (inverse) relation comes closer to fitting the high–σ_v, high–L_X end of the distribution of groups, which is the end that we are most interested in when looking for groups at $z \sim 0.5$. Still, although the inverse regression is probably preferable to the direct regression, since there is scatter in both variables, this situation calls for least squares orthogonal distance fitting (described in Appendix I). If we assume that the typical relative errors on $\log(\sigma_v)$ and $\log(L_X)$ are the same (i.e., the error is shared equally between the two variables), then the distance–fit line,

$$\log\left(L_X/1 \text{ erg s}^{-1}\right) = 31.5 + 4.25\log\left(\sigma_v/1 \text{ km s}^{-1}\right), \tag{2.7}$$

is nearly as steep as the inverse regression line, as shown in Figure 2.4.

2.5.2 Limits on Group Luminosities

Since none of the groups was visible at 5σ above the background, we may place limits on their fluxes and therefore, since we know their redshifts, on their luminosities. The total number of background counts TC_{BG} at the location of a group is estimated from the smoothed background image. Assuming Poisson statistics, the standard deviation on this number is $\sigma_{TC_{BG}} = (TC_{BG})^{1/2}$. The maximum number of

group counts TC_G, then, is limited by

$$TC_G < 5\sigma_{TC_{BG}} = 5\sqrt{TC_{BG}}, \tag{2.8}$$

We estimate each group's temperature with the standard relation given in, e.g., Mulchaey (2000):

$$T = \frac{m_p}{\mu}\sigma_v^2 \times (k\beta), \tag{2.9}$$

where m_p is the proton mass, μ is the mean molecular weight, k is Boltzmann's constant, and β, whose empirical value tends to be around 2/3, is the eponymous parameter of the isothermal–β model for density and may be considered to be defined by this equation. (For low-density objects, β has been measured to be closer to 1, but for analytical simplicity we used $\beta = 2/3$ for most calculations.) The β density model for a spherical isothermal plasma is

$$\rho(r) = \rho_0\left(1 + \left(\frac{r}{r_c}\right)^2\right)^{-3\beta/2}, \tag{2.10}$$

where ρ_0 is the central density, r is the radial distance from the sphere's center, and r_c is the "core radius," a distance within which the density is nearly constant (approximately ρ_0). Using the *Astrophysical Plasma Emission Code* (*APEC*) model in *XSPEC*, we solve for what X–ray luminosity, at a given temperature, would lead to the number of group counts given by the maximum possible value of TC_G in (2.8).

In Figure 2.4, the X's show the maximum luminosity that each group could

have had during our observations (if the group were more luminous, it would have shown up at $> 5\sigma$ above the background). If a group is predicted to be less luminous than the luminosity represented by its X, then it is no surprise that we failed to see the group; conversely, if a group is predicted to be more luminous than the luminosity represented by its X, then the group is less luminous than predicted. The dashed line shows the L_X–σ_v direct regression line, the dashed–dotted line shows the inverse regression line, and the solid line shows the distance–fit line. All X's (all groups) lie above the direct regression line, so *if* that line represents the true relationship between the variables, there is no surprise in seeing none of the groups. Three groups, however, lie below both the inverse regression line and the distance–fit line; so if those lines represent the true relationship between the variables, it might be surprising that we failed to see three of the groups. In the remainder of this section, we will try to quantify how surprising it is.

2.5.3   Quantifying Surprise

The *CNOC2* catalog of small groups (Carlberg et al. 2001b) contains groups with as few as three galaxies in them. Any measurement of the velocity dispersion of such a group will necessarily have large uncertainties. In fact, for some groups, the estimated uncertainty in velocity dispersion is actually greater than the velocity dispersion itself. Because the symbol σ is frequently used to denote both uncertainty and velocity dispersion, it is inelegant to represent the uncertainty in velocity dispersion, but, eschewing elegance, we shall use σ_{σ_v} to denote uncertainty on σ_v.
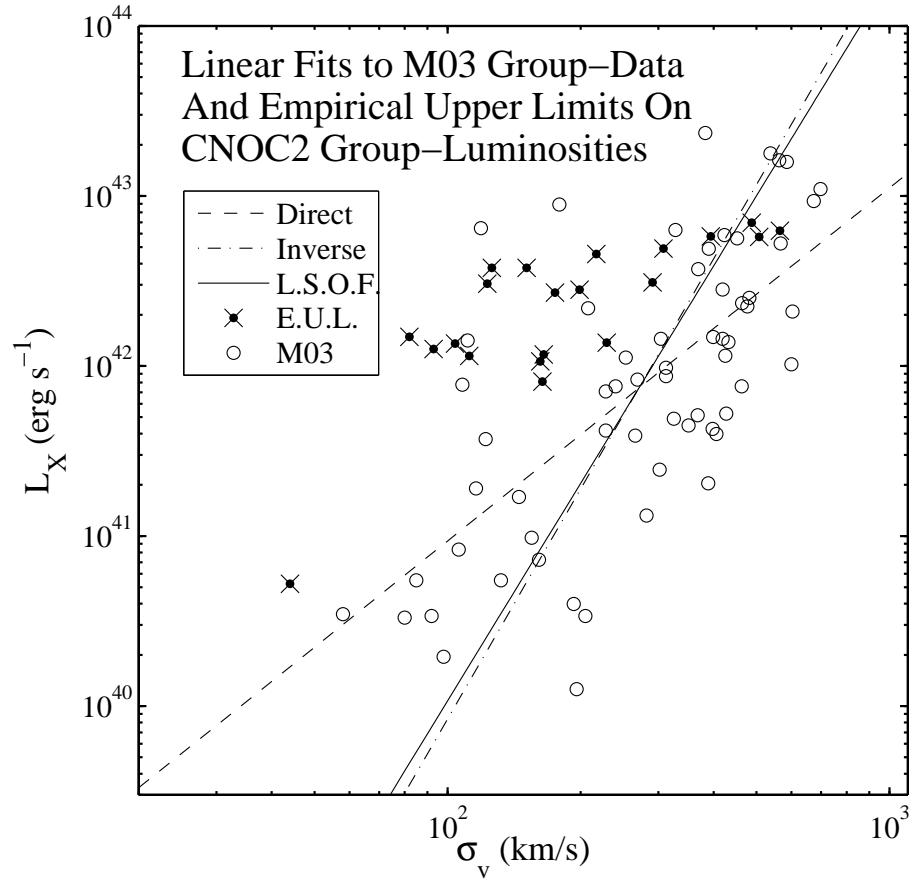
Figure 2.4 Linear Fits to M03 Group–Data And Empirical Upper Limits On CNOC2 Group–Luminosities: bolometric X–ray luminosity vs. line–of–sight velocity dispersion. The circles show the M03 groups. The dashed line is the direct regression line, the dashed–dotted line is the inverse regression line, and the solid line (L.S.O.F.) is the least squares orthogonal fit line. These three lines are fits to the M03 data. The X's (E.U.L.) display the "empirical upper limits" to groups' luminosities, i.e., the maximum luminosities that the *CNOC2* groups could have had during our observations without showing up at more than 5σ above the background. The uncertainty in velocity dispersion associated with the X's is typically nearly $\sim 100\%$, as shown in Table 2.1. Three of the X's lie beneath both the dashed–dotted and the solid lines, indicating that these three groups were predicted to be luminous enough to be detected, and yet were not.

It is possible that our null result – our failure to detect any of the groups in our survey – occurred because the velocity dispersions of the three groups predicted to be observed were overestimated. If so, then perhaps their predicted luminosities based on, e.g., (2.7), should have been low enough that the groups should not have been seen after all. Of course, if there were no bias to the measurements of σ_v, it is just as likely that the velocity dispersions of these three groups were *underestimated*, which would make their non–detection even more surprising. Furthermore, in order for this sort of statistical error to explain our null result, not only would the velocity dispersions of the three groups that we think we ought to have detected need to have been sufficiently overestimated, but no *other* groups could have have had their velocity dispersions underestimated by too much. In short, the statistical errors would need to be in particular directions on particular groups. If there is no non–statistical reason for measurements to be in error, then what is the likelihood that the types of error needed to explain our null result actually occurred?

To answer this question, we simulated 10,000 mock–*CNOC2* catalogs, each with its own set of mock–velocity dispersions ($\{\widehat{\sigma_v}(i)\}$) associated with the mock–groups. If the set of measured velocity dispersions from the actual *CNOC2* catalog is $\{\sigma_v(i)\}$, then, for each mock–catalog, the simulated velocity dispersion of the ith group $\widehat{\sigma_v}(i)$ was drawn from from a gamma distribution with mean $\sigma_v(i)$ and standard deviation $\sigma_{\sigma_v}(i)$. The gamma distribution has the attractive property that the mean and the standard deviation may be independently specified. Because the tail of the gamma distribution extends to infinity, however, a small fraction of mock–

groups were initially given unrealistically high velocity dispersions. Incidentally, this is preferable to the normal distribution, which has another tail that extends below zero to negative infinity – unphysical values for velocity dispersion. To deal with the gamma distribution's problem of predicting a few unrealistically high values, we, for each group i, eliminated from the set of mock–catalogs all mock–groups with velocity dispersions more than $4\sigma_{\sigma_v}(i)$ above the mean ($\sigma_v(i)$), and replaced these values with new gamma–distributed variables. We repeated this procedure iteratively until there were no more mock–groups with velocity dispersion more than $4\sigma_{\sigma_v}$ above the mean. Because this procedure involves replacing the high–end outliers with values closer to the mean, it had the effect of slightly depressing the mean. All values were therefore increased by the difference between the new mean and $\sigma_v(i)$, thereby raising the mean back to $\sigma_v(i)$, as it should be.

We also tried several other random distributions for assigning mock–velocity dispersions to our mock groups. We tried a gamma distribution with no high–end cut–off; and we tried several modifications of the normal distribution, all of which included a low–end cut–off at 0. We found that those results were not importantly different from results from the gamma distribution with the high–end cut–off, which is the analytically simplest distribution that produces realistic results. Results are presented only for the mock–catalogs generated with the high–end cut–off gamma distribution.

Each mock–group was assigned the same aperture size as the corresponding real group. Its mock–intensity was spatially distributed according to a spherical

isothermal–β model. Since emissivity is proportional to the square of density ρ, the surface brightness or intensity I is proportional to the integral along the line–of–sight of ρ^2. At a projected viewing "impact parameter" b from the sphere's center,

$$I(b) \propto \int_{r=b}^{r=r_h} \rho(r)^2 \mathrm{d}s,$$

where r_h, the "halo radius," is the outer boundary of the sphere of plasma, and s is the variable along the line–of–sight. Substituting for ρ with equation (2.10) and performing the integration results in the following expression for the surface brightness:

$$\mathcal{B}(x, A, \beta) \propto (1 + x^2)^{-3\beta}(A^2 - x^2)^{1/2}\, {}_2F_1\left(\frac{1}{2}, 3\beta, \frac{3}{2}, \frac{x^2 - A^2}{1 + x^2}\right), \qquad (2.11)$$

where $x \equiv b/r_c$ and $A \equiv r_h/r_c$ are substitutions to simplify the integration, and ${}_2F_1$ is the Gauss hypergeometric function. Three quantities go into the normalization of $\mathcal{B}$ for each mock–group: its mock–luminosity from e.g. (2.7); its luminosity distance $D_L(z)$; and its core radius. Finally, we assumed a foreground column density of neutral hydrogen of 2×10^{20} cm^{-2} (Dickey & Lockman 1990).

It was unclear what core radii our mock–groups should be assigned, because the β-model core radii of groups and clusters vary a great deal. Ota & Mitsuda (2004) report that in a sample of 79 clusters, including 45 "regular" clusters and 34 "irregular" clusters, the mean of all 79 values of r_c is $(163 \pm 202)\, h_{70}^{-1}$ kpc, midway between the mean r_c of the regular clusters $((76 \pm 60)\, h_{70}^{-1}$ kpc) and the mean r_c of

the irregular clusters ($(273 \pm 259)\ h_{70}^{-1}$ kpc). Because there is no strong correlation between r_c and group–mass, we think it is reasonable to adopt a uniform assumed value of r_c for the mock–groups in our mock–catalogs. We ultimately decided to set each mock–group's core radius to 125 kpc, because this value lies in the middle of the range of values reported by Ota & Mitsuda (2004). So long as the value of r_c that we adopted is physically plausible, the precise value does not matter much, because the predicted number of counts from a group, within an aperture of fixed size, is not very sensitive to the assumed value of r_c, as long as the angular–size of the aperture is larger than that of the core radius.

The SED of each mock–group (and hence the flux in a given energy band) is calculated with *XSPEC*, using *APEC*, with its mock–temperature calculated from equation (2.9). For A, we tried values between 2 and ∞, and for β, we tried values between 0.6 and 1.0. The results presented below are for $A = \infty$ and $\beta = 2/3$, and for D–band (0.2–1.0 keV).

In brief, the results of this Monte Carlo indicate that the probability appears to be small, but not vanishingly so, that statistical fluctuations in the measured values of σ_v explain our null result. Figure 2.5 shows a summary of the mock–catalog results. The median and the mode number of groups seen per catalog are both 3, and the mean number seen is 3.1, which is in good agreement with the prediction of section 2.5.2. The finding of our observation – that no groups were seen – occurred in only 3.5% of the mock–catalogs. If there were no systematic errors in the measurements of velocity dispersions of the *CNOC2* groups, then it

is unlikely but not impossible that the only reason we had expected some of the groups to be visible in the first place was erroneously high measurements of σ_v.

2.5.4   Redshifted Mulchaey Groups

Perhaps we have placed too much emphasis, in sections 2.5.1 and 2.5.2, on which best–fit line actually fits the M03 data the best – when the truth is that there will be significant scatter to the data around any line drawn through the cloud of points. An alternative way to determine whether it is surprising that no groups in our survey showed up at 5σ above the background is to ask whether the *CNOC2* groups and the M03 groups were drawn from the same population. We addressed this question by investigating how many of the groups in the M03 atlas would have been visible at the 5σ level if they had been in our observation.

The M03 groups are all in the local universe, approximately at redshift 0. If they had been in our survey, i.e., between redshifts 0.1 and 0.6, they would have both been dimmer and subtended a smaller solid angle on the sky. Furthermore, they would have been observed with *XMM–Newton*, which, owing to its greater focal ratio, has a higher particle background rate than *ROSAT* (the instrument with which they were originally observed).

There were 109 groups in the M03 catalog, of which 61 were seen in X–rays. We simulated observations of each of these 61 M03 groups at each of the 21 redshifts of the *CNOC2* groups, ranging from $z_1 = 0.165$ to $z_{21} = 0.543$ (ID numbers taken from
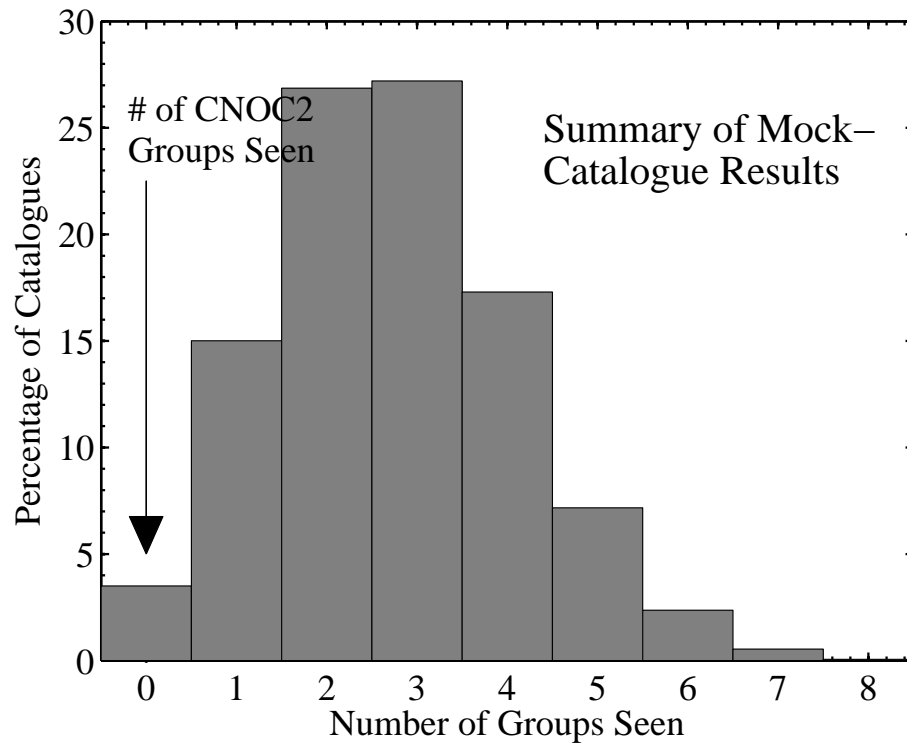
Figure 2.5 Histogram of the results of the mock–catalog analysis. Of the 10,000 mock–catalogs, 96.5% contained at least 1 mock–group visible at the 5σ level. In the actual *CNOC2* survey no groups were visible at this level.

Table 2.1). As before, our criterion was that a group must outshine the background by at least 5σ to qualify as being detected in our simulated observations. Each group was assumed to be at a location with detector–averaged characteristics: the mean exposure map value, the median background value (from Figure 2.5), the mean amount of smearing due to the point–spread function, and the mean amount of lost usable detector–area because of bright point sources. Much like in the mock–catalog calculations described in section 2.5.3, we calculated the count–rate of each group with *XSPEC*, using *APEC*, with the group's temperature calculated from equation (2.9). We distributed emission according to the isothermal–β profile using equation (2.11), with $\beta = 2/3$, $A = \infty$, and $r_c = 125$ kpc. For our simulated observations, we used an aperture of 250 kpc, and looked in D–band. For background, we used $\mathcal{R}_{BG} = 3.0 \times 10^{-6}$ cts arcsec^{-2} s^{-1}. Finally, as before, we assumed a foreground column density of neutral hydrogen of 2×10^{20} cm^{-2}. Within our simulation, the number of M03 groups that exceeded the 5σ detection criterion varied from 22 of the 61 total groups (or $p_1 = 36\%$) at the lowest redshift (z_1), to 2 of 61 ($p_{21} = 3\%$) at the highest redshift (z_{21}).

We identify two populations of the Mulchaey groups – all 109, and the 61 that were seen in X–rays. In Appendix II, we present in detail our estimate of the likelihood that the *CNOC2* groups and the M03 groups were drawn from the same population. The result is that the probability of obtaining our null result on the assumption that the *CNOC2* groups were drawn from the same population as the whole M03 catalog is 14.7%; and the analogous probability on the assumption that

the *CNOC2* groups were drawn from the same population as the 61 X–ray–bright M03 groups is 2.8%. These probabilities (especially the latter) are low enough to be interesting, but are certainly not dispositive.

2.6   Conclusion

The analysis in section 2.5.4 is mildly suggestive that the groups of the *CNOC2* survey constitute a sample that is on average less luminous than the sample of groups observed by Mulchaey and collaborators in the 2003 atlas. The interesting comparison, however, is not between two particular samples, but rather between the population of groups at intermediate redshift and the population in the local universe.

A first caveat is that the two samples were selected in very different ways, and there is no *a priori* reason to suspect that their underlying populations have similar luminosity functions. Furthermore, it is conceivable that for some reason the M03 groups should be expected to be a more luminous sample of their population than the *CNOC2* groups are of theirs. Perhaps the M03 groups are on average more massive (although the ranges and distributions of velocity dispersions for the two samples are nearly identical). Do the M03 groups contain more galaxies than the *CNOC2* groups? Certainly, Mulchaey reports more galaxies per group on average – including as many as 63 galaxies in one group. Of the *CNOC2* groups in our field of view, none contained more than 4 galaxies with identified redshifts. But most or all of the difference here is almost certainly due to the difficulty of measuring

spectroscopic redshifts of faint galaxies at large distances; we therefore see no reason to think that the M03 groups are on average more massive than the *CNOC2* groups.

An obvious criticism of the present work is that the *CNOC2* groups might not be groups at all, but just chance superpositions of galaxies. While this would clearly explain the lack of X–ray emission, we will argue that we find it unlikely to be the reason why we detected no hot gas.

Carlberg et al. (2001b) identified as a group any object in the *CNOC2* Survey that contains at least 3 galaxies in sufficiently close redshift–space proximity. This identification procedure does leave open the possibility of confusion: if two galaxies are separated by a large distance, but have appropriately large peculiar velocities, they could have similar recessional velocities (redshifts) despite not being part of a gravitationally bound structure. Or, even if a few galaxies are close to one another in physical space, they will not be part of a bound system without a dark matter halo that is massive enough to bind them.

In searching for groups and clusters, it is often assumed that where there are associations of galaxies, there is dark matter to bind them, but this assumption could be wrong. With rich clusters, there are multiple independent checks of this assumption: the temperature of intergalactic gas (which can be detected through its X–ray emission or through inverse Compton distortions to the cosmic microwave background) is one indicator of the mass; the velocity–distribution of galaxies is another indicator of the mass, and in particular the distribution should be Maxwellian

(or Gaussian) for a virialized structure; and the degree of gravitational lensing of background galaxies is a third indicator of the mass.

The first check has obviously failed in this case – for these 21 groups, there is no apparent hot gas component. The second check is useless for an association of very few (3–6) galaxies – nearly any velocity distribution is consistent with an underlying Gaussian distribution. How about lensing? For structures of $\lesssim 10^{13.5}M_\odot$, noise prevents detection of a lensing signal. As a result, in the *CNOC2* fields, it is only possible to determine a statistical lensing mass for the groups by stacking them and determining their combined lensing signal. This analysis has been performed by Hoekstra et al. (2001, 2003): their calculated estimate of the average velocity dispersion per group based on the mass inferred from the average shearing of background galaxies (258^{+45}_{-56} km s^{-1} for $\Omega_m = 0.2$ and $\Omega_\Lambda = 0.8$) is in good agreement with the average velocity dispersion of 230 km s^{-1} from the kinematics of the groups' member–galaxies. The observed lensing signal was sufficiently strong that chance projections of unbound galaxies alone would have difficulty mimicking it. Some of the detected lensing mass is therefore almost surely in the form of galaxies orbiting within a parent dark matter halo – groups. It is entirely reasonable to expect part of this group–component of the lensing mass to consist of hot gas trapped within the group potential. On the other hand, it is not unlikely that a portion of the lensing mass is in the form of projected, yet physically unassociated, galaxies. Although there may be limited amounts of gas associated with the individual galaxies, such projections will produce no extended group emission. Therefore,

the presence of a lensing signal does not automatically lead to the expectation of group–like X–ray emission that is at sufficient surface–brightness to be detectable in our survey.

It is conceivable that there were several relatively massive groups, or dark clusters (i.e., cluster–mass objects with very low mass–to–light ratios), that dominated the mean mass measurement. Although we cannot rule out the possibility that a large fraction of the mass is in only a few objects that happened to fall outside the field of view of our observations, and that the remaining objects are not gravitationally bound structures but rather chance associations of galaxies that just happen to be near one another in redshift–space, this seems somewhat unlikely: Combinatorial statistics dictate that if there were a few massive halos scattered among the 40 alleged groups and none of the rest are true groups, then the probability is low that all the massive ones ended up among the 19 groups that did not fit into our field of view. Consider, for instance, if there were only four massive halos: since 19 is roughly half of 40, if the four massive halos were randomly distributed the probability that all four would be out of our field of view is $\lesssim (1/2)^4$. Furthermore, a calculation in Carlberg et al. (2001b) shows that the number of alleged groups in the field in the redshift range $0.1 < z < 0.6$ (40 in the *CNOC2* field, 21 in the *EPIC–PN* field of view) is roughly consistent with predictions. The concordances both of the observed number of groups with the predicted number, and of the observed velocity dispersions with the mass inferred from weak lensing, together constitute a strong indication that the *CNOC2* group sample does not suffer significant con-

tamination from chance superpositions of galaxies. An important caveat is that the number of groups in the *CNOC2* fields in the high σ_v range is in excess of the aforementioned predictions derived from the basic formalism of Press & Schechter (1974) by a factor of ~ 4. Based on the Carlberg et al. (2001a) results, ~ 80% of such high σ_v groups could be false, and our own null results are consistent with this hypothesis.

If we take seriously the conclusion, for which our analysis has generated some modest evidence, that galaxy groups at intermediate redshift are less X–ray luminous than those at redshift zero, we can think of several potential explanations. Groups at $z \sim 0.5$ are younger than their relatives in the local universe, which leads to the following three possibilities: (*i*) Since simulations and observations indicate that groups and clusters at the present are accreting gas from cosmic filaments, groups at a younger stage of the universe would presumably have been accreting gas for less time than the more evolved groups of the present–day universe and might therefore contain less gas. (*ii*) Even if the intragroup gas in young groups is not lower density than the gas in groups today, if the young groups are still in the process of heating up – if they have not yet reached their virial temperatures – their luminosities would be reduced. (We find this possibility to be unlikely, because the heating timescale is short relative to a Hubble time, even at $z \sim 0.5$.) (*iii*) Even if the gas in young groups is on average the same density and the same temperature as the gas in nearby groups, if it is of lower metallicity its X–ray emissivity would be reduced. There are undoubtedly feedback processes that enrich the intragroup

medium, and pristine gas of primordial abundance is also surely constantly falling onto groups, so it is not clear whether to expect higher metallicity in the IGM of groups that are at $z \sim 0.5$ or at $z \sim 0$. For warm gas, observed in a soft X–ray band such as our D–band, line–emission can be important, depending on abundances. For example, if a $T = 1$ keV group at redshift 0.5 contained primordial gas, its D–band X–ray flux would be only two thirds as great as if its gas were at 10% solar abundance and only half great as if its gas were at 20% solar abundance. It is possible, therefore, that the relation between velocity dispersion and X–ray luminosity ought to be shifted down by a factor of ~ 2 in Figure 2.4. Such a shift would make it less surprising that we detected no X–ray emission from the groups.

Future data on high–redshift clusters that make it possible to examine the time–evolution of the density and metallicity of the intracluster medium in low–mass clusters, and deeper searches for X–ray gas in low–mass objects at intermediate redshift, will help to answer why our survey failed to detect such gas.

APPENDIX

I Regression Reconsidered

In this Appendix, we address the question of which regression represents the "true" relation between the velocity dispersion and the X–ray luminosity of groups of galaxies. It is instructive first to consider a statistical effect that is common to all regression analysis. It has been known at least since Karl Pearson's seminal

paper in 1901 that if there is scatter (say, due to measurement error) in the variable considered to be the independent one, this will in general cause the recovered slope from regression analysis to be shallower than the slope of the true underlying relation between the variables. (PT04 demonstrate a nice example of this effect.) An important consequence of this effect is that when there is scatter in both variables, the slopes from both the direct and the inverse regression will be shallower than the respective true slopes.

For example, suppose two variables x and y are related by $y = \alpha x$ (or $x = \beta y$, where $\beta = 1/\alpha$). In an experiment, a large number of (x, y) pairs are measured, and there are statistical uncertainties in the measurements of both variables. If x is considered to be the independent variable, regression analysis will lead to $y = \hat{\alpha} x$, where in general $\hat{\alpha} < \alpha$ because of the scatter in x. If y is considered to be the independent variable, regression ("inverse regression") analysis will lead to $x = \tilde{\beta} y$, where in general $\tilde{\beta} < \beta$ because of the scatter in y. Of course, if we wish to represent y as a function of x even in the case of the inverse regression, we will write $y = \tilde{\alpha} x$, where $\tilde{\alpha} = 1/\tilde{\beta} > \alpha$. So, although the recovered slopes are in general less than the respective true slopes for both regressions ($\hat{\alpha} < \alpha$ and $\tilde{\beta} < \beta$), when we fix the abscissa and ordinate variables we get a shallower slope than the true one in the case of the direct regression and and a steeper slope in the case of the inverse regression. Note that this effect holds whether the scatter comes from measurement error or is intrinsic to the variables.

When there is scatter in both variables, so that neither one is a proper inde-

pendent variable, it is sometimes appropriate to use a type of best–fit line that depends on the geometric orientation of the points. Instead of finding the line that minimizes the summed, squared *vertical* deviations of the data from the line, as in the case of regression, some situations are better analyzed by finding the line that minimizes the summed, squared *perpendicular* distances of the data from the line. This is sometimes called "least squares orthogonal distance fitting," and was the subject of Pearson (1901). Because this type of fitting, in contrast to linear regression, is unfortunately sensitive to the units of measurement, it is advisable to normalize the data prior to fitting. The most common normalization is by the estimated errors or uncertainties (σ_x and σ_y) on the variables (Press et al. 1992; Akritas & Bershady 1996). The direct regression line will have the minimum slope, the inverse regression line will have the greatest slope, and the "distance–fit" line will have an intermediate slope (closer to the direct regression slope if the scatter in x is smaller; closer to the inverse regression slope otherwise). All three lines intersect at the centroid of the data.

For completeness, we present the equation of the least squares orthogonal distance fit line. Let us express the measured data as a set of N points $\{(x_i, y_i)\}_{i=1}^N$. To simplify the form of the line's equation, we define the following sums of data:

$$S_x \equiv \sum_{i=1}^{N} x_i \qquad S_y \equiv \sum_{i=1}^{N} y_i$$

$$S_{x^2} \equiv \sum_{i=1}^{N} x_i^2 \quad S_{xy} \equiv \sum_{i=1}^{N} (x_i\, y_i) \quad S_{y^2} \equiv \sum_{i=1}^{N} y_i^2; \tag{2.12}$$

and we define the following combinations of these sums:

$$A \equiv S_x S_y - N S_{xy}$$

$$B \equiv (S_x)^2 - (S_y)^2 - N(S_{x^2} - S_{y^2})$$

$$C \equiv -S_x S_y + N S_{xy}. \tag{2.13}$$

If the equation of the least squares orthogonal distance fit line is

$$y = mx + b, \tag{2.14}$$

then the slope of the line is given by

$$m = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}, \tag{2.15}$$

and the y–intercept is given by

$$b = \frac{S_y - m S_x}{N}. \tag{2.16}$$

The expression for m in equation (2.15) contains a $\pm$–sign from the solution to a quadratic equation. One solution for m gives the line that minimizes the summed squared orthogonal distances to the points, and the other solution gives the line passing through the centroid of the data that maximizes the summed squared orthogonal distances to the points.

II Comparison: *CNOC2* vs. M03 Groups Samples

In this Appendix, we address how to compute the probability that the *CNOC2* groups and the Mulchaey (M03) groups derive from the same population. In order to do this, we find the properties of the hypothetical parent group population $\mathcal{P}$ that maximize the joint probability of our results and Mulchaey's results.

Let the redshifts of the *CNOC2* groups be labeled $\{z_k\}_{k=1}^{21}$. At each redshift z_k, let N_k be the number of M03 groups that our simulation predicts would have been visible in our 110 ksec observation, and let p_k be the probability that a group drawn from $\mathcal{P}$ will be visible in our survey if it is in our field of view and at redshift z_k.

At redshift z_k, the probability of obtaining our result – of not seeing any groups – is

$$q_k \equiv (1 - p_k); \tag{2.17}$$

and the probability of obtaining the M03 result – of seeing N_k groups, as predicted by our simulation – is

$$B(M, N_k, p_k) = \binom{M}{N_k} p_k{}^{N_k} q_k{}^{M-N_k}, \tag{2.18}$$

where M is the total number of groups in the M03 sample, either 109 or 61, as described above.

The joint probability, then, of obtaining our result and of obtaining the M03 result is

$$P_{\text{joint}}(z_k) = q_k B(M, N, p_k) = \binom{M}{N_k} p_k{}^{N_k} q_k{}^{M-N_k+1}. \tag{2.19}$$

At each redshift z_k, we find the binomial probability $\hat{p}_k$ that maximizes $P_{\text{joint}}(z_k)$.

This set of optimal probabilities $\{\hat{p}_k\}$ can be thought of as defining the parent population $\widehat{\mathcal{P}}$ of groups that maximizes the probability that we would obtain our null result and that the M03 groups would have their observed properties. Each percentage $\hat{p}_k$ in turn can be thought of as an estimate of the probability that a group would be visible in our survey if it were randomly selected from $\widehat{\mathcal{P}}$ and placed at redshift z_k. The value $\hat{q}_k \equiv (1 - \hat{p}_k)$, then, is the probability that such a group would *not* be detected in our survey.

If 21 groups were selected at random from $\widehat{\mathcal{P}}$ and placed at the 21 redshifts z_k, the probability that none would be seen in our survey is the product of the 21 quantities $\hat{q}_k$. A maximal estimate of the conditional probability that none of the *CNOC2* groups would be detected at the 5σ level, given that the *CNOC2* groups were drawn from the same population as the M03 groups, may therefore be represented as follows:

$$P(\varnothing \mid \text{same pop.}) \leq \prod_{k=1}^{21} (1 - \hat{p}_k), \tag{2.20}$$

where $\varnothing$ denotes our null result of not detecting any *CNOC2* groups. The result of our simulation is that the product in equation (2.20) is $P \leq 14.7\%$ for $M = 109$ and $P \leq 2.8\%$ for $M = 61$; this indicates that the probability is small, but not vanishing, that the *CNOC2* groups in our survey have the same luminosity function as the M03 groups.

Part II

Constraining Properties of Exoplanets

Chapter 3

Can We Probe the Atmospheric Composition of an Extrasolar Planet from its Reflection Spectrum in a High–Magnification Microlensing Event?

3.1 Introduction

With the first discovery, 16 years ago, of a planet orbiting a star other than our Sun (Wolszczan & Frail 1992), astronomy finally entered an age in which we could hope to answer scientific questions about distant planets, with the ultimate aim

of detecting and characterizing an extrasolar Earth-like planet. Among the most tantalizing questions is: What is the chemical composition of an extrasolar planet? In this paper, we suggest a way to test for the presence of certain compounds in the atmospheres of extrasolar planets at much greater distances than has previously been discussed.

While the vast majority of the > 280 currently known extrasolar planets have been discovered in radial velocity surveys (e.g., see Marcy & Butler, 1998 or Woolf & Angel, 1998 for reviews [1]), two methods have recently been proposed to search for extrasolar planets via their gravitational microlensing signatures. These methods are complementary to the radial velocity surveys, in that they can detect planets at larger distances, well beyond our solar neighborhood, and one of these methods has the advantage of potentially providing information about the spectrum and therefore the composition of the planet (transit surveys have the same two advantages; e.g. Charbonneau et al. (2002)).

Mao & Paczynski (1991) and Gould & Loeb (1992) suggest that as a background star passes behind a lens–star with a companion planet, the planet could be detected as lens, since it will cause a secondary, sharp spike in the source star's light–curve. Indeed, two years ago, a lens-plane planet was finally discovered (Bond et al. 2004). Recently, Graff & Gaudi (2000, hereafter GG00) and Lewis & Ibata (2000) have suggested that extrasolar planets might instead be detected in the source-plane, as they cross the caustics of a foreground lens system and are highly magnified

---

[1]For up–to–date information on the current status of these searches, see http://exoplanet.eu and http://exoplanets.org.

relative to their parent star (Heyrovsky & Loeb (1997) also discuss from a theoretical perspective the possibility of using microlensing light–curves to probe structures in the source plane, in their case the structure of a background star behind a point–like lens, and other authors since then have carried out such studies, e.g. Albrow et al. (1999), and Castro et al. (2001)).

While detecting the planet as a lens, as Mao & Paczynski suggest, has the potential to reveal a statistically important sample of extrasolar planets, the drawback is that we receive no information about the planets except for perhaps their masses and projected separations from their host stars. Reflected light from a planet, however, contains information about physical parameters of the planet (presence and sizes of rings, satellites, spots and bands, for example). Detecting a planet as a lensed source therefore holds the promise of allowing these parameters to be measured, as suggested by Gaudi et al. (2003, hereafter GCH03) (Lewis & Ibata (2000) suggest further that polarization fluctuations during microlensing events could be indicative of properties of planetary atmospheres). In the present work, we investigate the viability of detecting an extrasolar planet as a microlensed source, and the extent to which a measurement of the magnified reflection spectrum can be used to glean information about the planet's atmospheric composition.

An unperturbed, isolated point–like lens (such as a single planet or a star) produces a point–like caustic. A binary lens, however, can produce a closed caustic curve, consisting of a set of piecewise concave curves that meet in cusps. In the present context, a binary lens, then, has several advantages over a point–lens: first,

the relatively large spatial extent (compared to a point) of the binary lens caustic implies a much larger region in the sky in which for a high–magnification event to occur; second, since the caustic of a binary lens is a closed curve, caustic crossings come in pairs, and the second crossing can be anticipated; third, both star and planet can cross the caustic of a binary lens, while it is unlikely that both would cross the point–caustic of a point lens. When a background star with a companion planet crosses the caustic of a binary lens, a unique observational signature will be produced in the light–curve. If such a signature is detected on ingress (or, if the lensed light–curve shows, at least, that the star has entered the inter–caustic region of a binary lens), GG00 suggest that many observatories could train their telescopes on this system so as to obtain dense sampling of the light–curve at egress (exiting the caustic region). If the planet's reflected light is sufficiently magnified, multi–color light–curves, or even detailed time-dependent spectra might, in principle, be obtained. Such spectral binning of the signal would shed light on the wavelength-dependence of the planet's albedo, which could in turn yield information about the chemical composition of the planet's atmosphere.

GCH03 suggest that morphological features such as moons or rings around extrasolar planets may be detectable, and they find a signal-to-noise ratio of ~ 15 for *I*-band detection of a planet in a typical planet-star-lens configuration with a 10m telescope. If the light (in a given wavelength range) is split up into N bands, the signal-to-noise ratio should go down roughly as $1/\sqrt{N}$. Signal-to-noise is also directly proportional to the diameter of the telescope's aperture. This suggests

that with a 10m–class telescope, light could be split up into a few broad spectral bins before signal-to-noise becomes unacceptably low, and motivates us to examine whether useful information about the atmospheric composition of the planet could be obtained with this method.

The rest of this paper is organized as follows. In § 3.2, we present our model of a planetary caustic–crossing, including a detailed discussion of both the model of the planet and the computation of the caustic–crossing light–curve. In § 3.3, we discuss the detectability of extrasolar planets through microlensing. In § 3.4, we describe the albedos and reflection spectra of gas–giant planets in our own solar system. In § 3.5, we analyze the possibility of determining the wavelength–dependence of the albedo of a microlensed extrasolar planet. In § 3.6 we present a detailed discussion of the factors that affect the S/N of the detection of planets with various features in their reflection spectra. Finally, in § 3.7, we discuss the limitations of current technology, and conclude with projections of what may be possible with future instruments.

3.2   Modeling Planetary Caustic–Crossing Events

3.2.1   The Planet–Star System

We consider a star with a companion planet as it crosses the fold–caustic of a binary lens. Figure 3.1 shows an illustration of the configuration we model. The observed surface–brightness of the planet at a given wavelength depends on properties of
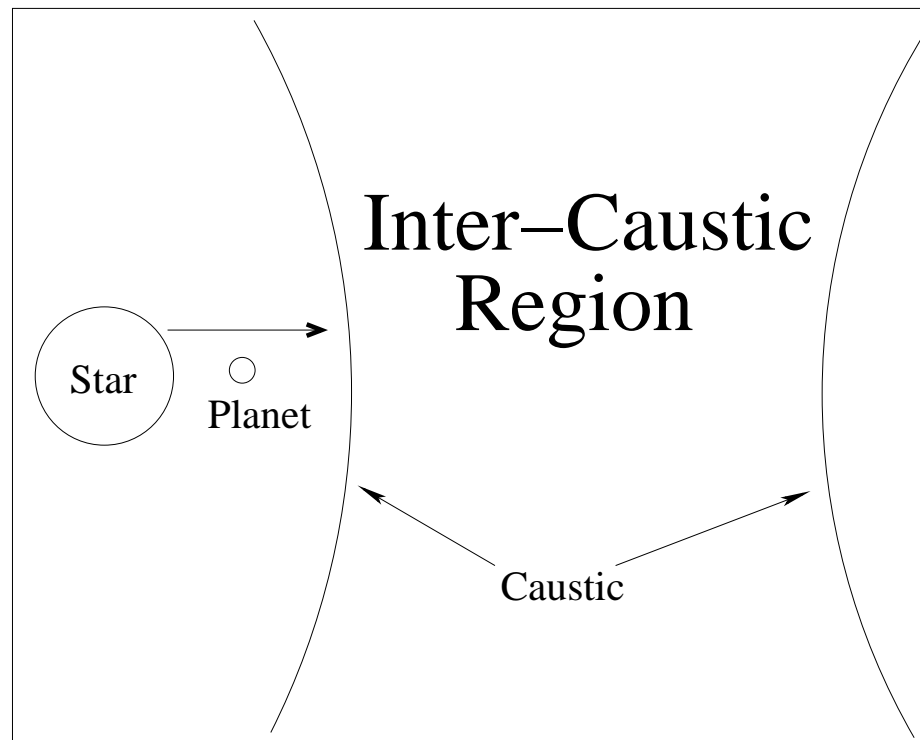
Figure 3.1 Schematic illustration of the planet–star system in the "planet–leading" configuration. The planet–star system is moving to the right, while the caustic stays still. In the reverse ("planet-trailing") configuration, the planet would be to the left of the star as they move to the right.

the star, the planet, and their relative geometry – specifically, the stellar flux, the albedo and phase of the planet, and the reflection or scattering properties of its atmosphere. For the stellar spectrum, we adopt that of a G0V star, and for the wavelength–dependent albedo of the planet's atmosphere, we use the gas giants in our own solar system as a guide (both to be described in more detail in § 3.4 below, where spectral features are considered). The planetary phase is given by the angle ϕ between the line–of–sight and the ray from the star to the planet (e.g. $\phi = 0°$ corresponds to the "full-moon" phase), as described in, e.g., GG00, GCH03, and Ashton & Lewis (2001)). GCH03 adopted and compared two simple reflectance models (uniform and Lambert scattering) that prescribe the angular dependence of reflectivity; and Ashton & Lewis (2001) considered the effects of planetary phase. Neither of these studies, however, considered simultaneously the effects of the planet's phase and its reflectance model on the lensed light–curve. In our studies, we compared three different reflectance models: uniform, Lambert, and Lommel-Seeliger reflection (see, e.g., Efford (1991) for more detailed discussions of these models, and see the Appendix for details on the computation of planet-models).

In Figure 3.2, we illustrate the surface brightness maps of three planets, one for each reflectance model described, each at fixed phase $\phi = 45°$. The maps were created numerically on a square grid of 401×401 pixels that we find to be sufficiently fine to converge on the light–curves we obtain below. We concur with GCH03 that with current technology it would be impossible to infer the true reflectance model of an extrasolar planet during a microlensing event, and so we use only one model,

Lommel-Seeliger reflectance – which we expect to be the most realistic one, in calculating the light–curves that we present below.

3.2.2 Modeling the Caustic–Crossing Event

For a good description of the details of gravitational microlensing, including see, for example, Mao & Paczynski (1991); for details regarding the generation and shape of fold-caustics, see GG00. We assume that the planet–star system described in § 3.2.1 above is in the source plane of a binary lens. The lensing stars are massive enough and close enough to one another that they generate a fold–caustic in the source-plane, a closed curve of formally infinite magnification. The caustic is considered to be a straight line (we follow GG00 and GCH03 and assume a low probability of crossing the caustic near a cusp) that sweeps across the planet and star. We assume that the plane of the planet's orbit is edge–on and is normal to the caustic at the point where the star-planet system crosses; we will argue in § 3.5 below that this simplification is not critical to our results. A source is magnified by the binary lens proportionally to the inverse square-root of the source's distance from the caustic when it is in the inter–caustic region (ICR) and not otherwise. We use the discretized magnification equation given by Lewis & Belle (1998) and Ashton & Lewis (2001):

$$\mathcal{A}_{\mathrm{pix}}(x_k) = 2\frac{\kappa}{\Delta x}\left(\sqrt{x_k + \Delta x} - \sqrt{x_k}\right) + \mathcal{A}_0, \tag{3.1}$$
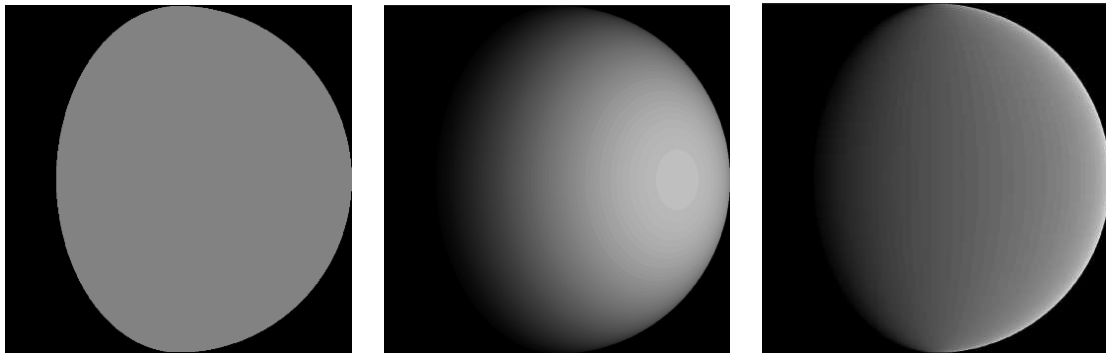
Figure 3.2 Illustration of the surface brightness of planets illuminated by a star at a phase angle of $\phi = 45$ degrees. The three panels assume different reflectance models. *Left panel:* uniform illumination; *Middle panel:* Lambert reflectance; *Right panel:* Lommel–Seeliger reflectance. The color maps differ slightly between the different models; all three scatter approximately equal amounts of light toward the viewer.

where $\mathcal{A}_{\text{pix}}$ is the magnification of the pixel, κ is a constant close to unity that represents the "strength" of the caustic, Δx is the width of a single pixel, and x_k is the distance of the kth pixel from the caustic. In equations (3.1), distances are measured in units of the Einstein radius of the lens system: $\theta_E = \sqrt{2R_{\text{Sch}}/D}$, where $R_{\text{Sch}} = 2GM/c^2$ is the Schwarzschild radius of the lens, and $D \equiv D_{\text{os}}D_{\text{ol}}/D_{\text{ls}}$ (D_{os}, D_{ol}, and D_{ls} are the distances between the observer and source, the observer and lens, and the lens and source, respectively). In our model, we use equation (3.1) to compute the brightness of each pixel across the face of the planet. We then sum the contributions from all of the 401×401 pixels to determine the total brightness at a given position x (corresponding to a given time during the lensing event).

3.2.3   Preview of Results

Using the model described above, we study a number of different scenarios. In all cases, we assume that the source star is a clone of HD209458, a G0V 8 kpc away with a companion planet that has the properties of HD209458b, i.e. a "Hot Jupiter," with radius $R_p = 1.35$ times the radius of Jupiter, and with orbital radius $a = 0.046$ AU (for details of the discovery of HD209458b, see Henry et al. (2000); Charbonneau et al. (2000)). In order to reproduce the published results of GCH03, we assume generous viewing conditions with albedo $A = 1$ (all incident light is reflected). To model a more realistic situation, we examine several other albedo models, including Jupiter's albedo and a "gray atmosphere" – a constant, wavelength–independent albedo. The lens stars are assumed to be typical bulge stars, 6 kpc away, each with

a mass of $0.3 M_\odot$.

We first consider a simple estimate of the flux from such a system. The un–magnified flux from a solar-type star 8 kpc away is quite low; $F_* = L_\odot/4\pi(8\text{kpc})^2 \approx 5 \times 10^{-13}$ erg cm^{-2} s^{-1}. Since a typical photon (~ 500 nm) from such a star carries about 4×10^{-12} ergs, this flux corresponds to a photon number flux of approximately 0.1photons cm^{-2} s^{-1}. Even a large, close–in planet, such as the one under consideration, subtends a small solid angle from the star's perspective; so even with an albedo of $A = 1$, the flux of photons from the planet is reduced by a factor of $\gtrsim 10^4$ from the stellar flux. As a result, the flux from the planet is $\approx 6 \times 10^{-6}$ photons cm^{-2} s^{-1}. This is the well–known reason why gravitational microlensing is essential for detecting the reflected light of a planet around such a distant star.

Crossing a fold-caustic can lead to impressive magnification. In the situation under consideration, the Einstein radius of the lens is approximately 4000 times the radius of the planet, which means that, according to equation (3.1), a magnification factor of $\mathcal{A} \sim \sqrt{4000} \approx 60$ can be achieved.[2] As a result, as shown in the lightcurves below, the planet can perturb the total flux by as much as $\sim 1\%$.

Blending of background starlight in crowded fields makes detecting microlensing events more difficult. For details, see GG00 and GCH03. We follow GCH03 and ignore blending, for with good seeing its effect is negligible.

Note that the order in which the star and planet cross the caustic matters a

---

[2]Our calculations agree with those of Kayser & Witt (1989), and indicate that the maximum effective magnification of a uniform disk is $\sim 1.4/\sqrt{\rho}$, where ρ is the disk radius in units of the Einstein radius of the lens. The maximum effective magnification is slightly greater for Lambert and for Lommel-Seeliger scattering. Thus, an effective magnification factor of $\sim 1.5 \times \sqrt{4000} \approx 100$ can be achieved.

great deal for the detectability of the planet. There are two basic ways in which the planet-star system can be configured as it crosses the caustic region – planet leading star or planet trailing star. Since a planet will almost surely not be detected on ingress, as the system enters the ICR, the favorable configuration for detecting a planet is the planet trailing the star on egress, so that the star is not magnified as the planet crosses the caustic. If the planet is leading the star on egress, the configuration is much less favorable for detecting the planet.

Finally, consider a future microlensing survey that uses a telescope large enough to discern the ingress signature of a planet crossing the caustic. Then, in a fraction of star-planet systems that cross fold-caustics, the system could be in the favorable configuration for both caustic–crossings (i.e., with the star outside the ICR when the planet crosses the caustic). This is because the ICR–crossing–time ($\sim 3-4$ days) is comparable to the semi-orbital period of a close-in extrasolar planet. For example, consider a planet with orbital period ~ 6 days, twice the ICR–crossing time of ~ 3 days. In this case, there should be a $\sim 50\%$ chance that the planet will be in the planet-leading configuration on ingress and in the planet-trailing configuration on egress after having traversed half an orbit (and an equal chance of being unfavorably oriented both in ingress and egress); and so a planet *could* be detected on both ingress and egress. Clearly, the actual likelihood of catching the same planet on both ingress and egress crossings depends on the poorly known distribution of orbital radii for both the planets and for the binary lenses, but it is unlikely that the probability is negligibly small. While the coincidence between the

orbital and intra-caustic-region-crossing timescales is interesting, we note that, in practice, a planet is unlikely to be detected on ingress – unless a deep future survey is devoted to blind monitoring of stars for lensing at ~ hour time–resolution.

3.3   Detectability

Detecting the presence of a planet is, of course, challenging, since even when the planet is on the caustic, its flux is a small fraction ($\lesssim$ 1%) of even the un–magnified flux from the star. As an example, in the inset in Figure 3.3, we present a model R–band light–curve for a 10m telescope, showing first the star, and then the planet exiting the ICR in the favorable configuration for detecting the planet. The planet is modeled with Jupiter's albedo (described in § 3.4 below), corresponding to $A \approx 0.45$. The solid dots show simulated data points. The broad peak between 0–2 hrs results from the star crossing the caustic. The three large dots at 3.0–3.3 hrs correspond to the planet crossing the caustic. On this scale, the magnified planetary flux is invisible against the un–magnified star-flux. Nevertheless, we next show that, as we suggest in § 3.2.3 above, with the current generation of 10m telescopes, it is possible to detect a planet when the star is outside the ICR (but not when the star is inside the ICR).

In Figure 3.3, the top panel shows the tail of the light–curve (after the star has exited the ICR) for the planet's caustic–crossing egress at $\phi = 10°$ (i.e. a zoom–in version of the planet signal from the inset). The bottom panel in this figure shows a random realization of the flux from the planet for the planet's caustic–

crossing egress at $\phi = -45°$ (i.e. in the unfavorable orientation). In both panels, we show error–bars corresponding to the $\sqrt{N}$ shot noise from the total photon–flux. (We ignore instrumental noise because shot noise will dominate for bright bulge stars.) We sum the signal and the $\sqrt{N}$ photon noise over five 8–minute integrations around the planetary caustic crossing. In the favorable orientation (the top panel), we find that the planet is detectable with a total $S/N \sim 15.3$ while in the unfavorable orientation, the planet is essentially undetectable ($S/N \sim 3.6$).

The relationship between the signal-to-noise of detection and the phase angle is summarized in Figure 3.4 below. The planet–flux/star–flux ratio is maximized when the planet is in the "full moon" phase ($\phi \sim 0°$). When the star is outside the ICR, therefore, the planet's detectability is maximized for low phase angles. Phase $\phi \approx 10°$ is optimal (the star intersects a fraction of the planet's surface for $\phi \lesssim 8°$, leading to a rapid decrease in the S/N ratio for still smaller phase angles). When the star is inside the ICR, however, there is a more delicate balance. Since magnification in the ICR decreases with distance from the caustic, the planet's detectability is improved when the the projected impact parameter b is large, which happens for $\phi \sim \pm 90°$. These two competing factors (planet–flux and star–flux) balance to maximize the planet/star flux ratio at about $|\phi| \sim 45°$.

The reader can estimate from Figure 3.4 what fraction of the orbit will yield acceptable signal–to–noise. The S/N of detection exceeds 5 for approximately 90° of phase, or 1/4 of the orbit. Since we are inquiring what will be possible in good viewing conditions, we hereafter will consider only the favorable orientation

($\phi = 10°$).

Note that, for most phases, the S/N of detection is decreased if the plane of the planet's orbit is inclined less than 90°, but is unaffected if the plane of the orbit is not normal to the caustic.

3.4   Planetary Reflection Spectra

HD209458 is one of a relatively small number of stars confirmed to have a transiting companion (HD209458b). By carefully comparing the spectrum of this star during a planetary transit against its spectrum outside transit, Charbonneau et al. (2002) measured how the opacity of the transiting planet's atmosphere varies with wavelength, and inferred the presence of sodium in the atmosphere. Performing a similar analysis, Vidal-Madjar et al. (2003) claim to find an extended hydrogen Lyα-emitting envelope surrounding the planet.

We here investigate the prospects of analogously observing, instead of a transmission spectrum during a transit event, a reflection spectrum during a caustic–crossing event. Although near–future ground–based coronographs (such as Lyot[3]) and more distant future space projects, such as the Terrestrial Planet Finder[4], will be able to probe the spectra of planets around nearby stars using coronographs or nulling interferometers (e.g., Kuchner & Traub (2002)), we are not aware of any other ways at this time to study the reflection spectra of extrasolar planets.

[3]http://lyot.org
[4]http://planetquest.jpl.nasa.gov/TPF/tpf_index.html
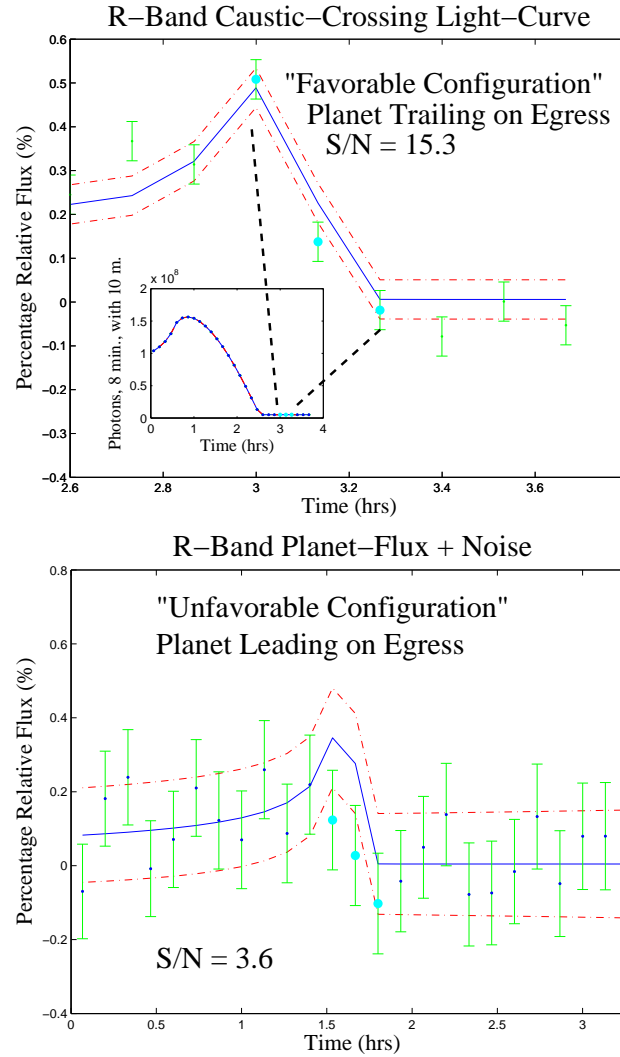
Figure 3.3 Planetary lensing light–curves on egress in the *R*-band, assuming Lommel–Seeliger reflection off a planet similar to HD20y9458b and with Jupiter's albedo, expressed as percentage change in the total flux caused by the presence of the planet. Solid (blue) curves show theoretical light–curves; dashed (red) curves show 1-σ errors; solid (blue) dots show random mock data (theoretical light–curve plus noise) with 1-σ error bars; large (cyan) dots denote times when the planet's surface intersects the caustic. *Top panel:* Favorable (planet–trailing) orientation for detection of the planet on egress. The inset shows the entire caustic-crossing lightcurve for both the star and the planet, with the ordinate showing total photons collected in 8 minute observations with a 10 m. telescope; the big plot is a zoom-in of the tail end of the inset. Dashed lines indicate the relationship between the inset and the parent image. An optimal phase angle of $\phi = 10°$ was assumed. *Bottom Panel:* Unfavorable (planet–leading) orientation on egress, with the optimal value of $\phi = -45°$.
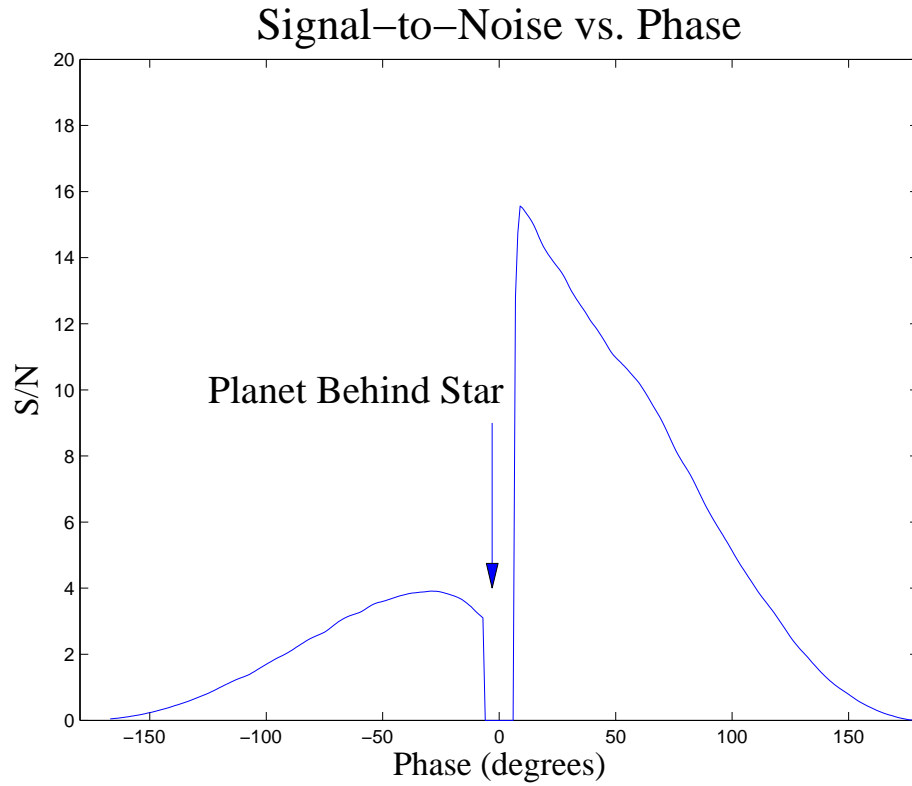
Figure 3.4 This figure shows the dependence of *S/N* on phase angle for *R*-band detection of a planet in a clone of the HD209458 system at 8 kpc, lensed by binary 0.3 $M_\odot$ stars at 6 kpc, in 8-minute integrations with a 10 m telescope. The plane of the planet's orbit is assumed to be inclined 90° to the line-of-sight (edge–on) and normal to the fold-caustic at the point of crossing. Typical proper-motion of lens and source (e.g. GCH03) is assumed. Since the plane of orbit is at inclination 90°, the planet disappears behind the star for phase angles $|\phi| \lesssim 8°$ and detectability drops to zero.

We emphasize that the technique we present in this paper is not in competition with current coronographic work, but is rather in several senses complementary (for more information on coronographic techniques, see, e.g., Oppenheimer et al. (2001)). First, current and near–future coronographic studies are not sensitive to close–in planets, because these planets lack sufficient angular separation from their host–stars, but these are precisely the planets that are most readily seen in the source–plane of a microlensing event. Second, the population of planets available to microlensing is in the Galactic bulge, or at a distance $\sim$ 8 kpc, and is therefore complementary to the nearby population of planets that will eventually be available to the other methods just mentioned.

As a first step toward modeling the reflection spectrum of an extrasolar planet around a solar-type star, we adopt the reflection spectra of the Jovian planets in our solar system, because these are the only gas-giant planets whose wavelength-dependent albedos have ever been measured. Atmospheric conditions, and hence reflection spectra, of hot Jupiters (extrasolar giant planets with short orbital periods) are likely to be much different from those of Jupiter, Saturn, Uranus, and Neptune (for detailed discussions of hot Jupiter atmosphere models, see, e.g., Sudarsky et al. (2003, hereafter SBH03); Burrows et al. (2004); and Seager et al. (2000)). However, given the uncertainty and differences among published atmospheric models of extrasolar giant planets, we prefer to base our calculations on the unambiguously measured albedos of the solar–system gas giants. We will then discuss (at the end of § 3.6 below) the expected differences for the hot–Jupiter atmospheres, and identify

features in the theoretical spectra that could be detected at a similar significance.

To obtain our desired reflection spectra, we need the spectrum of a G2V star, and the albedos of the gas giants in our solar system (with albedo defined as the ratio of reflected flux to incident flux). We obtained an incomplete G2V spectrum from Greg Bothun's webpage[5], that had data missing at wavelengths of strong atmospheric absorption. Regions of missing data up to 1050 nm were filled in with a best-fit $T = 6000$ K blackbody spectrum, and the spectrum was normalized to a peak value of unity for clarity of presentation (see Fig. 3.5). Planetary albedos are taken from Karkoschka (1994), interpolated on a cubic spline (every 5nm) to the G2V reference wavelengths, and are shown for the four gas giants in Figure 3.6. Reflection spectra (in arbitrary units), then, are just the product of the albedo and the solar spectrum (shown by the bottom curve in Fig. 3.5).

While a reflection spectrum of an extrasolar planet with a high signal-to-noise ratio, covering a full range of wavelengths from the visible into the near infrared (NIR), would be ideal (see discussion in § 3.5 below), certain bands of the visible and NIR spectrum provide more information about chemical composition than others. By comparing the reflected flux from wavelength ranges where Jupiter's albedo is low with reflected flux from comparable wavelength ranges where Jupiter's albedo is much higher, we can infer the presence of those compounds responsible for the low albedo. It is clear from the bottom (blue) curve in Figure 3.5 that in a narrow band around 900nm (880nm-905nm) and in a slightly wider band around 1000nm (980nm-1030nm), Jupiter's albedo is quite low ($\sim$ 0.05 and $\sim$ 0.1,

---

[5] *http://zebu.uoregon.edu/spectrar.html*
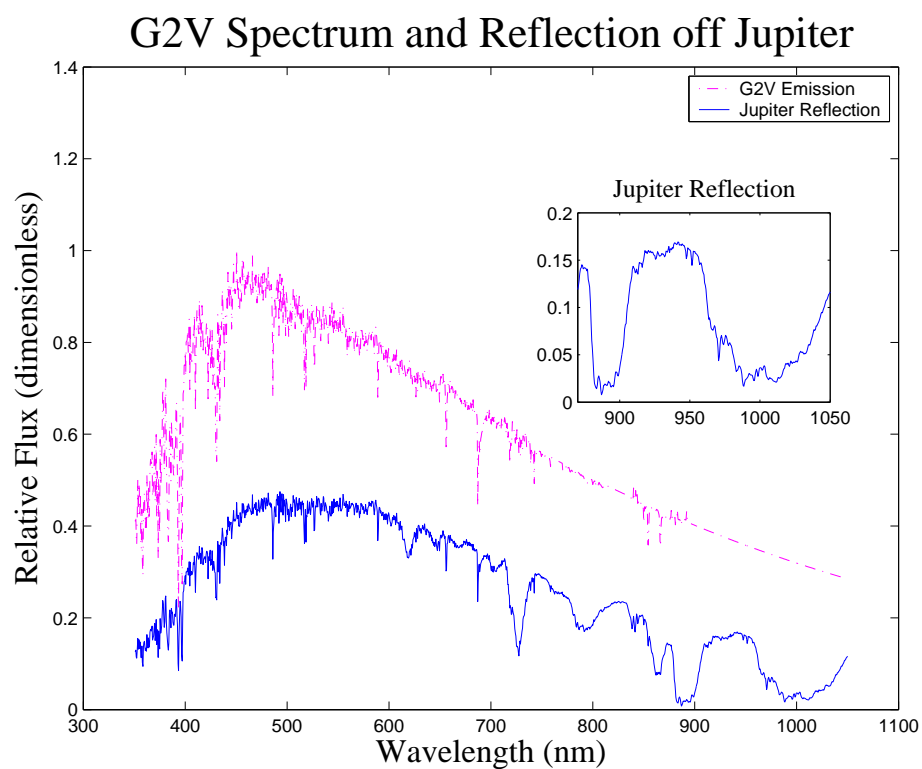
Figure 3.5 Spectrum of a G2V star (top, dashed-dotted magenta curve) and reflection spectrum of a G2V off a planet with Jupiter's albedo (bottom, blue curve). Inset: zoom–in of Jupiter's reflection spectrum over the wavelength range 870–1050nm.
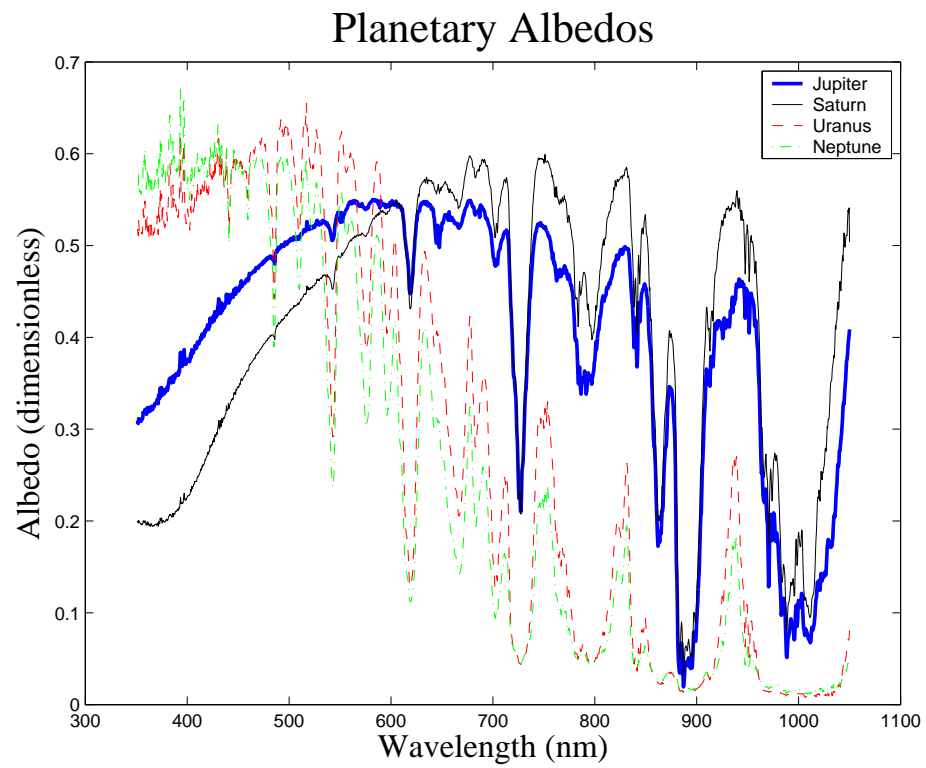
Figure 3.6 Albedos of Jupiter (thick blue curve), Saturn (thin black curve), Uranus (dashed red), and Neptune (dashed-dotted green), adopted from Karkoschka (1994).

respectively); while in–between (920nm-950nm), its albedo is much higher (∼ 0.45). These troughs are caused by absorption by methane in the Jovian atmosphere (Karkoschka 1994). This stark contrast in albedo between adjacent wavelength bands suggests a way to search for, e.g., methane (or other elements or compounds that are expected, in theoretical models for hot Jupiters, to cause features with a similar equivalent width; see discussion below) in the atmosphere of an extrasolar planet.

We note that extrasolar giant planets can orbit very close to their host star (0.05 AU or less), but thermal emission from a planet would nevertheless contribute negligibly to the reflected flux at wavelengths ∼ 1μm even for a hot planet (∼ 1500K).

3.5   Modeling Spectra During Caustic–Crossing

If the (unlensed) flux from the star has spectrum $F_*(\lambda)$ and the planet has wavelength-dependent albedo $A(\lambda)$, then the flux from the planet may be written

$$F_p(\lambda, t) = F_*(\lambda)A(\lambda)f(t), \tag{3.2}$$

where the multiplicative function $f(t)$ depends on various geometrical factors (the solid angle that the planet subtends from the perspective of the star, whether the planet has moons or rings, how far the planet is from the caustic, etc), and also on the reflectance model. The total observed flux, therefore, can be written (in the

favorable orientation, with the star un–magnified, as discussed above) as

$$F_T(\lambda, t) = F_*(\lambda) + F_p(\lambda, t)$$

$$= F_*(\lambda)(1 + A(\lambda)f(t)). \tag{3.3}$$

The observables are F_T and F_*. The physically interesting characteristics of the planetary system, however, are $A(\lambda)$ and $f(t)$, and these may be solved for as

$$A(\lambda)f(t) = \frac{F_T(\lambda, t)}{F_*(\lambda)} - 1 \equiv G(\lambda, t), \tag{3.4}$$

where we define the function G as the observable quantity constructed on the right–hand side of equation (3.4). With perfect data, the time-difference between the star's and the planet's caustic crossings breaks the apparent degeneracy between A and f in the general solution,

$$A(\lambda) = k_1 \exp\left[\int \frac{\partial G/\partial \lambda}{G} d\lambda\right] \tag{3.5}$$

$$f(t) = k_2 \exp\left[\int \frac{\partial G/\partial t}{G} dt\right], \tag{3.6}$$

(where k_1 and k_2 are constants of integration such that $Af = G$).

In practice, with data as noisy as can be expected with the current generation of telescopes, it is impossible to separate A from f, and A may be determined only given a model for f. Still, it is possible in principle to posit a model for f (as outlined in § 3.2 above) and then to solve for $A(\lambda)$. In this case, since the signal–to–noise

ratio for the detection of the planet we find is only ~ 15.3, it is still only possible to split the light into a few broad spectral bands, rather than into a resolved spectrum.

In order to test the idea that we could look for the spectral signature of a particular compound in the reflected light from a distant extrasolar planet, we model a planet with Jupiter's reflection-spectrum and scrutinize the model data for evidence of methane. In order to maximize signal–to–noise, we assume an egress caustic–crossing with planetary phase $\phi = 10°$.

To search for signatures of methane, we construct a mock "methane band filter" (hereafter "MBF"), that allows complete transmission from 880nm-905nm *and* from 980nm-1030nm (the bands where Jupiter's albedo is low because of methane, as discussed in §3.4) and zero transmission elsewhere (MBF is therefore a "double top-hat" filter). Note that we do not necessarily mean a physical filter here; we effectively assume that the flux in a low-resolution spectrum can be binned and computed in these wavelength ranges. A more realistic analysis would have to take into account the additional instrumental noise in any physical implementation of such a filter (such as read–out noise in the case of a spectrograph). We then compare the MBF light–curve of a model planet with Jupiter's albedo to the MBF light–curve of a model planet with the methane feature removed – i.e., a model planet where the albedo is replaced by a constant equal to Jupiter's mean albedo $\bar{A} = 0.45$. Figure 3.7 shows this comparison: the top panel shows the MBF light–curve for a planet with Jupiter's albedo (here, the planet is detected at $S/N = 1.8$, which counts as a non-detection); the bottom panel shows the MBF light–curve for

a planet with constant albedo $A = 0.45$ (here, the planet is detected at $S/N = 8.4$).

In practice, the observational strategy would involve employing a "high albedo filter" (hereafter "HAF") that uses a region of the spectrum that is relatively unaffected by methane and that is comparable in width to the MBF filter (e.g. the adjacent 920nm-950nm region, and/or other regions where Jupiter's albedo is high). The flux measured through the HAF filter would then be used to predict the expected MBF flux according to the no–methane null–hypothesis. In practice, then, a *non*–detection of the planet in the MBF band together with a simultaneous *detection* in the HAF band, would be evidence for the presence of methane in the atmosphere of the planet. The S/N is proportional to the square root of the number of photons collected, and to the diameter of the telescope. With a future 30 or 100m telescope, therefore, it would be possible to achieve a S/N of ~ 25 or ~ 80, respectively, in detecting the presence of methane.

3.6 Discussion

Note that, strictly speaking, our S/N calculations are for a space-based observatory, because we do not include sky brightness. Detailed data on sky brightness are available in Leinert et al. (1998). In R-band, the contribution to total flux from the sky is small for good seeing (for seeing $\sim 0.75''$, the star is more than an order of magnitude brighter than the sky within the aperture subtended by the star). At 900nm, the star is still several times brighter than the sky for good seeing conditions, but by 1μm the sky is comparably bright to the star, which would increase the noise
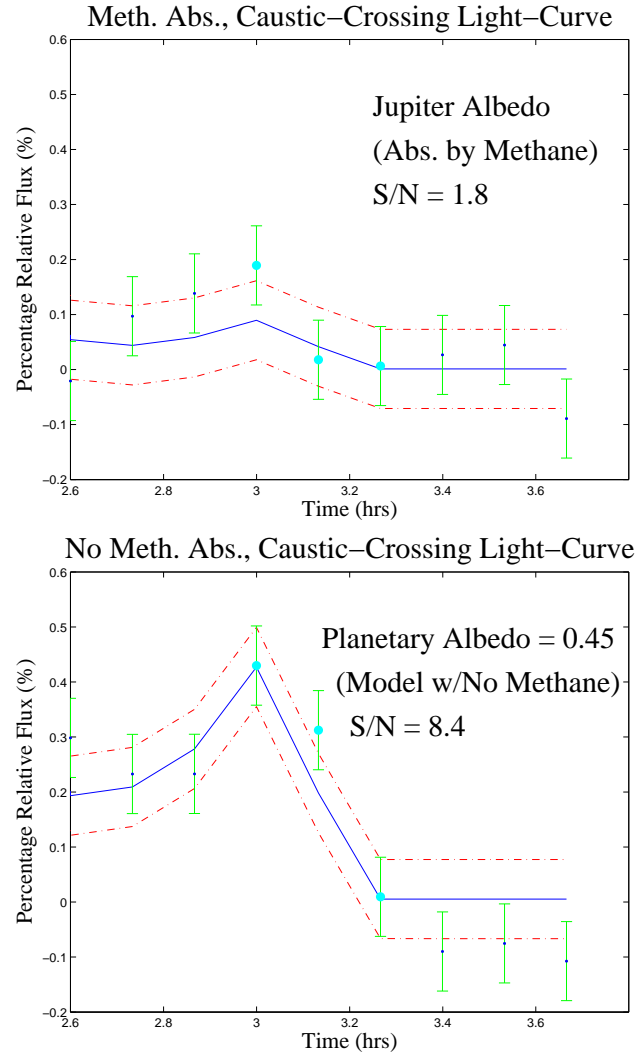
Figure 3.7 Egress light–curves in the MBF band, which covers two deep methane absorption features and includes light from 880nm-905nm and from 980nm-1030nm. The meaning of the symbols are as in Figure 3.3. *Top panel:* Jupiter's low albedo is adopted, which leads to a non–detection of the planet. *Bottom panel:* A constant albedo of $A = 0.45$ is used showing what the light–curve would look like if there were no methane present ($S/N = 8.4$). This plot is quite similar to the light–curve that would be obtained through a filter in a band where there is low methane absorption and Jupiter's albedo is much higher ($\sim 0.4 - 0.5$).

in an observation by a factor of $\sim \sqrt{2}$ and would therefore decrease the S/N by the same factor. For a 10m telescope, this still indicates $S/N \sim 6$ for good seeing conditions in the situation modeled above. Note that if the plane of the planet's orbit is inclined less than 90°, then for nearly all phases the S/N for detection of the planet is reduced, for any spectral filter. The ratio of the flux from the planet through the MBF to that through the HAF, however, is independent of both inclination and phase.

Although some models of close–in extrasolar giant planets predict significantly less methane than is present in Jupiter's visible cloud-layer, this prediction is not universal. Seager et al. (2000), for example, present a model of 51 Peg b that has spectroscopically significant methane–levels. They point out, however, that these methane features might be present only in the coolest, least absorptive models.

Even if future models should converge upon the conclusion that the gaseous methane content of hot Jupiters is very low, there are spectral features due to other chemicals that are predicted to be present in the atmospheres of close–in extrasolar giant planets, that are predicted to be comparably strong to Jupiter's methane features. A search for these other predicted features would be analogous to the methods described above. SBH03 identify five classes of extrasolar giant planets, ranging from class I (Jupiter–like) through classes IV ($a \sim 0.1\text{AU}$) and V ($a \sim 0.05$). A caveat introduced by SBH03 is that, for class IV and V planets, the planet's spectrum redward of $\sim 500\text{nm}$ includes increasing levels of thermal emission, and so it makes more sense to discuss the "emergent spectrum" rather than the reflection

spectrum. Their model emergent spectra for class IV planets include several strong absorption features. In the visible, sodium ($\sim$ 600nm) and potassium ($\sim$ 800nm) are predicted to induce absorption features with a comparable equivalent width to the methane features we consider above; in the NIR ($\sim 1.4\mu$m), water, which is thought to condense too deep in Jupiter to affect the cloud-top albedo, is predicted to cause an even deeper (factor of $\sim$ 100) trough in the emergent spectrum. This water feature is at a wavelength where the Earth's sky is fairly bright (an order of magnitude or more brighter than the star), which would make it difficult to discern from ground-based observations but which should pose no difficulty for a space-based telescope. Table 3.1 summarizes the strengths of the three absorption features predicted in SBH03 class IV planets described above and compares them to the methane features previously considered. If a planet is detected with high S/N in a HAF but is (un)detected in a filter centered on a spectral feature with S/N much less than the number quoted in the last column of Table 3.1, this would be evidence for the presence of the chemical responsible for the feature.

The right-hand side of Figure 3.4 is summarized and generalized in Equation 3.7 below, which gives a rough estimation of the expected signal–to–noise for detection of a planet whose orbit is at inclination 90°, in which the source is a Main-Sequence star. The lensing configuration is taken to be the one described above. D is the diameter of the telescope's aperture, EW_f is the equivalent width of

the spectral filter used, and EW_l is the equivalent width of a spectral line or feature.

$$\frac{S}{N} \sim \left(16 - \frac{\phi}{9^\circ}\right) \times \Theta(\bar{A}, D, EW_f, EW_l) \times \Psi(M_*, R_p, a) \tag{3.7}$$

where Θ and Ψ are the functions given below:

$$\Theta(\bar{A}, D, EW_f, EW_l) = \frac{\bar{A}}{0.45} \times \frac{D}{10 \text{ m}} \sqrt{\frac{EW_f}{150\text{nm}}} \left(1 - \frac{EW_l}{EW_f}\right) \tag{3.8}$$

$$\Psi(M_*, R_p, a) = 10^{-3} \left(\frac{2M_*}{M_\odot} - 0.8\right) \left(\frac{R_p}{R_J}\right)^{3/2} \left(\frac{a}{1 \text{ AU}}\right)^{-2}. \tag{3.9}$$

Observe that both Θ and Ψ are unity in the case of an R-band observation through a 10 m telescope of the planetary system considered above.

Equation 3.7 slightly over-predicts S/N for filters on the red side of R-band and slightly under-predicts S/N for filters on the blue side; furthermore, as above, it does not include sky-brightness, which is particularly important redward of 1μm.

3.7 Conclusions

In the Galactic bulge there is a large number of stars and, presumably, a comparably large number of planets. With current and future microlensing surveys in the direction of the bulge, we expect that some solar systems will cross the fold-caustics of binary lenses. Unfortunately, although in such events there will be two caustic–crossings, it appears that current technology will only allow for detection of a planet orbiting the source-star during the egress caustic–crossing – and furthermore only

when the star-planet system is in the favorable configuration. Still, with its expected mean albedo, the planet should reflect enough light that, in the case we consider, it should be detectable for roughly 1/4 of its orbit with a 10m telescope.

Our results suggests that the strategy outlined by GG00 and Lewis & Ibata should be viable: each time a bulge star is seen to cross a fold-caustic into the ICR, the egress event should be closely monitored in order to detect a planet in the trailing (favorable) configuration, should such a favorable orientation occur. If 10% of bulge stars have hot Jupiter companions, then, since a quarter of planet-star systems will be have appropriate in the planet-trailing configuration on egress, $\sim 2 - 3\%$ of bulge stars that cross fold-caustics will be seen, under close monitoring (in ~ 8–min integrations) during the egress crossing, to have planetary companions.

If such planets are detected, it will be possible, in principle, to determine various properties of the planet, including physical (reflectance model, phase, angular orientation relative to the caustic, presence of moons or rings; see GCH03) and chemical characteristics (the presence of specific constituent compounds of the atmosphere, as suggested by our results). Since the expected perturbations to an observed light–curve from the physical characteristics are either small (moons, rings, angular orientation) or degenerate with other effects on the total brightness of the planet, such as the planet's albedo or the solid angle it subtends from the perspective of its star (reflectance model, phase), it will be difficult in practice to determine these physical characteristics. For example, if we were to observe the egress caustic-crossing light–curve of a planet in the bulge that has rings around

it, is at illumination phase $\phi = 45°$, and obeys Lommel-Seeliger reflection, we would most likely not be able to infer the presence of the rings, the phase, or the nonuniform reflectance because the data could be fit equally well (within error bars) by a best-fit $\phi = 0°$ model with no rings (at $\phi = 0°$, a Lommel-Seeliger planet is uniformly illuminated). The expected perturbations from some atmospheric compounds, however, are much greater (a factor of ~ 5 or more) and do not suffer from analogous geometrical degeneracies.

In our example, using 8–minute observations on a 10m telescope, we found that the presence of methane could be inferred from a non–detection of the planet in the strong broad methane absorption band at $\approx 0.9\mu$m, accompanied by a $S/N \sim 10$ detection in adjacent bands. Observations such as the ones described in this paper will provide a crucial constraint on models of roaster atmospheres. Then, in turn, as more accurate atmosphere models become available, this S/N could improved by fitting the data to spectral templates with free parameters corresponding to variable compositions. Future generations of large telescopes might therefore be able to probe the composition of the atmospheres of distant extrasolar planets.

APPENDIX

In this Appendix, we present the specifics of our model of the planet, including details regarding the three reflectance models we consider.

The viewing geometry of a planet–star binary can be described in a three–dimensional coordinate system, centered on the planet, as illustrated in Figure 3.8.

Table 3.1   This table shows the line-center and equivalent width for each of 5 spectral features. Three of these features, the Sodium (Na), Potassium (K), and Water features, are expected in planets classified by SBH03 as class IV ($a \sim 0.1\text{AU}$). The other two features are the methane features considered in detail in this paper, with data taken from Jupiter's reflection spectrum, available in Karkoschka (1994). The last column shows the predicted S/N ratio given the absence of the chemical, and for Na and K it is computed from Equation 3.7 (using $\bar{A} = 0.35$ and $EW_l = 0$), while for the Methane features it is computed from our simulations. This ratio is not applicable to the case of water, because the sky is too bright for ground-based observations at this wavelength.

Spectral Feature	SBH03 Class	Line Center	Eq. Width	S/N Ratio
Sodium	IV	$\sim 600\text{nm}$	$\sim 80\text{nm}$	14
Potassium	IV	$\sim 780\text{nm}$	$\sim 20\text{nm}$	7
Methane	I	$\sim 990\text{nm}$	$\sim 20\text{nm}$	5
Methane	I	$\sim 1.00\mu\text{m}$	$\sim 40\text{nm}$	7
Water	IV	$\sim 1.40\mu\text{m}$	$\sim 200\text{nm}$	N/A
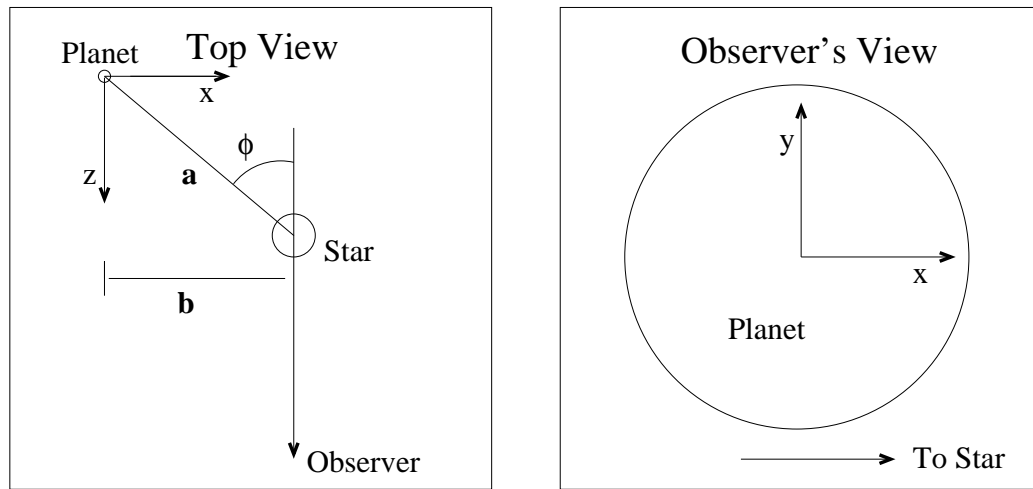
Figure 3.8 Viewing Geometry. *Left panel:* Top view ("God's eye view") of the viewing geometry, showing the phase angle (ϕ), the orbital radius (a), and the projected impact parameter ($b = a\sin\phi$). *Right panel:* Observer's view of the planet.

The z–axis is defined to point toward the observer; the x–axis is in the direction from the planet to the star, as projected on the sky from the perspective of the observer; and the y–axis is defined by the x and z-axes and the usual right–hand rule. The phase is as defined in Section 3.2.1 given by the angle.

The reflectance models considered are the following: *Uniform reflection:* the planet has uniform surface–brightness as seen from the observer, regardless of phase. *Lambert reflection:* the surface–brightness of a patch of projected surface is proportional to the cosine of the angle between the incident radiation and the surface–normal vector: $B \propto \cos(\theta_{\text{ill.}})$. Let $\hat{x}$, $\hat{y}$, and $\hat{z}$ be the dimensionless coordinates on the planetary surface ($\hat{x} = x/R_p$; $\hat{y} = y/R_p$, and $\hat{z} = z/R_p$, where R_p is the planet's radius). The unit vector to the star is $\mathbf{s} = (\sin\phi, 0, \cos\phi)$, and the unit surface-normal vector is $\mathbf{N} = (\hat{x}, \hat{y}, \hat{z}) = (\hat{x}, \hat{y}, \sqrt{1 - (\hat{x}^2 + \hat{y}^2)})$, so the desired cosine is given by $\mathbf{s} \cdot \mathbf{N} = \hat{x}\sin\phi + \sqrt{1 - \hat{x}^2 - \hat{y}^2}\cos\phi$. A failing of the Lambert reflectance model is that in the "full-moon" phase, the specific intensity from the edge of the projected disk drops to zero, in conflict with the appearance of the Moon and other planets in our Solar system. *Lommel–Seeliger reflection* is a phenomenological model, designed to reproduce the reflectance of the Moon, that also mimics well the appearances of a number of other bodies in the Solar System. Neither the Lambert nor the Lommel–Seeliger model – and certainly not the uniform model – can capture in detail the appearance of a patch of planetary or Lunar surface at high resolution; but the Lommel-Seeliger model in particular is successful at reproducing at low resolution the whole planetary disk. The surface–brightness of a patch

of projected surface in the Lommel-Seeliger model is proportional to the cosine of the illumination angle, and inversely proportional to the sum of the cosines of the illumination angle and the viewing angle: $B \propto \cos(\theta_{\text{ill.}})/[\cos(\theta_{\text{ill.}}) + \cos(\theta_{\text{view}})]$. The cosine of the viewing angle is the dot product of $\mathbf{N}$ with the unit vector to the observer, $\mathbf{z} = (0, 0, 1)$, or $\mathbf{N} \cdot \mathbf{z} = \sqrt{1 - (\hat{x}^2 + \hat{y}^2)}$.

Within our coordinate system, the un–magnified flux from a patch of surface at projected coordinates $(\hat{x}, \hat{y})$ can be represented as follows

$$dF = KP\frac{B(\hat{x}, \hat{y})}{4\pi r^2}d\hat{x}\, d\hat{y}. \tag{3.10}$$

Here P is the total incident stellar power that the planet reflects, or $L_*A(\pi R_p^2)/4\pi a^2$, where L_* is the luminosity of the star; A and a are the planet's albedo and orbital semi-major axis, respectively; $B(\hat{x}, \hat{y})$ gives the spatial dependence of the apparent brightness of the planet, and depends on the reflectance model; r is the distance of the observer from the system; and K is an overall scale-factor so that the total reflected light equals the total intercepted light times the albedo A. What remain to be given, then, are K and B for each reflectance model.

The resulting constants and formulae are:

$$K_U = 2/\pi$$

$$K_L = 4/\pi$$

$$K_{LS} = 1.556, \tag{3.11}$$

and

$$B_U(\hat{x}, \hat{y}) = 1$$

$$B_L(\hat{x}, \hat{y}) = \hat{x}\sin\phi + \sqrt{1 - (\hat{x}^2 + \hat{y}^2)}\cos\phi$$

$$B_{LS}(\hat{x}, \hat{y}) = \frac{\hat{x}\sin\phi + \sqrt{1 - (\hat{x}^2 + \hat{y}^2)}\cos\phi}{\hat{x}\sin\phi + \sqrt{1 - (\hat{x}^2 + \hat{y}^2)}(\cos\phi + 1)}, \tag{3.12}$$

where the U, L, and LS subscripts refer to uniform, Lambert, and Lommel-Seeliger

reflectance, respectively.

Chapter 4

On Constraining A Transiting Exoplanet's Rotation Rate With Its Transit Spectrum

4.1 Introduction

Since the early 1990s, more than 280 planets have been discovered orbiting other stars. Roughly four fifths of the planets that have been discovered are at least half as massive as Jupiter, and about a quarter of the known planets orbit extremely close to their parent star ($\lesssim 0.1$ AU). The Jupiter–mass planets that are in close orbits are highly irradiated by their stars and are therefore called "Hot Jupiters." Twenty Hot Jupiters transit along the line of sight between Earth and their star (Charbonneau et al. 2000; Henry et al. 2000; Konacki et al. 2003; Bouchy et al. 2004; Pont et al.

2004; Konacki et al. 2004, 2005; Udalski et al. 2002a,b,c, 2003, 2004; Alonso et al. 2004; Bouchy et al. 2005b; McCullough et al. 2006; O'Donovan et al. 2006; Bakos et al. 2006; Collier Cameron et al. 2006; Burke et al. 2007; Gillon et al. 2007; Bakos et al. 2007b,a; Johns-Krull et al. 2007)[1]. Observations of the transiting planets have confirmed their similarity to Jupiter by revealing that these are gas giant planets, with radii comparable to, or somewhat larger than Jupiter's (Charbonneau et al. 2000; Henry et al. 2000; Gaudi 2005).

The effects of the tidal torques experienced by an orbiting body have been studied for a long time – for an early seminal analyis, see Goldreich & Peale (1966). Such torques tend to synchronize a satellite's rotation rate to its orbital rate, and if the torque is sufficient this synchronization is achieved and the orbiter is said to be "tidally locked," as the Earth's Moon is. The Hot Jupiter–class extrasolar planets are thought to orbit sufficiently close to their stars that their tidal locking timescales are much shorter than the ages of the planets. The planets, then, are expected to be tidally locked to the stars, with one hemisphere in permanent day and the other in permanent night (Harrington et al. 2006).

A tidally locked Hot Jupiter will have a permanent sharp contrast in temperature between the substellar point and the night side, which must have a profound influence on the atmospheric dynamics. Showman & Guillot (2002) make simple predictions of the day/night temperature difference (~ 500 K) and the speed of winds (up to ~ 2 km s^{-1}), and their detailed, three–dimensional simulations agree with their estimates. Shallow–water simulations by Cho et al. (2003) predict

---

[1]Corot–Exo–1b is reported at http://www.esa.int/esaCP/SEMCKNU681F_index_0.html.

longitudinally averaged zonal wind speeds of up to 400 m s^{-1}, with local winds approaching 2.7 km s^{-1} (under some assumtions). Simulations by Cooper & Showman (2005) predict a super–rotational jet (i.e., blowing eastward, where north is defined by the right–hand rule) that blows the hottest part of the planet downstream by about 60° from the substellar point. Their simulations predict supersonic winds exceeding 9 km s^{-1} at high latitudes, high in the atmosphere (where the optical depth is low) and winds exceeding 4 km s^{-1} at pressures near the photosphere. A Spitzer Space Telescope phase curve for υ Andromedae b rules out a phase–shift as large as 60° between the substellar point and the hottest spot (Harrington et al. 2006), but a Spitzer phase curve for HD189733b favors a $\sim$ 30° shift for that planet (Knutson et al. 2007a), so it remains unclear to what extent available data indicate very strong photospheric winds.

Transmission spectroscopy is a way to probe the atmospheres of these planets. Charbonneau et al. (2002) were the first to detect an absorption feature in what is probably the atmosphere of HD209458b, when they found that the effective radius of the planet increases slightly at the wavelength of a strong sodium absorption doublet (the sodium D lines) at $\sim$ 590 nm. In addition, Vidal-Madjar et al. (2003, 2004) have reported a number of absorption features in HD209458's transit spectra that are due to various species (including hydrogen Lyman alpha, neutral carbon, oxygen, and sulfur, and some ionization states of carbon, nitrogen, and silicon) in a hot exosphere that is probably only loosely bound to the planet. Intriguingly, through analyzing the red and near–IR portion of HD209458b's transit spectrum

Barman (2007) found a 10σ detection of atmospheric water vapor. Several measurements of the planet's emission spectrum, however, have found results that seem to be inconsistent with high water abundance high in the atmosphere (Grillmair et al. 2007; Richardson et al. 2007; Swain et al. 2007).

Initial work by Seager & Sasselov (2000) and a comprehensive study by Brown (2001, hereafter B01) have described various other considerations that should affect the details of transit spectra, including the orbital motion of a planet (a few tens of kilometers per second in the radial direction), the rotation of the planet (a few kilometers per second at the equator, according to the hypothesis that the planet is tidally locked), and winds on the planet's surface (in B01's analysis, up to ~ 1 km s^{-1}). These physical effects should tend to broaden or impose Doppler shifts on absorption features due to the planet's atmosphere. B01 constructed an impressively detailed model of radiative transfer through a Hot Jupiter's atmosphere, assuming various models of zonal windflow superimposed on an equatorial bulk rotation speed of $v_{\mathrm{eq}} = 2$ km s^{-1}, which is approximately the value for HD209458b under the assumption that it is tidally locked in its 3.5 day orbit. He finds the height of the cloud deck to be the most important parameter that affects the transmission of light through the planet's atmosphere.

The original discovery of the roughly Jupiter–mass planet in a close, ~ 4 day orbit around 51 Pegasi (Mayor & Queloz 1995) prompted interest in the dynamics and structure that must govern a highly insolated gas giant planet (Guillot et al. 1996). Observations of the transiting Hot Jupiters heightened this interest when

they revealed a puzzling feature of these planets: at least several of them are a bit puffier than Jupiter, with diameters ranging from slightly larger than Jupiter's to as much as $\sim 80\%$ larger. It is not clear what allows some planets to maintain such large radii. It has been suggested that if a Jovian planet migrates very quickly, from its presumed formation location at least several AU from its star, to its eventual several day orbit, then it might reach its final home before it has cooled enough to shrink to Jupiter's radius. Accordingly, some authors have investigated the migration processes that lead gas giant planets to such close orbits as have been found (e.g. Trilling et al. 2002). Others have investigated various ways in which a gas giant could either be heated once it ends up near its star, or otherwise maintain sufficient internal energy to sustain its inflated size (Guillot & Showman 2002; Burrows et al. 2003; Laughlin et al. 2005; Bodenheimer et al. 2003; Guillot 2005; Burrows et al. 2007b; Chabrier & Baraffe 2007). Although various physical mechanisms have been suggested as the apparently missing energy source that allows the unexpectedly large radii sometimes seen, the lesson of these investigations *in toto* is that it is not easy to explain the inflated sizes, either in terms of the greater stellar flux that these planets experience by virtue of being so close to their stars, or in terms of their evolutionary migratory histories. A recent paper by Winn & Holman (2005) propose that, contrary to the commonly accepted paradigm, Hot Jupiters might be trapped in a Cassini state with large obliquity, in which the spin–axis precesses in resonance with the orbit, but lies nearly in the orbital plane. Such a state might be stable against perturbation, and yet able to generate sufficient internal energy to

increase a gas giant planet's radius to the observed values. In light of an even more recent analysis by Levrard et al. (2007), however, it appears that the probability of capture into a Cassini state 2 resonance is quite small for a planet with semi–major axis $a < 0.1$ AU. Furthermore, Fabrycky et al. (2007) argue that even if a planet is captured into Cassini state 2, it is likely to remain there for a time that is short relative to the age of the system.

High–resolution transit spectra that have high signal–to–noise ratios will allow us to distinguish between various models of orbit, rotation, and weather, as discussed by B01. Because the orbit is known to high accuracy, and the predictions of the effects of weather (or climate) are highly uncertain, as described above, we will focus in this paper on the much more easily predicted effect of a planet's rotation on a transit–spectrum. If we neglect winds, then the large–obliquity Cassini state described by Winn & Holman (2005) should have a spectral signature that is very similar to that of a non–rotating model. In contrast, the rotation of a tidally locked planet should impose a Doppler distortion on spectral lines arising from the planet's atmosphere that is roughly an overall redshift during ingress, as the planet is just entering the stellar disk, and a similar distortion that is roughly an overall blueshift during egress, as the planet is just exiting the disk. During mid–transit, the spectral distortion is more similar to rotational broadening. In the present investigation, we address whether there is any hope that these spectral distortions from tidally–locked rotation can be observed. In our study, we focus only on the sodium doublet detected by Charbonneau et al. (2002). As we will show below,

the sensitivity of a measurement of rotation scales with the square root of the number of lines under consideration. Model spectra from, e.g., Sudarsky et al. (2003) and Barman (2007) predict a strong potassium doublet at ~ 770 nm, strong water absorption features in the near–infrared, and a handful of near–UV lines. If some of these are confirmed in the atmosphere of a transiting planet, they will provide a modest increase in S/N. Since the sodium lines are expected to be the strongest, however, it seems unlikely that observing multiple lines will yield a boost in S/N by more than a factor of a few.

We emphasize that it may not be at all justified to neglect winds. It is quite likely that there are super–rotational winds on Hot Jupiters, which are probably necessary to heat the "night" side. As indicated above, some models predict, and the observed phase curve for HD189733b suggests, that at the photosphere these winds might be significantly (100% or more) greater than the equatorial rotation rate, and therefore might contribute importantly to the Doppler distortion induced by the motion of the planet's atmosphere. Nevertheless, in order to isolate the contribution of rotation, we do neglect winds in this study. The Doppler distortions that we predict can therefore probably be taken as a lower bound on the distortions that would be observed for a tidally–locked transiting Hot Jupiter.

We find that the spectral shifts induced by rotation will be difficult to detect with current technology, but perhaps not insurmountably so, at least with technology that might be available in the not–to–distant future. The measurements we will describe are limited by a paucity of photons. As such, their signal–to–noise

ratio will be enhanced by a bright star and a puffy planet (i.e., a planet with a large scale–height). HD209458 is at least a magnitude brighter than any other star with a known transiting planet except HD189733, and its planet is larger than HD189733b; so HD209458b should be a better target than any other known transiting planet except possibly HD189733b. In this paper, we model the HD209458b system because it is the best–studied system, and it is unlikely that any currently–known planets would be significantly better targets. In a single transit, observations of HD209458 with a 6 m telescope that has a high–resolution ($> 50,000$) optical spectrograph with good throughput ($\sim 18\%$) could only show the influence of tidally locked rotation at the $\sim 0.2\sigma$ level. With ultrahigh–resolution ($\gtrsim 700,000$) and good throughput ($\sim 4\%$) this effect would still only show up at the $\sim 0.6\sigma$ level. In less than a year, the signal of rotation could be present at five times the noise ($S/N = 5$). Of course, a telescope with larger collecting area, higher spectral resolution, or better throughput would cause the signal to be apparent at that significance level in less time.

Other studies have approached the problem of determining the rotation rate from a different angle. Seager & Hui (2002) and Barnes & Fortney (2003) suggest that an oblate spheroid will have a different transit light curve from a perfect sphere, and so measuring the oblateness from transit photometry will provide a handle on the rotation rate. The oblateness is somewhat degenerate with several other parameters that are not perfectly known, however, so they conclude that it would be difficult to actually determine that rotation rate in this manner. The method

we describe here could eventually prove to be an important complement to other observations to constrain the rotation rate.

In the remainder of this paper, we address this idea in detail. One complication that we discuss below is that the technique of this paper is not immune from several near–degeneracies among the many attributes of transiting extrasolar planets that influence light curves or spectra. Although it is likely that current or near–future instruments will be sensitive enough that the spectral distortion imposed by HD209458b's rotation (if it is tidally locked) is visible, it might still be very challenging to discern the fingerprint of rotation from other attributes that affect the spectra at a similar level. In this paper, we tackle the forward problem of calculating the amount of distortion that is caused by rotation. The inverse problem – determining from observations whether a planet is tidally locked – is more difficult and should be the topic of a future study.

The structure of the rest of this paper is as follows: In § 4.2, we describe qualitatively what happens to the starlight received on Earth when a planet transits its star; we give a rough order of magnitude estimate of the the magnitude and detectability of the spectral distortions caused by tidally locked rotation; and we briefly describe some technological progress and remaining challenges relevant to our task of deducing bulk motions in a planet's atmosphere from transit spectra. In § 4.3, we describe our computational model of a transit spectrum. In § 4.4, we describe the results of our model according to various assumed input parameters. In § 4.5, we discuss the scaling of S/N on various model parameters and we address

the prospects of actually observationally determining whether a transitting planet is tidally locked. In § 4.6, we conclude by describing various ways to boost our predicted S/N to a more optimistic value.

4.2   Overview of the Problem

The practical feasibility of the investigation we undertake depends on a few factors: understanding the various detailed processes that affect the starlight that reaches Earth when a planet transits its star; the magnitude of the distortion that tidally locked rotation induces; and the technology available to measure such distortions. In this section, we give an overview of these three factors – in particular, in § 4.2.2, we give a simple estimate of the results that we will later (in § 4.4) calculate in detail.

4.2.1   Relevant Processes

A planet transiting in front of its star affects the starlight that ultimately reaches Earth in many ways. The motion of the planet's atmosphere (rotation and winds) is a small perturbation on top of several more dominant effects. We therefore summarize below the physical processes that are at least as significant as the effect of tidally locked rotation. Figure 4.1 schematically represents this situation, and captures nearly all of the processes described below: a rotating planet (of exaggerated relative size) transits in front of a rotating star. The figure depicts a snapshot
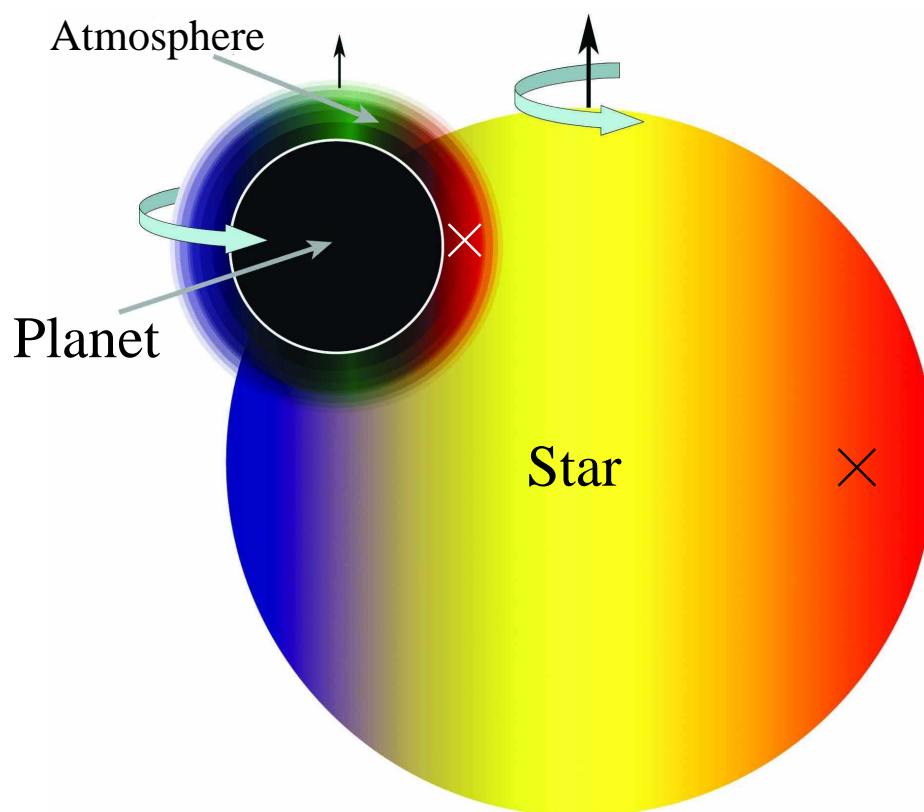
Figure 4.1 Rotating planet beginning to transit in front of rotating star. The vertical black arrows represent the rotation axes of the planet and the star, and the curved arrows indicate the direction of rotation for each. The X's on the right–sides of both the planet and the star indicate regions that are receding from the observer and are therefore redshifted; the unmarked left sides of the planet and the star are moving toward the observer and are therefore blueshifted. The white circle surrounding the opaque black part of the planet denotes the cloud deck, or the boundary between the partially transparent and the fully opaque portions of the planet's disk. The planet is orbiting in the same sense as both it and the star are rotating. The planet is shown above the star's midplane to represent the inclination of the orbit relative to the line–of–sight.

partway through ingress, when half of the planet is in front of the stellar disk[2]. The white circle indicates a hypothetical sharp demarcation between the opaque part of the planet (in black) and the optically thin part, labeled "Atmosphere" (described further below).

1. *Geometric Occultation*:

   The largest effect is an overall dimming by a factor of roughly the ratio of the area of the planet to that of the star: $(R_p/R_*)^2$. Since stars are not perfectly uniform disks, but instead tend to darken toward the limb at most visible wavelengths, the fractional dimming due to being in the planet's shadow tends to be slightly less than the ratio of the areas when the planet is near the edge of the stellar disk and slightly more than this ratio when the planet is near the center.

2. *Stellar Wobble*:

   The primary spectral effect of the planet orbiting the star is the radial velocity wobble induced by the planet's gravity. This periodic spectral shift is of course in effect during the transit, when, for a close–in planet like HD209458b, it has an influence on the order of $\sim \pm 10$ m s^{-1}. This effect is a redshift as the planet begins to transit across the disk (during ingress) and a blueshift during egress.

3. *Rossiter–McLaughlin Effect*:

   A more subtle effect arises because, during the transit, the planet moves across – and therefore blocks – parts of the star that have different recessional

[2]The planet is above the star's equator to represent a slight inclination in its orbit.

velocities. If (as is expected) the planet's orbit is aligned with the star's spin, then during ingress the planet is blocking a part of the star that is slightly blueshifted, and during egress it is blocking a part of the star that is slightly redshifted. Figure 4.1 illustrates the planet blocking some of the bluest parts of the star during ingress.

The parts of the star that are occluded during ingress/egress have spectra that are blue/redshifted by a velocity that is approximately the equatorial rotational speed of the star, or about ∼ 1-2 km s^{-1} for a Sun–like star. As the figure indicates, during ingress/egress, the integrated spectrum of the remaining (unblocked) parts of the star is on average slightly redder/bluer than it would be if the planet were entirely transparent. Therefore, during ingress, the centroids of stellar lines are shifted slightly to the red, and during egress the centroids are correspondingly shifted to the blue.

This so–called Rossiter–McLaughlin effect (RME), described originally by Rossiter (1924) and McLaughlin (1924) in the case of eclipsing binary stars, adds to the shifts already caused by the radial velocity induced by the planet's gravity, described in (2.) above. The RME has been described in depth more recently in the context of extrasolar planets by Ohta et al. (2005), Giménez (2006), and Gaudi & Winn (2007). These centroid–shifts are expected to be comparable in magnitude to the radial velocity wobble from the planet's

gravity, and can be roughly estimated as

$$|\delta v_{\text{R-M}}| \sim 1 \text{ km s}^{-1} \times (R_p/R_*)^2 \sim 10 \text{ m s}^{-1},$$

In fact, the amount of the shift can be predicted precisely for a given orientation of the planet's orbit, and so measuring the shift is tantamount to measuring the alignment between the star's spin and the planet's orbit. Three years ago, Winn et al. (2005) first found that the spin of HD209458 and the orbital plane of its planet are nearly aligned. The degree of alignment has been measured for two other systems – Winn et al. (2006) found that the spin of HD189733 and its planet are also nearly aligned, and Narita et al. (2007) measured a mis–alignment between these two vectors by $\sim (30 \pm 20)°$ in the TrES–1 system.

4. *Planet's Atmospheric Opacity*:

   Furthermore, a gas–giant planet's opacity surely does not have a perfectly sharp discontinuity at an outer boundary, as a billiard ball does. Instead, it has an extended atmosphere in which the opacity must vary more or less smoothly. There may be a cloud layer, below which the planet is entirely opaque to tangential rays and above which the opacity varies smoothly. Most critical to our investigation, at a given radius, the planet's opacity to tangential lines of sight must vary with wavelength, depending on the contents of its atmosphere. At wavelengths of strong atomic or molecular transitions, the

planet's atmosphere will be more opaque than at other wavelengths. As a result, the effective radius of the planet, or the radius at which the optical depth along a tangential ray is of order unity, is greater at some wavelengths than at others. These effects have been described in detail by B01.

5. *Planet's Orbital Motion*:

   The motion of the planet's atmosphere must influence the transit spectrum in several delicate ways. As B01 points out, there are three main mechanisms by which the motion of a planet's atmosphere relative to its star can affect the spectrum: the planet's orbital velocity along the line–of–sight, the planet's (possibly tidally locked) rotation, and winds in its atmosphere. The largest effect of these three is the orbital velocity, which imposes a bulk blue/redshift during ingress/egress of ~ 15 km s^{-1} to spectral lines arising from the planet's atmosphere. These shifts are of opposite sign to the radial velocity wobble and to the shifts from the RME, and therefore tend to lessen the apparent RME slightly.

6. *Planet's Atmospheric Motion*:

   The most dynamically interesting (and subtlest) effects are those caused by the planetary rotational velocity and atmospheric winds. Since a tidally locked planet rotates in the same sense as it orbits, the rotational velocity of its outside edge has the same sign as its orbital velocity, and the rotational velocity of its inside edge has the opposite sign. As a result, during the beginning of ingress and the end of egress, when only the inside edge of the planet is over the star,

tidally locked rotation will impose a spectral distortion that is in the opposite sense of that caused by the bulk orbital velocity described in (5.) above, and that is in the same sense as the RME: the distortions are roughly equivalent to a relative redshift during ingress (graphically represented in Figure 4.1) and a relative blueshift during egress. During mid–transit, with some parts of the rotating planet's atmosphere moving toward and other parts away from the star relative to an otherwise identical but non–rotating planet, the overall influence of the planet's rotation is approximately equivalent to rotational broadening.

Winds complicate the picture even further. It is likely that winds tend to rush from the substellar hot spot to the colder night side of the planet. With the substellar point on the opposite side of the planet from Earth during a transit, this corresponds to winds rushing toward us at several hundred to several thousand meters per second. This would tend to blueshift the spectrum throughout the transit. Zonal wind bands, somewhat similar to those on Jupiter but with much higher speeds, or other more detailed winds, can have an even more intricate effect.

7. *Additional Effects*:

If a transiting planet were to have nonzero orbital eccentricity, or rings, these could complicate a measurement of rotation rate. Nonzero eccentricity would break the symmetry between ingress and egress. Still, if the orbit were well-known, this could be modeled and taken into account. It seems unlikely

that a Hot Jupiter could maintain rings: Icy rings would sublimate, and, if not continuously replenished, dusty/rocky rings would quickly succumb to the Poynting–Robertson effect (Poynting 1903; Robertson 1937). But if, somehow, a ring were to find a way to persevere around a transiting Hot Jupiter, it could confound – perhaps hopelessly – a measurement of rotation. The consequences of rings for the Rossiter–McLaughlin effect are addressed in Ohta et al. (2006). Saturn's rings are nearly four–times the area of the planet, so for a planet (with equatorial rings that are as relatively large as Saturn's) whose orbit is tilted an angle 0.1 (in radians) from edge-on, the rings would be ∼ 40% the area of the planet, which would increase the RME by ∼ 40%. Uncertainty about the presence and size of a ring introduces an uncertainty in the size of the RME effect that is probably larger than the size of the rotation effect. Furthermore, a ring would occlude a (small) part of the planet's atmosphere, which would (slightly) reduce the strength of the rotation signal.

Other interesting phenomena that primarily affect a transit light curve, rather than the spectrum, include star–spots (Silva 2003), atmospheric lensing (Hui & Seager 2002), and finite–speed–of–light effects (Loeb 2005). Although Winn & Holman (2005) describe a possible configuration (Cassini state 2) that would produce a spectral signature that is nearly identical to what would be expected from a non–rotating planet, the likelihood that any Hot Jupiters are in such a configuration might be low, and it seems quite likely that *some* transiting planets are not in this

state. Nonetheless, the motion of a transiting planet's atmosphere – rotational, wind, or other – is clearly interesting, and the basic technique that we describe below is applicable to any model of atmospheric motion.

4.2.2 Preview of Results

A rough estimate of the velocity–shift that is imposed during ingress to the centroids of the stellar Na D–lines by the planet's tidally locked rotation (on top of the RME and the shift from the planet's orbital velocity, both of which would be present even if the planet were not rotating) is the following:

$$
\begin{aligned}
\delta v \;\sim\; & \left(\langle \cos[\phi] \rangle_{-\pi/2}^{\pi/2} \right) \times \left(\frac{1}{2} \times \frac{R_p{}^2}{R_*{}^2} \right) \times \left(\frac{2\pi R_p \Pi_{\text{atm}}}{\pi R_p{}^2} \right) \times v_{\text{eq}} \\
\sim\; & 0.64 \times 1\% \times 15\% \times 2000 \text{ m s}^{-1} \\
=\; & 1.9 \text{ m s}^{-1}.
\end{aligned}
\tag{4.1}
$$

In this equation, ϕ is a planet–centered azimuthal angle, R_p and R_* are the planet's and star's radius, respectively, Π_{atm} is the height of the planet's atmosphere, and v_{eq} is the equatorial rotation speed. The rotation speed at angle ϕ is $v_{\text{eq}} \cos[\phi]$. We take the average of $\cos[\phi]$ from $-\pi/2$ to $\pi/2$ to get the average planetary rotation speed. We have used $\Pi_{\text{atm}} = 7500$ km, or 15 times the presumed scale height of 500 km, because the sodium lines are so heavily saturated that at the assumed abundance and cloud deck height in our model the line cores do not become optically thin until that height. Burrows et al. (2004) and Fortney (2005) describe how the optical

depth along tangential rays is greater than the optical depth on normal rays. The product

$$\delta_{\text{atm}} \approx \left(\frac{1}{2} \times \frac{R_p{}^2}{R_*{}^2} \right) \left(\frac{2\pi R_p \Pi_{\text{atm}}}{\pi R_p{}^2} \right) = \left(\frac{R_p}{R_*} \right)^2 \left(\frac{\Pi_{\text{atm}}}{R_p} \right)$$

is the ratio of the area of the portion of the planet's atmosphere that is in front of the star halfway through ingress to the total area of the disk of the star. Based on this estimate, we expect a maximum velocity shift of $\delta v \sim 190$ cm s^{-1}. If we take into account that HD209458b's orbit is actually slightly inclined relative to the line of sight, the cosine average decreases to ~ 0.45, and the total estimate decreases to ~ 140 cm s^{-1}. This estimate is in reasonably good agreement with the centroid–shifts predicted by the full model calculation below (~ 60 cm s^{-1}); the difference between the estimates is most likely due to the difference between the shapes of the stellar and planetary lines.

We now estimate the signal–to–noise ratio for the detectability of this effect in an observation of duration Δt, with a telescope that has diameter D and throughput efficiency η. The signal is the distortion of the spectrum relative to a non–rotating planet, and for now we will assume that the noise is dominated by photon noise. If a spectrum $F[\lambda]$ with a symmetric absorption feature of depth ΔF centered at λ_0 is redshifted by an amount $\Delta \lambda$ to $\widehat{F}[\lambda] \equiv F[\lambda - \Delta \lambda]$, what is the integrated absolute difference $|F - \widehat{F}|$ over some wavelength range $2L$ centered on λ_0? If the absorption feature is wide compared with $\Delta \lambda$, then, by symmetry,

$$S \;=\; \int_{\lambda_0 - L}^{\lambda_0 + L} |F[\lambda] - F[\lambda - \Delta \lambda]| \, d\lambda$$

$$\approx \ 2 \int_{\lambda_0}^{\lambda_0+L} (F[\lambda] - F[\lambda - \Delta\lambda]) \, d\lambda; \tag{4.2}$$

and if $\Delta\lambda$ is small then

$$S \ \approx \ 2\Delta\lambda \int_{\lambda_0}^{\lambda_0+L} F'[\lambda] d\lambda$$

$$\approx \ 2(\Delta\lambda)(\Delta F). \tag{4.3}$$

We may now estimate the S/N of our effect (for a single absorption line) using the lesson of equation (4.3), provided we know the absolute normalization of the stellar spectrum (the number of photons per unit wavelength). A spherical blackbody of radius R_* and temperature T_*, at distance d from the telescope, has a photon number flux at wavelength λ of

$$\frac{d\dot{N}_\gamma}{d\lambda} \ \sim \ B_\lambda[\lambda, T]\left(\frac{1}{hc/\lambda}\right)\left(\frac{\pi R_*^2}{d^2}\right) \times \eta\pi(D/2)^2$$

$$= \ \frac{\pi^2 c}{2\lambda^4 \left(\exp[(hc)/(\lambda k T_*)] - 1\right)} \times \eta\left(\frac{R_* D}{d}\right)^2, \tag{4.4}$$

where B_λ is the Planck function. Since the fractional decrease in the spectrum at the line–center is approximately δ_{atm}, we may express the parameter ΔF from equation (4.3) as $\Delta F \approx \delta_{atm}(d\dot{N}_\gamma/d\lambda)$. Similarly, since the root–mean–square velocity shift during ingress is $\langle v^2 \rangle^{1/2} \sim (1/2) \times (2000 \text{ m s}^{-1}) = 1000 \text{ m s}^{-1}$,[3] we may express the parameter $\Delta\lambda$ as $\Delta\lambda \sim (\langle v^2 \rangle^{1/2}/c) \times \lambda_0$. The distortion (the signal) from a single

[3] We write $(1/2) \times (2000 \text{ m s}^{-1})$ because the mean value of $\cos^2$ from $-\pi/2$ to $\pi/2$ is $1/2$

line can therefore be estimated as

$$
\begin{aligned}
S &= \delta N_\gamma \sim 2(\Delta\lambda)\left(\delta_{\text{atm}}\frac{d\dot{N}_\gamma}{d\lambda}\right)\Delta t \\
&= \frac{\pi^2 c\,(\delta_{\text{atm}})\,\Delta\lambda}{\lambda^4\,(\exp[(hc)/(\lambda k T_*)]-1)} \times \eta\left(\frac{R_* D}{d}\right)^2 (\Delta t).
\end{aligned}
\tag{4.5}
$$

The shot–noise is the square root of the number of photons in a wavelength range $2L$ roughly equal to the FWHM of the line, or about 7 km s^{-1} for a heavily saturated line such as the Na D lines under consideration:

$$
\begin{aligned}
N &\sim \sqrt{\frac{d\dot{N}_\gamma}{d\lambda}(2L)(\Delta t)} \\
&\sim \sqrt{\frac{\pi^2 L c(\Delta t)}{\lambda^4\,(\exp[(hc)/(\lambda k T_*)]-1)}} \times \sqrt{\eta}\left(\frac{R_* D}{d}\right).
\end{aligned}
\tag{4.6}
$$

We estimate the total signal–to–noise ratio arising from a single absorption line, during an ingress integration of duration Δt, to be roughly

$$
\begin{aligned}
S/N &\sim \frac{\pi(\delta_{\text{atm}})}{\sqrt{\exp[(hc)/(\lambda k T_*)]-1}}\left(\frac{\Delta\lambda}{\lambda}\right)\left(\sqrt{\frac{c\Delta t}{L}}\right)\left(\frac{R_* D}{d\lambda}\right)\sqrt{\eta} \\
&\sim \left(6.6\times10^{-4}\right)\left(3.3\times10^{-6}\right)\left(2.1\times10^{11}\right)\left(5.0\times10^{-3}\right)\sqrt{\eta} \\
&\sim 2.3\,\sqrt{\eta}.
\end{aligned}
\tag{4.7}
$$

The above calculation uses parameters for HD209458 and its planet, a sodium D line, and a 6 m telescope: $\lambda = 600$ nm; $\Delta t = 1000$ s; $R_* = 7.3\times10^{10}$ cm; $T_* = 6100$ K, $d = 47$ pc; and $D = 600$ cm. For two identical absorption lines, we gain a factor of $2^{1/2}$ in S/N, and for egress we gain another factor of $2^{1/2}$, giving a total one–transit

S/N of roughly $4.6\eta^{1/2}$, not counting the additional signal available during mid–transit (see further discussion below). This S/N ratio is in principle independent of the spectral resolution of the spectrograph, for sufficiently high spectral resolution. For low spectral resolution, however, the S/N could be lower than this estimate (below, we conclude that the S/N loses its dependence on resolving power for spectral resolution $\gtrsim$ 500,000).

There were several optimistic assumptions that went into this estimate. Still, this rough estimate of the degree to which a planet's rotation influences its transit spectrum indicates that the more in–depth study that we perform below is warranted.

4.2.3 Available Technology

Detecting the centroid–shifts caused by tidally locked rotation ($\lesssim 1\,\mathrm{m\,s^{-1}}$) will require very precise measurements of stellar transit spectra. Obtaining such high precision spectra will be quite challenging, for a number of reasons, several of which were described in the groundbreaking paper by Butler et al. (1996) that analyzes the limits of Doppler precision. Of particular concern, stellar pulsations and turbulent motions in stellar photospheres can cause small regions of the stellar disk to move at up to 300 m s^{-1} (Dravins 1985; Ulrich 1991). These motions tend to average out to produce stellar spectra that are largely stable; but it is likely that at least some giant convection cells are not small relative to the size of a planet, and these could introduce a contaminating source of noise when they are located

behind the planet or its atmosphere. Butler et al. (1996) reviewed what was then known about the variability of stellar line–profiles; the upshot is that line–widths may vary by up to several meters per second over several years, but it is not clear to what extent spurious apparent velocity shifts may be induced by convection, and such stellar jitters may prove to be a significant source of noise that would render it difficult to measure sub meter–per–second velocity–shifts. More recently, Bouchy et al. (2005a) have actually achieved sub meter–per–second accuracy with the HARPS instrument (spectral resolution of 115,000), and they have found a dispersion in night–averaged radial velocity measurements for a particular star (HD160691) of ~ 0.4 cm s^{-1} for nights when they took many ($\gtrsim 200$) observations. Since in our situation (taking spectra during ingress, say) we have minutes, not hours, available, the rms scatter in ingress–averaged radial velocity measurements is likely to be larger than what they found. In addition to the difficulties posed by several systematic sources of noise, achieving sufficient photon statistics will be difficult for two reasons: for a given throughput efficiency η, higher spectral resolution means fewer photons per bin; and η tends to decrease with increasing spectral resolution R_S.

By the mid–1990s, the timeless quest for high–resolution spectrographs reached a milestone at the Anglo–Australian Telescope with the development of UHRF and its resolving power of up to $1,000,000$ (Diego et al. 1995). Despite impressive throughput relative to previous endeavors, however, its efficiency was insufficient to obtain the sub decameter–per–second Doppler precision on a $V \geq 7$ star that

would be required for planet searches. With a $R_S = 600,000$ spectrograph built at Steward Observatory, Ge et al. (2002) obtained stellar spectra with $R_S \sim 250,000$ and throughput of 0.8%. Furthermore, they predicted that by optimizing their technology they could increase the throughput to 4%. More recently, Ge et al. (2006) detected a new planet, around HD 102195, with the Exoplanet Tracker instrument at Kitt Peak. This instrument has resolution of $R_S \sim 60,000$ and total throughput of 18%. Plans for a spectrograph that has resolving power of 120,000 on a thirty meter telescope (Tokunaga et al. 2006) give cause for optimism that increased aperture area and efficiency feeding high and ultrahigh–resolution spectrographs will, in coming years, provide accurate enough spectra that tidally locked rotation of HD209458b has a detectable influence.

4.3 A Model of a Planetary Transit

We consider the spectrum of a star whose companion planet transits across the face of the stellar disk from Earth's perspective. The primary effect of the planet is to reduce the stellar flux at all wavelengths, but the planet's chemical composition, internal structure, and rotation rate influence the spectrum in wavelength–dependent ways. Since each of these factors – and others too, such as the star's rotation – influences the observed spectrum, we built a model that incorporates the many parameters related to each process. The star and the planet are both assumed to rotate as solid bodies, with no other (nonthermal) motion in their atmospheres. Since deviations from pure solid body rotation are likely to be no more than 25%

over the disk of the star – e.g., the Sun's equator–to–pole variation in rotation rate is about 21%, as per Howard et al. (1984), this is probably a reasonable assumption for the star. For the planet, this assumption might fail, because wind–speeds in excess of the equatorial rotation speed of $v_{\mathrm{eq}} \approx 2 \mathrm{\ km\ s^{-1}}$ are predicted by many models, as described in § 4.1 above. Still, when making this initial study of the spectral effect of the motion of a transiting planet's atmosphere, separating rotation from other processes makes the problem more tractable. We set parameter values to match measured values from the HD209458b system where possible.

The planet is modeled as an inner component that is entirely opaque and an outer component that is isothermal and drops off exponentially. We compute the wavelength–dependent optical depth due to the sodium D-doublet at ≈ 590 nm in the planet's atmosphere; important parameters include the temperature and density of the planet's atmosphere and its Na–content. We use the Voigt profile – as described by, e.g., Press & Rybicki (1993) – to calculate $\tau[\lambda]$, the optical depth to absorption along the line of sight.

As the planet transits the star, there are four points of "contact" between the planet and the star (really between their projections on the sky): when the disk of the planet first touches the disk of the star; when the planet is first entirely over the stellar disk; when the planet is last entirely over the stellar disk; and when the planet last touches the stellar disk. We will additionally sometimes refer to "1.5[th]" contact (half–way between first and second contact), and analogously to "2.5[th]" and "3.5[th]" contact.

As described in § 4.2 above, the type of distortion that a planet's rotation imposes relative to a non–rotating planet changes depending on when during the transit the observation is made. During ingress or egress, the rotation of a tidally locked planet's atmosphere will impose a distortion similar to an overall shift relative to a non–rotating planet: redshift during ingress; blueshift during egress. When the planet is in mid–transit, in the middle of the stellar disk, the overall distortion to the spectrum imposed by its rotation is akin to a star's rotational broadening. Since the line–centers of the lines we are considering are heavily saturated and therefore flat at their cores, rotational broadening has the somewhat counterintuitive effect of steepening the cores of the profiles while broadening the wings. We will discuss this in greater detail in the next section. Although the type of distortion is different during ingress and egress from during mid–transit, it turns out that the amount of distortion, in terms of S/N ratio, is nearly constant throughout transit. This, too, we will discuss in § 4.4 below.

We simulate the HD209458b system, with a $1.32 R_J$ planet in a 3.5 day orbit, orbiting a G0 star at with radius $1.05 R_\odot$ that is 47 pc away. Our model star has the limb darkening profile that Knutson et al. (2007b) measured for HD209458. In order to approximate the fits to the data in Charbonneau et al. (2002), we assign our model planet's atmosphere a sodium–content and cloud deck height (1% solar, and 0.01 bars) that are comparable to the parameter–combinations that result in the best fits in that paper. Finally, we present results at our simulation's spectral resolution ($R_S = 700,000$), and we simulate transit events observed using two different lower

resolution spectrographs, one with spectral resolution $R'_S = 50,000$ and one with $R'_S = 150,000$. All spectrographs (and associated optical paths) in our simulations have 100% throughput efficiency. In the remainder of this section, we provide a detailed description of our parameterization of the problem.

4.3.1 Parameters of the Star

The parameters related to the star are listed in Table 4.1. They are set to match measured parameters for HD209458, and we use the limb–darkening profile from Knutson et al. (2007b). We normalize the flux to that of a blackbody of temperature T_* of the size and at the distance of HD209458.

4.3.2 Parameters of the Planet

The parameters related to the planet are in Table 4.2. We model the planet as an inner component that is essentially a billiard ball (completely opaque at all wavelengths) and an outer component that is an isothermal atmosphere with scale height $H = R_{\mathrm{gas}} T_p / \mu g$, where R_{gas} is the gas constant, μ is the molar mass, and g is the acceleration of gravity. The density of our model planet's atmosphere varies as $\rho = \rho_0 \exp[(r - R_{p_0})/H]$, where R_{p_0} is the radius of the optically thick part (some authors have called this radius the "cloud–deck" (Charbonneau et al. 2002)). This hypothetical cloud deck could cause the planet to be optically thick at a higher altitude than would otherwise be expected, as discussed in, e.g., Richardson et al. (2003) and Sudarsky et al. (2000). The cloud deck causes the optical depth as a

function of radius in our model to have a singular discontinuity at radius R_{p_0}.

4.3.3 Spectral Parameters

The parameters pertaining to the shape of the observed spectrum are in Table 4.3. In addition to the location of the planet within the stellar disk, the shape of the stellar spectrum and the wavelength–dependent opacity of the planet's atmosphere together influence the transmission spectrum.

`Spec_Shape` is a parameter that can take on the values "Flat", "Blackbody", or "Solar", and determines the rest–frame spectrum of the model stellar photosphere. (The integrated stellar spectrum is the convolution of the rest–frame spectrum with the stellar rotation profile.) When "Flat" is chosen, the rest–frame model stellar spectrum intensity is set to the mean value of the blackbody intensity in the specified wavelength range $[\lambda_{\min}, \lambda_{\max}]$, which, in our simulation, is set to [580 nm, 600 nm]. When "Solar" is chosen, the model stellar spectrum instensity is set to a high–resolution solar spectrum that is normalized to the flux from HD209458[4]; but the Na D lines in this high–resolution spectrum have been replaced by Gaussian fits to the solar lines.

The planet's atmosphere has N_{abs} absorption features, each of which is due to an element with a given fraction of the solar abundance. In the models presented in this paper, $N_{\mathrm{abs}} = 2$: we consider the Na doublet at 588.9950 nm and 589.5924 nm, with sodium at fractional abundance $f_\odot \equiv X_{\mathrm{Na}\ p}/X_{\mathrm{Na}\ \odot} = 0.01$ of the solar

---

[4]From *ftp://solarch.tuc.noao.edu/*.

abundance. Each line is modeled as a Voigt profile, as described in, e.g., Press & Rybicki (1993).

4.3.4   Parameters of Observing and Computing

The final set of parameters, listed in Table 4.4, includes those that specify the observer and those that determine how the observation is discretized for the purpose of numerical computation. The model observational setup is determined by three parameters: the telescope's diameter D (6 m in our simulations) and efficiency η (100%), and the spectrograph's spectral resolution R_S (we set R_S to 700,000 for the purpose of computing the model, and we re–bin to lower, more easily achieved resolutions – 150,000 and 50,000 – after computing a model). These three parameters prescribe the sizes of the spectral bins and the rate at which those bins are capable of collecting light.

In order to compute the flux at Earth as a function of wavelength, we begin by dividing the stellar disk into n_{b*} concentric annuli, and we divide each annulus into n_{ϕ_*} azimuthal sections. In each section, the redshifted spectrum and the normalization must both be computed.

Knowing the stellar rotation rate and axis, we may calculate the recessional velocity of any point on the star's surface as a function of its location on the projected stellar disk, and we redshift the spectrum from each part of the star accordingly. When the planet is in–transit, we separate the stellar disk into an annulus that contains the planet and the rest of the disk that we treat as described above. The

annulus that contains the planet is treated almost as above – divided into n_{b*A} sub–annuli, each of which has $n_{\phi_{*A}}$ azimuthal sections – but the sub–annuli are incomplete, interrupted by the planet.

In order to sample the planet's atmosphere, we divide the region that overlaps the star into n_{b_p} concentric annuli around the planet's center, each of which is divided into n_{ϕ_p} azimuthal sections. In each section, we must determine the optical depth and multiply by $\exp(-\tau)$. In calculating the optical depth, we note that in the case that the planet's rotation axis is entirely normal to the line–of–sight, if the planet rotates as a solid body then the radial component of its recessional velocity is constant along a ray:

$$\tau[b_p, \phi, \lambda] = N[b_p] \times \sigma \left[\frac{\lambda}{1 + (v_p[b_p, \phi_p]/c)} \right], \tag{4.8}$$

where the column density is calculated in terms of a function G that is specified below: $N[b_p] = n_0 G[b_p, R_{p_0}, H]$. In equation (4.8), $v_p[b_p, \phi_p]$ is the recessional velocity of the planet, as a function of radius and azimuth, which depends upon the orbit and the rotation. Note that there is a single v_p along a given line–of–sight defined by a (b_p, ϕ_p) pair only under the assumption of solid body rotation. The rest–frame cross–section $\sigma[\lambda]$ is computed according to the Voigt profile. The function G is defined as the following integral:

$$G[b_p, R_{p_0}, H] = \begin{cases} \int_{-\infty}^{\infty} \exp\left[-\frac{\sqrt{b^2+l^2}-R_{p_0}}{H}\right] dl & b_p > R_p \\ \infty & b_p \leq R_p \end{cases} . \tag{4.9}$$

4.4 Model Transit Spectra

As described in § 4.2.2, we seek the expected signal–to–noise ratio for distinguishing between the spectrum that would be observed due to a non–rotating planet (or one that is is in a Cassini state with its rotation axis nearly in the plane of orbit) and the spectrum that would be observed due to a tidally locked planet. The computed model spectrum $\mathcal{N}[\lambda]$ is the time integral of the instantaneous spectrum $\dot{\mathcal{N}}[\lambda]$ and consists of the number of photons detected per wavelength bin:

$$\mathcal{N}[\lambda] \approx \dot{\mathcal{N}}[\lambda]\Delta t_{\mathrm{obs}},$$

for some small exposure time Δt_{obs}.

The model signal (of rotation) per bin that we are looking for is the difference between the rotating model spectrum $\mathcal{N}_{\mathrm{rot}}$ and the non–rotating model spectrum $\mathcal{N}_{\mathrm{no\ rot}}$:

$$S_b = \left(\dot{\mathcal{N}}_{\mathrm{rot}}[\lambda] - \dot{\mathcal{N}}_{\mathrm{no\ rot}}[\lambda]\right)\Delta t_{\mathrm{obs}}. \tag{4.10}$$

We make the optimistic approximation that the noise per bin is just the photon–noise:

$$N_b = \sqrt{\dot{\mathcal{N}}_{\mathrm{no\ rot}}[\lambda]\Delta t_{\mathrm{obs}}}. \tag{4.11}$$

The total signal–to–noise ratio in a single exposure, then, is the sum in quadrature

of S_b/N_b for all wavelength bins λ_i:

$$S/N = \sqrt{\sum_{i=1}^{\#\text{bins}} \left(\frac{\dot{\mathcal{N}}_{\text{rot}}[\lambda_i] - \dot{\mathcal{N}}_{\text{no rot}}[\lambda_i]}{\sqrt{\dot{\mathcal{N}}_{\text{no rot}}[\lambda_i]}} \right)^2} \times \sqrt{\Delta t_{\text{obs}}}. \qquad (4.12)$$

A similar summation in quadrature applies over all exposures. Note that, in principle, the expression in equation (4.12) is insensitive to the sizes of bins and hence to the spectral resolution R_S, as long as the bins are small relative to the Gaussian width of the absorption feature under consideration. Our simulations indicate that the spectral resolution must be $\gtrsim 500,000$ in order for S/N to be nearly independent of R_S.

The effect of rotation, both during ingress and during mid–transit, is illustrated in Figure 4.2. For illustrative purposes, in this figure we assume a uniform star (flat spectrum, non–rotating, no limb–darkening). On the left panels of Figure 4.2, we show a snap–shot during ingress (at 1.5$^{\text{th}}$ contact) and on the right, we show a snap–shot during the middle of a transit (2.5$^{\text{th}}$ contact). The quantity plotted is $\mathcal{R}' = \mathcal{R} - 1$ from B01, where

$$\mathcal{R}[\lambda, t] = \frac{\dot{N}_{\text{in transit}}[\lambda, t]}{\dot{N}_{\text{out of transit}}[\lambda, t]} \qquad (4.13)$$

The bottom panels of Figure 4.2 show the difference spectra between the models with a tidally locked planet and the models with a non–rotating planet ($\mathcal{R}_{\text{rot}} - \mathcal{R}_{\text{no rot}}$).

As described in § 4.2 above, a planet's rotation causes the centroids of stellar
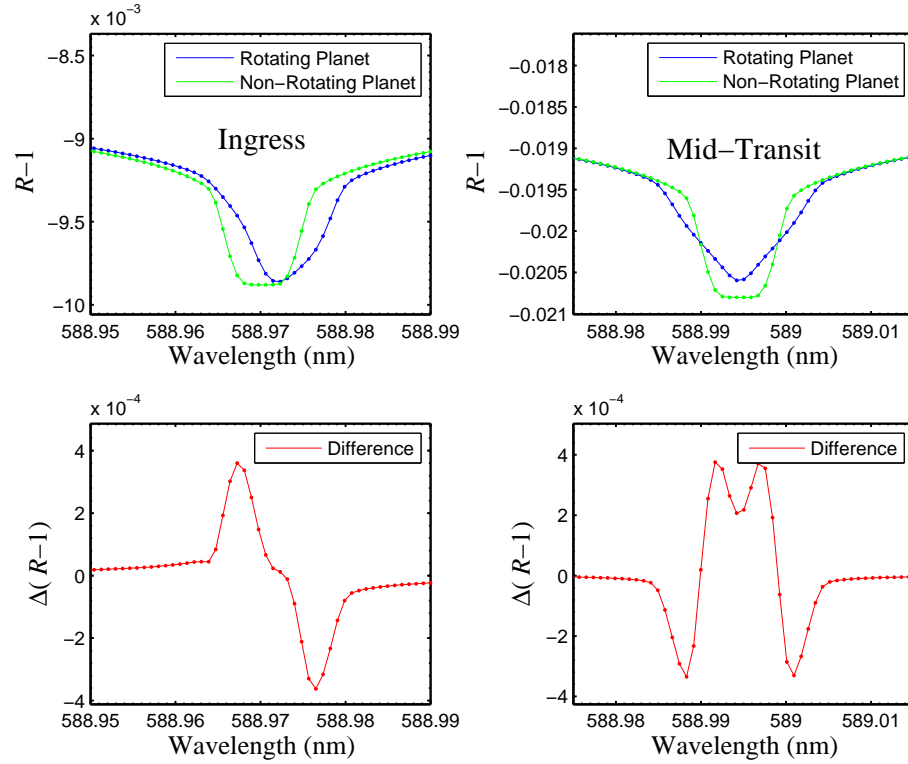
Figure 4.2 Upper panels show snap–shot spectra for one of the the Na D lines for two different model planets (tidally locked and non–rotating); lower panels show the difference between the two model spectra.The quantities plotted are $\mathcal{R}' = \mathcal{R} - 1$ (upper panels) and $\Delta\mathcal{R}'$ (lower panels), where $\mathcal{R}[\lambda, t] = \dot{N}_{\text{in transit}}[\lambda, t]/\dot{N}_{\text{out of transit}}[\lambda, t]$. In the upper panels, the blue curve is the tidally locked planet's transit spectrum, and the green curve is the non–rotating planet's transit spectrum. In the lower panels, the difference between the rotating and non–rotating planet's spectra. *Left*: Halfway through ingress (at 1.5[th] contact). *Right*: Halfway through the whole transit (2.5[th] contact).

Table 4.1   Model Transit Parameters: Star

Parameter	Description	Value
M_*	Star Mass	$1.05\ M_\odot = 2.09 \times 10^{33}$ g
R_*	Star Radius	$1.05\ R_\odot = 7.35 \times 10^{10}$ cm
T_*	Star Temperature	6100 K
d_*	Distance to star	47 pc
τ_*	Stellar Rotation Period	1 month

Table 4.2   Model Transit Parameters: Planet. Parameter values are set to match measured values from the HD209458b system where possible.

Parameter	Description	Value
M_p	Planet Mass	$0.69 M_J = 1.31 \times 10^{30}$ g
R_{p_0}	Optically Thick Planet Radius	$1.32 R_J = 9.44 \times 10^{5}$ km
P_0	Planet Pressure at R_{p_0}	0.01 bars
H	Planet Atmosphere Scale Height	500 km
T_p	Planet Atmosphere Temperature	1300 K
f_{TL}	Frac. Tidal Locked Rot. Rate	0 or 1 ($v_{\mathrm{eq}} = 0$ or 2 km s^{-1})
a	Semi–Major Axis	0.046 AU
$\#_H$	Number of Scale Heights in Atm.	15

Table 4.3   Model Transit Parameters: Spectral Features.

Parameter[a]	Description	Value
Spec_Shape	Shape of Star Spectrum	Flat, Blackbody, or Solar
$\lambda_{\min}$	Min. Wavelength in Sim.	580 nm
$\lambda_{\max}$	Min. Wavelength in Sim.	600 nm
N_{abs}	# Abs. Features in P. Atm.	2
$f_{\odot 1}$	Frac. Solar Abund., First Line	0.01
λ_{01}	First Line–Center	588.9950 nm
A_{ki1}	Transition Prob. First Line	6.16×10^{-9} s^{-1}
g_{i1}	Stat. Wt. Lower Level First Line	2
g_{k1}	Stat. Wt. Upper Level First Line	4
$f_{\odot 2}$	Frac. Solar Abund., Second Line	0.01
λ_{02}	Second Line–Center	589.5924 nm
A_{ki2}	Transition Prob. Second Line	6.14×10^{-9} s^{-1}
g_{i2}	Stat. Wt. Lower Level Second Line	2
g_{k2}	Stat. Wt. Upper Level Second Line	2

[a]In parameters that have i and k subscripts, i indicates the lower level ($3s_{1/2}$ for both lines) and k indicates the upper level ($3p_{3/2}$ for the bluer line and $3p_{1/2}$ for the redder line). The fractional solar abundance is set to 0.01 in order to achieve modest agreement with data observed for the Na D doublet in HD209458b's atmosphere.

Table 4.4 Model Transit Parameters: Observation and Computing. Parameter values are set to match measured values from the HD209458b system where possible.

Parameter	Description	Value
D	Telescope Diameter	2.4 m $-$ 30 m
η	Spectroscope Efficiency	1.00
R'_S	Obs. Spec. Resolution	50,000 - 700,000
$\mathcal{T}_{\text{int}}$	Integration Time	932.088 s
R_S	Comp. Spec. Resolution	700,000
Δt	Time–Step in Integration	50 s
n_{b*}	# of Star Annuli	10
n_{ϕ_*}	# of Star Azimuthal Sections	16
n_{b*A}	# of S. Annuli in P. Annulus	10
$n_{\phi_{*A}}$	# of S. Azim. Sec.'s in P. Ann.	10
n_{bp}	# of Planet Atm. Annuli	20
$n_{\phi p}$	# of Planet Atm. Azim. Sections	20
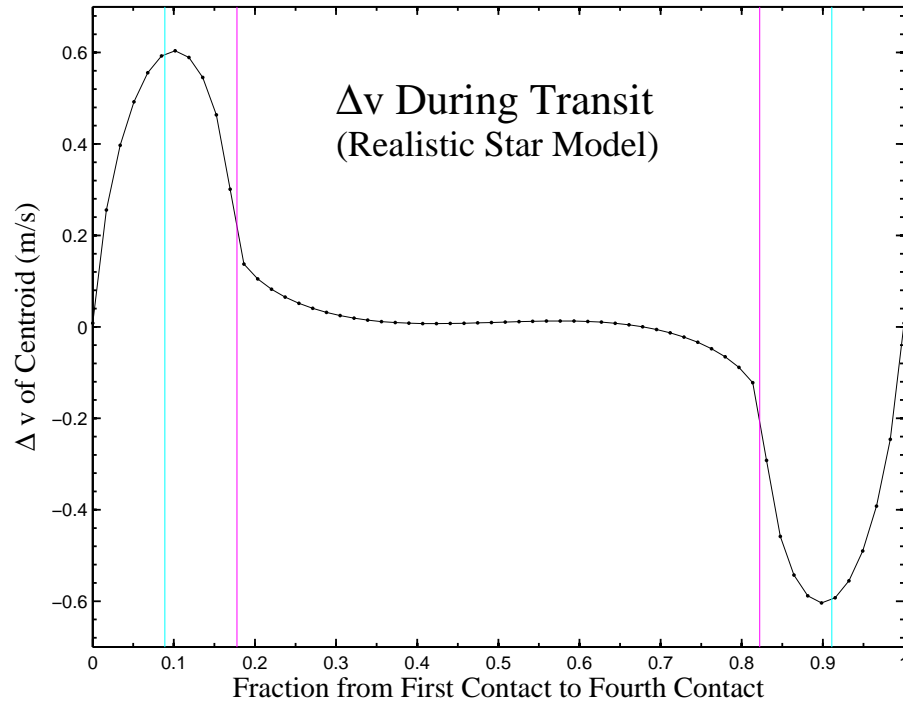
Figure 4.3 Centroid–shift of Na D lines from tidally locked rotation from the beginning to the end of a transit, relative to an identical but non–rotating planet; Sun–like stellar spectrum. The vertical lines denote 1.5th and 3.5th contact (cyan) and second and third contact (magenta). Between first and second contact, the spectrum with the rotating planet is redshifted relative to the non–rotating planet by up to about 60 cm s^{-1}; between third and fourth contact, it is blueshifted by the same amount. This plot samples the transit at 60 regularly–spaced points. Parameters were chosen to represent the HD209458 system.

absorption features to shift relative to a non–rotating planet. In Figure 4.3, centroid shifts (in velocity units) are plotted as a function of position in transit, for a planet transiting in front of a realistic star model with a Sun–like spectrum. The recessional velocity increases roughly sinusoidally during ingress, reaching a peak of about 60 cm s^{-1} at 1.5th contact. During mid–transit, between 2nd and 3rd contacts, the net velocity shift is much smaller. Egress is nearly perfectly symmetrical with ingress, though the velocity shifts have the opposite sign.

The cumulative and incremental signal–to–noise ratio across the transit are shown in the top and bottom panels, respectively, of Figure 4.4. As a planet proceeds in its transit from first contact to fourth contact, the S/N builds up steadily throughout the transit, as can be seen in the top panel of this figure. This cumulative increase reflects the steady incremental S/N per second of observation, which, for a planet crossing in front of a uniform star, is shown in the bottom panel. The three curves in the bottom panel represent, in decreasing order of S/N, the incremental S/N curves that are expected for a spectrograph with ultrahigh–resolution (the simulation's resolution $R_S = 700,000$), for a spectrograph with as high–resolution as bHROS on Gemini ($R'_S = 150,000$), and for a spectrograph with more standard high–resolution ($R'_S = 50,000$). In the top panel, we show the cumulative S/N (assuming the simulation's resolution) for a planet in front of a uniform star, and for a planet in front of a more realistic sun–like star that rotates once per month, has a limb–darkening profile, and has a solar spectrum. It is apparent in the top panel of Figure 4.4 that a spectrograph with our simulation's resolution would see

the effect of rotation at $S/N = 7.1\eta^{1/2}$ in a single transit in the case of the simplified uniform star, and at $S/N = 3.2\eta^{1/2}$ in one transit of the realistic star. The effect of including in the simulation the realistic features of stellar rotation, limb–darkening, and a solar spectrum is therefore to depress the S/N of the effect of tidally locked rotation by a factor of slightly more than 2. The bottom panel of the figure indicates that our predicted S/N for one transit of a realistic star ($3.2\eta^{1/2}$) might be adjusted downward by $\sim 50\%$, depending on the spectral resolution, indicating a total one–transit S/N for the case of the realistic star of about $\sim 1.7\eta^{1/2}$. Finally, we do not show the incremental S/N for the realistic star model, but, as the top panel of the figure indicates, a larger proportion of the total signal comes from ingress and egress in the realistic star model than in the uniform star model.

4.5   Discussion

The S/N scales with the square root of the number of absorption lines under consideration, with the square root of the number of transits, and with several other parameters, as follows:

$$
\begin{aligned}
S/N \quad \sim \quad & \left(\sqrt{\frac{N_{\text{abs}}}{2}}\right)\left(\sqrt{\text{\# transits}}\right)\left(\frac{D\sqrt{\eta}}{6\text{ m}}\right) \times \\
& 10^{-0.2(V_*-V_0)} \times \left(\frac{R_*}{R_0}\right)^{-2}\left(\frac{R_pH}{R_{p_0}H_0}\right)\left(\frac{v}{2000\text{ m s}^{-1}}\right) \times \\
& \left(\psi\left[R'_S\right]\right)(S/N)_0
\end{aligned}
\tag{4.14}
$$
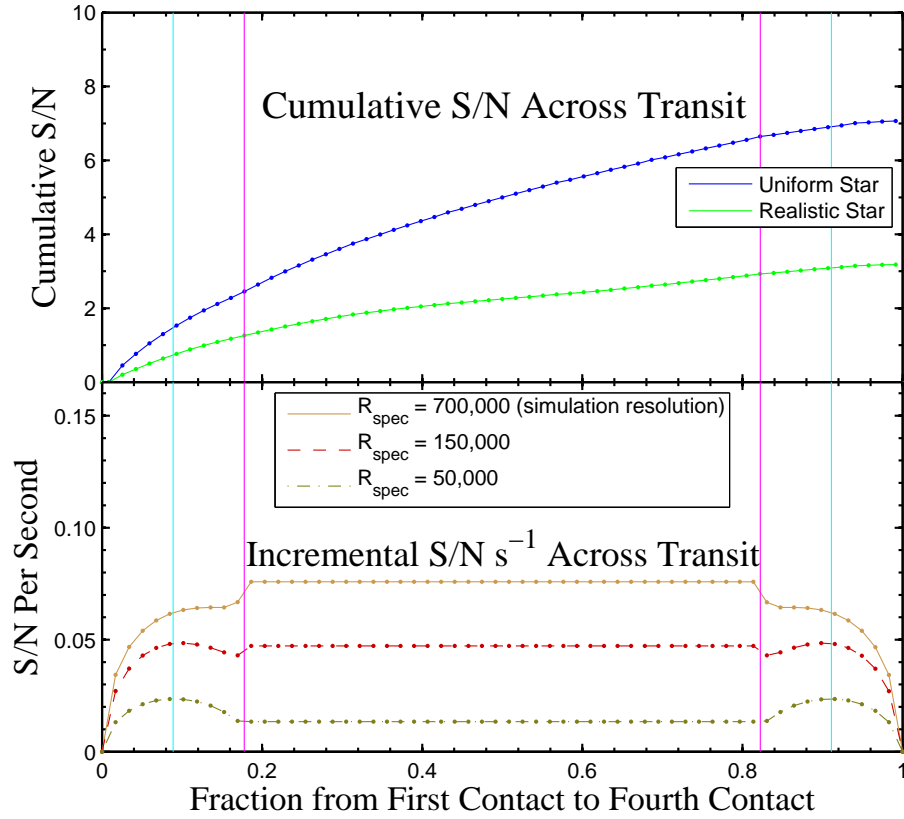
Figure 4.4 *Top*: S/N for distinguishing between a rotating and a non–rotating planet, accumulated during a single transit. Upper curve displays the cumulative S/N for the distinction of a tidally locked planet from a non-rotating model, assuming a completely uniform star: the model star has a flat spectrum, is not rotating, and there is no limb–darkening. Lower curve displays the same, but for a more realistic star: the model star rotates once per month, has the limb–darkening profile determined by Knutson et al. (2007b), and has a solar spectrum. These curves assume the spectral resolution used in the simulation ($R_S = 700,000$); for more easily achievable spectral resolution, these curves are adjusted downward by 30% - 80%, as displayed in the bottom panel. The vertical lines denote 1.5th and 3.5th contact (cyan) and second and third contact (magenta). Also, both top and bottom panels show model results assuming 100% throughput efficiency ($\eta = 100\%$). In practice, these curves would be adjusted downward by a factor $\eta^{1/2}$, or $\sim 0.04^{1/2} = 0.2$ for an ultrahigh resolution spectrograph. *Bottom*: Incremental S/N for sixty 1–second observations across the transit, assuming a uniform, non–rotating star with a flat spectrum. Top curve shows the incremental S/N, assuming a spectrograph with the spectral resolution of the simulation ($R_S = 700,000$). Middle and bottom curves show the same, assuming spectral resolutions $R'_S = 150,000$ and $R'_S = 50,000$, respectively. Parameters of the system (except for the stellar spectrum in the bottom panel) were chosen to represent the HD209458 system.

In this equation, N_{abs} is the number of absorption lines, D is the telescope's diameter, η is the efficiency of the instrumental setup, V_* is the star's magnitude, V_0 is HD209458's magnitude (about 7.6), R_* is the stellar radius, R_0 is HD209458's radius, R_p and H are the planet's radius and scale height, and R_{p_0} and H_0 are HD209458b's radius and scale height. $\psi[R'_S]$ is a function that characterizes the (nonlinear) response of S/N to the spectral resolution of the spectrograph; for instance, $\psi[700,000] \approx 1$, $\psi[150,000] \sim 0.6$ and $\psi[50,000] \sim 0.2$. Finally, $(S/N)_0 = 7.1$ for a uniform star and $X_{\mathrm{Na}\,p}/X_{\mathrm{Na}\,\odot} = 0.01$; and $(S/N)_0 = 3.2$ for a more realistic (rotating, limb–darkened) star with a solar spectrum.

Culling the signal from mid–transit, however, is much more difficult than from ingress and egress, because the shape of the distortion depends more sensitively on the structure of the planet. For the realistic star, our model predicts $S/N \sim 1.3\eta^{1/2}$ (with the 6 m telescope, with Poisson–dominated photon noise) for ingress alone, and therefore a factor $2^{1/2}$ greater, or $S/N \sim 1.8\eta^{1/2}$ for ingress and egress.

In Table 4.5, we present the number of orbits that must be observed in order to make a 5σ detection of rotation, for various combinations of parameters. In all cases, we assume a realistically achievable optical setup, with spectral resolution $R'_S = 150,000$ and throughput $\eta = 4\%$. Improvement in R'_S will yield a modest improvement in S/N, and improvement in η could be quite significant. In the third column of Table 4.5 (A), we present the total available S/N for the whole transit, according to our model, while in the fourth column (B), we present the available S/N from just ingress and egress. The fifth and sixth columns are based on the S/N

in the fourth column.

Finally, as a sanity check, our model can be tested by comparing it to the analysis of Charbonneau et al. (2002). That analysis suggests that the sodium content of a planet's atmosphere can be determined by comparing the flux in a narrow band centered on the sodium resonance lines with the flux in a wider band surrounding but excluding the lines. In that paper, the decrement in the narrow band containing the sodium features is named n_{Na}, and was measured to be -2.32×10^{-4} for HD209458 during the middle of transit. They presented several models, all of which over–predicted the sodium decrement. The model that predicted the smallest magnitude of the decrement had 1% solar metalicity and cloud deck at 0.0368 bar; this model predicts $n_{Na} \sim -3.4 \times 10^{-4}$ in mid–transit. Our model (1% solar metalicity, cloud deck at 0.01 bar) predicts $\sim -4.1 \times 10^{-4}$ in mid–transit, as shown in Figure 4.5. We conclude that our model is in reasonable agreement with both Charbonneau et al.'s model ($\sim 20\%$) and with the actual data ($\sim 40\%$).

4.6   Conclusion

Our investigation indicates that, with currently available instruments, it will be difficult to obtain the sensitivity needed to achieve a minimal $S/N \gtrsim 5$ detection of tidally locked rotation of the planet HD209458b. Nevertheless, it appears that the effect of rotation will have significant (though small) influence on transit–spectra taken with current or near–future instruments. Because this influence is so small,

Table 4.5 Required number of transits to achieve a $S/N = 5$ detection of tidally locked rotation of HD209458b, assuming a planetary Na–content that is 1% solar. Assumptions: we are monitoring two spectral lines (the Na doublet) with a spectrograph with resolving power $R'_S = 150,000$, and an efficiency in the optical setup of 0.04. The third column "(S/N)/Tr. (A)" is the S/N per transit, as shown in Figure 4.4, and the fourth column is the S/N per transit, considering only ingress and egress. The "(A)" in the fifth and sixth columns indicates that the numbers of transits and the required duration are tabulated for S/N values from column three.

Spec_Shape	D	(S/N)/Tr. (A)	(S/N)/Tr. (B)	# Trans. (A)[a]	Dur. (A)[b]
Flat (unif. star)	6 m	1.1	0.52	~ 25	~ 3 months
Solar	6 m	0.48	0.27	~ 110	~ 1.25 years
Solar	10 m	0.79	0.45	~ 40	~ 5 months
Solar	30 m	2.4	1.3	~ 5	~ 3 weeks

[a]The required number of transits for a 5σ detection of tidally locked rotation, using S/N values from column three.

[b]HD209458b completes the number of orbits indicated in column 5 in the amount of time indicated in this column.
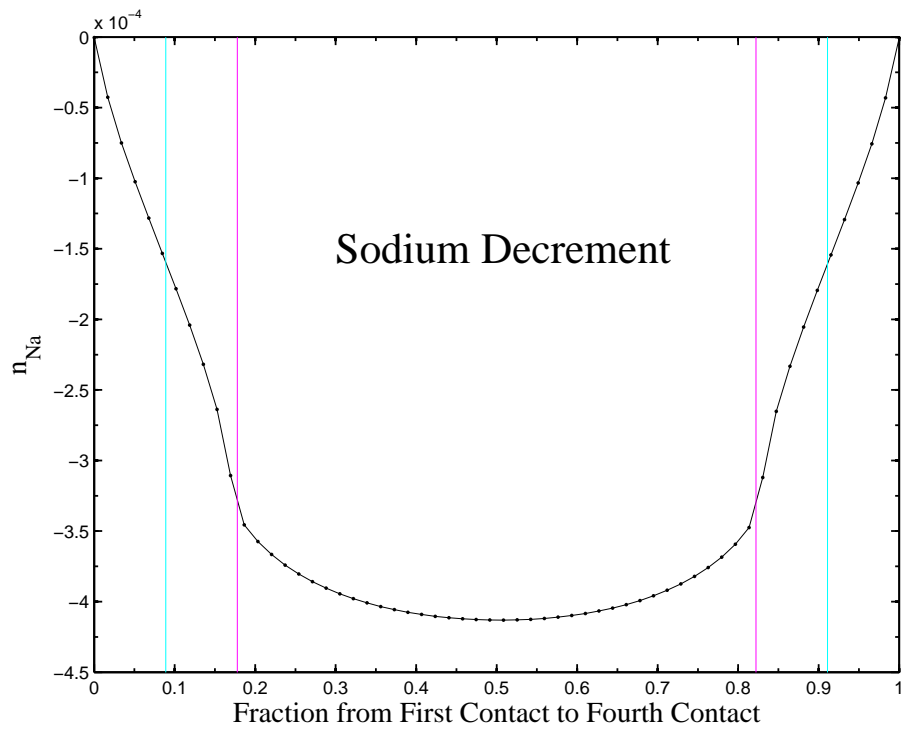
Figure 4.5 The flux decrement due to sodium in the planet's atmosphere, as shown by the quantity n_{Na} defined by Charbonneau et al. (2002). This decrement reaches a maximum in magnitude of -4.1×10^{-4}, halfway through transit, in our model.

it is worth considering ways that the *S/N* could be improved, according to the scalings summarized in equation (4.14).

- The most obvious way to boost S/N is to increase the collecting area of the telescope, the effect of which is shown in Table 4.5. With an extremely large (30 m–class) telescope, the required discrimination power could be achieved in ∼ 5 − 15 transits, if our optimistic assumptions about the noise are not too far off.

- As the abundances of other elements or molecules in the planet's atmosphere are identified, the number of absorption lines N_{abs} that can be utilized can increase. Observing four absorption features instead of two will boost the *S/N* per orbit by $2^{1/2}$ and will therefore cut in half the required observing time to achieve a fixed target *S/N*.

- The HD189733 system is only 19.3 pc away (Bouchy et al. 2005b), but the star is smaller and dimmer than HD209458, and has the same apparent magnitude. We find that, overall, it has no relative advantage in terms of *S/N*. It may be unlikely that a star significantly brighter than HD209458 will be found to have a transiting giant planet companion, but if this should happen the signature of rotation will be more readily apparent in the system with the brighter star.

- Finally, since S/N is proportional to the size of the atmosphere, represented by δ_{atm} in equation (4.7), a planet with a more extended atmosphere would show the distortion due to rotation more readily. The size of the atmosphere

depends on the product of R_p and the height of the atmosphere $\Pi_{\text{atm}} \sim 15H$, so the S/N could be increased in the case of a larger planet, a hotter one, or one with lower surface gravity. Unfortunately, although some other planets are somewhat larger, none of the puffier planets that have been discovered so far happen to transit a star as bright as HD209458. If a more bloated planet is found to transit a star as bright as HD209458, such a planet would probably be a more promising target for the type of study described in this paper.

Although the optical transit spectrum of an extrasolar giant planet contains many precious clues regarding the nature of rotation and climate or weather, we emphasize that we have so far addressed only the forward problem: deep, high spectral resolution transit spectra *will* be distorted by HD209458b's tidally locked rotation, if indeed it is tidally locked. But seeing the effect of rotation in a spectrum and *knowing* that we are seeing the effect of rotation are two very different things. The inverse problem – identifying the rotation rate – will be much more challenging, and is beyond the scope of this paper's investigation. In brief, it is impossible to decode the clues that the spectrum holds without a Rosetta stone – namely, without already possessing accurate knowledge of abundances in the planet's atmosphere. We note that when it comes to solving the inverse problem, it may be possible to make do without a perfect Rosetta stone by comparing the star's spectrum at two different times of transit instead of comparing it to an accurate input model. Specifically, comparing the ingress spectrum to the egress spectrum should eliminate various systematic uncertainties that may effect both ingress and

egress spectra. This strategy involves foregoing spectra between second and third contacts, at a cost of a factor of ~ 2 in S/N, but has the advantage of being more robust with respect to errors in modeling the planetary atmosphere. Furthermore, if detailed analysis along the lines of Charbonneau et al. (2002) should reveal many more planetary absorption features, a large current–generation telescope could detect the motion of an extrasolar planet's atmosphere in of order a month.

A final point to keep in mind is that the technique presented in this paper is not limited to discerning the difference between a tidally locked planet and a non–rotating planet; this technique can be used to investigate the applicability of any model of the motion of a transiting planet's atmosphere. Some atmospheric circulation models of Hot Jupiters predict extremely fast winds, up to ~ 9 km s^{-1}, or more then four times the predicted equatorial rotation rate. It might therefore turn out to be easier to detect winds than to detect rotation. Moreover, if an analysis similar to the one described here were to find a deviation from the hypothesis of tidally locked rotation, this would mean that *something* interesting is going on – the planet could be rotating roughly as a solid body but at a different rate; there could be significant non–rotational motion in the atmosphere; or there could be some combination of these phenomena.

Chapter 5

Habitable Climates

5.1   Introduction

Planetary science is being challenged by extrasolar planetary system discoveries (Mayor & Queloz 1995; Marcy & Butler 1996).[1] The orbital architectures of many of these systems, with massive planets located either very near to their parent stars or on highly eccentric orbits, are strikingly unusual. These discoveries have been surprising, however, mostly as a consequence of our own Solar-System centric point of view. In fact, planet searches at multi-AU orbital distances are now providing tentative evidence that Solar System giant planets, with their nearly circular orbits, are more the exception than the norm, by simple comparison with the ensemble properties of known extrasolar planetary systems (Marcy et al. 2005). If these preliminary trends hold, they may have a profound impact on our perception of

---

[1]See Reipurth et al. (2007) for a review. See also http://exoplanet.eu and http://exoplanets.org.

our place in the Universe.

The pace of exoplanet discoveries has accelerated sharply in the last few years and the future looks bright. Two significant developments in exoplanet research have occurred within the last year or so. A micro-lensing discovery of a likely terrestrial planet, with a mass 5.5 times that of Earth, was reported by Beaulieu et al. (2006), making it the first discovery of a planet that is thought to be terrestrial (i.e. rocky). Several months ago, Udry et al. (2007) announced the Doppler-velocimetry discovery of a potentially habitable terrestrial planet around a low-mass M-dwarf. In considering future discoveries, of specific interest to this study are two dedicated space missions: *Corot* and *Kepler* (Baglin 2003; Borucki et al. 2003, 2007). A key objective of both missions is to monitor a large number of stars to detect the (repeatable) micro-eclipses generated when terrestrial planets transit in front of their host star. It is expected that a few Earth analogues (i.e. with Earth-like masses and comparable distances to their host stars), and possibly hundreds to thousands of additional terrestrial planets unlike Earth, will be identified by *Corot* and *Kepler* after a few years of operation (Borucki et al. 2007, 2003; Basri et al. 2005). The discovery of an Earth-like exoplanet, potentially hosting life as we know it, is therefore within the 5-year astronomical horizon. Ambitious missions are also being prepared to map in detail the orbits of exoplanets around nearby stars (with *SIM PlanetQuest* (Unwin et al. 2007)) and later to obtain spectra of nearby Earth-like planets in the hope that they would reveal the first unambiguous signatures of life on a remote world (with the *Terrestrial Planet Finder* and *Darwin* Leger & Herbst

2007).

The idea that liquid surface water is a prerequisite for a terrestrial planet to have the ability to host life is widely used as the key concept behind searches for habitable planets around other stars. This is because of the central and critical role that water plays in the biochemistry on Earth.[2] As the current exoplanet census indicates, notions based on an old Solar System centric view may only be relevant to a minority of planets (or planetary systems). What if most terrestrial planets discovered in the future have, like the vast majority of exoplanets currently known, highly eccentric orbits generating large seasonal variations? What if the atmospheric mass and composition, planetary spin rate, or land-sea coverage of these exo-Earths are generally different from what they are on Earth? If the last ten years of extrasolar planet discoveries offer any guidance, our Solar System appears to show but little of the general planetary diversity found around other stars.

The classical calculations of habitability on Earth-like planets by, e.g., Dole (1964) and Hart (1979) predated extrasolar discoveries. Apart from a few exceptions (Franck et al. 2000; Gaidos et al. 2005; see also von Bloh et al. 2008; Selsis et al. 2007 in the specific context of the Gleise 581 system) the subject of planetary atmospheric habitability has been revisited little since the seminal work of Kasting et al. (1993). Given the major developments expected in the next five years and beyond, it is important that the climate regimes expected on exotic versions of the Earth, and their consequences for habitability, be studied and better understood. This will

---

[2]It should however be noted that alternate molecules (e.g., ammonia, methyl alcohol) could conceivably perform equivalent roles in different environments, such as those of lower temperature and higher pressure (Haldene 1954; Firsoff 1963; Goldsmith & Owen 2002).

help interpret upcoming planet detections and will inform future long-term efforts on the best strategies to find robust signatures of life on exo-Earths.

The central role of astronomical forcings in determining the seasons and the climatology of the Earth is well known (e.g., the long-term Milankovitch cycles). Clearly, obliquity, precession, and eccentricity can all strongly affect global and regional habitability conditions on the Earth and, by extension, on any other potentially habitable terrestrial exoplanet. Previously, only preliminary investigations of the role of obliquity (Williams & Kasting 1997, hereafter WK97; and Williams & Pollard 2003) and eccentricity (Williams & Pollard 2002), for very specific Earth-like conditions, have been considered in some detail. The surface habitability of a terrestrial planet, however, must depend on the combination of obliquity and eccentricity with the planetary rotation rate, the continental coverage (from dry Earths to water worlds) and the overall mass and composition of the atmospheric layer (e.g., compare Mars vs. Venus), among other factors. A thorough exploration of how these various global planetary attributes combine to affect climatology and make a terrestrial planet seasonally or regionally habitable is therefore an important element in the search for signatures of life elsewhere. In the present study, we describe how energy balance climate models can contribute to this understanding.

In § 5.2, we introduce the concept of climate modeling hierarchy, describe our energy balance model and validate it on the Earth. In § 5.3, we reconsider various features of habitability for Earth–like planets with seasonally-forced climates. In § 5.4, we discuss several subtleties in the definition of habitability that emerge from

our work on seasonal climates, and we finally conclude in § 5.5.

5.2 Climate Modeling Hierarchy

Current modeling tools to study the Earth's climate are constructed to be specific to Earth. For example, their radiative transfer schemes are elaborate and specific to the conditions of Earth's atmosphere and are valid only within fairly narrow ranges of atmospheric composition, temperature and pressure. Their surface boundary conditions are also specific to a given land-ocean-ice configuration. The same is essentially true of advanced climate tools used to model the atmospheres of Mars or Venus. This strongly limits the region of planetary parameter space that can be explored with these computationally intensive tools. Such models are bound to produce a somewhat limited view of what constitutes a habitable planet. They have been useful in providing some initial insights on astronomical factors affecting habitability (e.g., Williams & Pollard 2002, 2003), but given the likely detection of terrestrial exoplanets with various global attributes in the near future, it is important that tools addressing the specific needs of astronomers be developed now, to help assess the potential habitability of discovered exoplanets.

A difficult aspect of assessing the habitability of discovered terrestrial exoplanets will be that many of the key planetary attributes that determine the climate regime will remain entirely unknown to astronomers, at least in the foreseeable future. Consider, for instance, the terrestrial planets that will be discovered by *Kepler*. While their semimajor axes, their radii, and thus indirectly their masses and

surface gravities (e.g., Valencia et al. 2006, 2007; Fortney & Marley 2007; Adams et al. 2007) may be reasonably well constrained by the data – but see Seager et al. (2007) – very little additional information on the global planetary attributes will be available to help us evaluate their potential habitability. In some special cases, it may be possible to set constraints on orbital eccentricities (Barnes 2007; Ford et al. 2008), but in general, except for tidally-locked planets, the eccentricity, obliquity and planetary rotation rate will all be essentially unknown. Perhaps even more important, the atmospheric mass and composition of detected planets will be unknown. Even though an incremental amount of information on rocky exoplanets will become gradually available with future generations of space missions (i.e., orbits with *SIM-PlanetQuest* and later spectra with *TPF* and *Darwin*), this bright future does not alleviate the need for an interpretation of *Kepler*, COROT, and radial velocity exoplanet discoveries in terms of habitability.

To address this interpretational challenge, one approach is to consider the climate problem at a fundamental level, using first principles to systematically classify planets whose climate regimes and associated surface physical conditions might permit habitability. As we shall see below, an approach based on 1-dimensional Energy Balance Models (EBMs) is particularly attractive because it is computationally efficient. This approach, therefore, permits a relatively thorough exploration of the multi–dimensional parameter space of planetary properties in order to identify the most promising regions of that space for habitability.

In physical climatology (e.g., Hartmann 1994; Ghil 2002), a hierarchical fam-

ily of climate models has been built from tools of increasing complexity. At the lowest level of the hierarchy, quantitative descriptions of planetary climate start with simple, global radiative balance models. These models focus on a steady radiative equilibrium solution, associate a single surface temperature to an entire planet, and ignore important equator-to-pole dynamical atmospheric fluxes. They cannot account for the time–variable property, nor the regional property of climate and habitability. It is significant that, even though much of Earth itself is only transiently or regionally habitable by the standard liquid water criterion, calculations of habitability for astronomical applications have almost exclusively relied on global radiative balance models, of the type described in Kasting et al. (1993).[3] As a result, existing work on habitability has largely emphasized radiative and chemical issues. Tackling the equally important dynamical ones requires regional and seasonal climate models.

At the next level of the hierarchy, 1-dimensional EBMs solve a 1D time-dependent diffusion equation to specify the evolution of the surface temperature as a function of planetary latitude, based on seasonal variations in incoming and outgoing radiative fluxes and the energetic redistribution due to atmospheric motions. In the presence of a thick enough atmosphere, dynamical transport is indeed an important contributor to the local thermal budget of a terrestrial planet[4] and therefore affects its regional habitability properties. Ever since the seminal work of

---

[3]Note that global radiative models do account for 1 dimension, in the vertical, by describing the vertical radiative-convective structure of the studied atmosphere, in an average sense.

[4]Atmospheric motions result precisely from an atmosphere being locally out of radiative equilibrium, even when radiative equilibrium is globally satisfied.

Budyko (1969) and Sellers (1969) on the Earth's climate, 1D EBMs have been recognized as useful tools in physical climatology, especially for studies of variations in external (astronomical) forcing conditions and their effects on climate stability (Hartmann 1994). Except for a few preliminary investigations (e.g., WK97; Franck et al. (2000); Gaidos & Williams (2004)), however, 1D EBMs have generally not been applied to the outstanding problem of habitability on terrestrial exoplanets.

In the present work, we describe EBMs for planets that are largely similar to Earth, partly because this permits us to validate our new modeling tool on this well known and understood case. In the course of presenting our models, we will use two terms, "Earth–like" and "pseudo–Earth", to distinguish between model planets that are, within the context of our 1-dimensional energy balance framework, as similar to Earth as we can achieve with very simple heating and cooling functions ("Earth–like") and model planets that are less specifically tuned to Earth by virtue of, e.g., having a different rotation rate ("pseudo–Earth"). We investigate how the habitability of our pseudo–Earths varies with their distance from a sun–like star, and with the efficiency of the latitudinal redistribution of heat. In this study, we consider only cases with 0 orbital eccentricity and an Earth-like 23.5-degree obliquity. As illustrated below, we already find that the issue of habitability for close Earth analogs becomes a rich one when addressed with a 1D time-dependent EBM.

5.2.1   Global Radiative Balance

It is instructive to start our investigation of habitable climates with a discussion of global radiative balance results. Assuming that the annual mean surface temperature, T, on an Earth–like planet is determined by radiative balance between energy sources and sinks, the following steady-state radiative equilibrium equation must be satisfied,

$$I[T] = S(1 - A[T]),\qquad\qquad(5.1)$$

where I is the infrared cooling flux, S is the annual mean stellar insolation flux and A is the global planetary albedo.

In this work, we consider three different formulations for the cooling and albedo functions. Compared to previous work on global radiative balance (e.g. Kasting et al. 1993), our treatment is extremely simple. Nevertheless, it accounts for the key physical elements involved in the absorption and reemission of incident stellar flux and is sufficient for our EBM work, which is focused on dynamical climate issues rather than radiative transfer ones.

In a planetary atmosphere, the greenhouse effect acts to reduce the infrared cooling flux, $I[T]$, at a given value of the surface temperature, T. Kasting (1988) and Kasting et al. (1993), for instance, describe in detail how the Earth's greenhouse effect scales with surface temperature. An important result is that higher temperatures cause greater humidity, which in turn leads to stronger atmospheric heat retention via the greenhouse effect from water vapor absorption (see also WK97). Our purpose here is not to accurately reproduce Earth's climate but rather to inves-

tigate dynamic climate issues for Earth-like planets. We therefore consider three different models for the infrared cooling function, I, and the albedo function, A, as listed in Table 5.1. While each of these models reproduces reasonably well the current climate of the Earth, (see § 5.2.2.2), they could each respond differently once we start exploring physical conditions that are different from the current Earth. These three atmosphere models are therefore useful in evaluating the sensitivity of our main results on habitable climates to detailed model assumptions for the infrared cooling and albedo functions.

Models 1 and 2 have infrared cooling functions inspired from a one-zone radiative transfer formulation that assumes ground-level blackbody emission (σT^4), effectively reduced at the atmospheric "photosphere" according to a near-Eddington approximation (inspired by Shu (1982)), for a given atmospheric optical thickness to infrared radiation, τ_{IR} (as shown in Table 5.1). Model 3, on the other hand, adopts the standard linear cooling function of North & Coakley (1979), which can be interpreted as a linearization of IR–cooling around the globally-averaged conditions for the Earth. It is in close agreement with other T-linearized IR models used in the geophysical literature (e.g., Budyko 1969; North et al. 1981). In our model 1, the IR optical thickness is fixed to unity, while in model 2, a cubic dependence on surface temperature, T, is adopted, to better match the linear model results (model 3) for Earth-equivalent conditions.

In each of our three atmospheric models, the albedo function, $A[T]$, is chosen to capture a rapid ice-water transition, with values of $A \sim 0.7$–0.77 well below 263 K

Table 5.1 Atmospheric Models.

Model	IR Cooling Function	Albedo Function
1[a]	$I_1[T] = \frac{\sigma T^4}{1+(3/4)\tau_{IR}^0}$	$A_1[T] = 0.5 - 0.2\tanh[\frac{(T-268\mathrm{K})}{5\mathrm{K}}]$
2[b]	$I_2[T] = \frac{\sigma T^4}{1+(3/4)\tau_{IR}[T]}$	$A_2[T] = 0.525 - 0.245\tanh[\frac{(T-268\mathrm{K})}{5\mathrm{K}}]$
3[c]	$I_3[T] = A + BT$	$A_3[T] = 0.475 - 0.225\tanh[\frac{(T-268\mathrm{K})}{5\mathrm{K}}]$

[a] Model with fixed optical thickness: $\tau_{IR}^0 = 1$.

[b] Model with T–dependent optical thickness: $\tau_{IR}[T] = 0.79(T/273\mathrm{K})^3$.

[c] Linearized model: $A = 2.033 \times 10^5$ erg cm^{-2} s^{-1}, $B = 2.094 \times 10^3$ erg cm^{-2} s^{-1} K^{-1}.

Note. — σ is the Stefan-Boltzmann constant.

(ice–covered surface), and values ~ 0.25–0.3 well above 273 K (ice–free surface). To avoid albedo discontinuities, which are known to result in spurious "small ice-cap" instabilities in EBMs (e.g., Held et al. (1981)), the albedo transition is performed smoothly over the range 263–273 K with a hyperbolic tangent formulation. This type of albedo prescription is standard in the context of EBMs (e.g. WK97). The slightly different asymptotic values adopted for the albedos in our three models (see Table 5.1) were adjusted to best reproduce the annual mean climate of the Earth. This is common practice with simplified EBMs which do not account for the detailed land-surface conditions that contribute to the global planetary albedo (see § 5.2.2.1 for a discussion of EBM limitations).

In what follows, we adopt model 2 as our fiducial model but we also show a few comparison results with models 1 and 3. Quite generally, we find that predictions from our three atmospheric models differ only modestly at a quantitative level. With a choice of IR-cooling and albedo functions, steady-state radiative equilibrium solutions for the annual mean climate are obtained by solving Eq. (5.1) for a specified value of the annual mean insolation flux, S, received by the planet. For a fast-spinning planet, such that a diurnal average is justified, on a circular orbit at a distance a (in AU) from a Sun-like star, the annual mean insolation is simply $S = (1/4)\,S_0/a^2$ (e.g. Rubincam 2004), where the solar constant $S_0 = 1.36 \times 10^6$ erg cm^{-2} s^{-1}.

With these assumptions, Eq. (5.1) reduces to the thermal balance equation $Q^- = Q^+$, where the annual mean cooling rate, $Q^- = I[T]$, and heating rate, $Q^+ =$
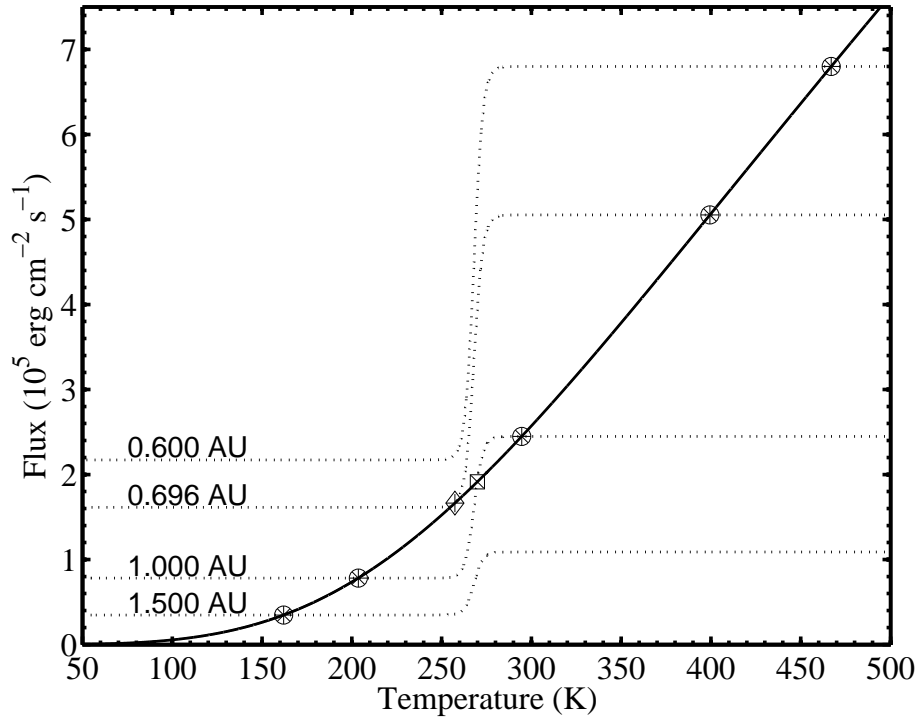
Figure 5.1 Heating and cooling fluxes, as a function of surface temperature, in a global radiative balance model using the IR-cooling and albedo functions listed as model 2 in Table 5.1. The solid line shows the IR-cooling flux while dotted lines show albedo-reduced, annual mean insolation fluxes at four different orbital distances from a Sun–like star (from 0.6 AU to 1.5 AU). Stable radiative equilibrium solutions are highlighted with circles. Other solutions are thermally unstable (see text for details). In the vicinity of ~ 1 AU, this class of global radiative balance models generally admits two stable solutions for the climate; one is ice-free and the other is ice-covered.

$(1-A[T])S$, are expressed per unit surface area of the planet. Due to the steep albedo variation with surface temperature, relative to the smooth IR-cooling dependence, it is possible for multiple climate equilibrium solutions to satisfy Eq. (5.1). Figure 5.1 illustrates this possibility, using our model 2 for definiteness. The cooling and heating rates, Q^- and Q^+, are shown as solid and dotted lines, respectively, for a wide range of surface temperatures. The heating rate is shown at four different orbital distances from the central Sun-like star, from 0.6 to 1.5 AU. The strong effect of the albedo transition on the heating rate is easily identified at $T \simeq 263–273$ K.

While unique climate equilibrium solutions satisfy Eq. (5.1) for $a = 0.6$ and 1.5 AU, two such solutions are found at $a = 0.696$ AU and three solutions are found at $a = 1$ AU, as shown by the intersections of dotted and solid lines in Fig. 5.1. Solutions to a thermal balance equation such as Eq. (5.1) are stable only if small increases in temperature cause cooling to exceed heating and small decreases cause heating to exceed cooling:

$$\frac{d\left(Q^+ - Q^-\right)}{dT} < 0, \tag{5.2}$$

If the above equation is satisfied, then any small surface temperature perturbation relaxes back to the starting equilibrium solution. Figure 5.1, then, shows that solutions found in the steep part of the albedo transition ($T \simeq 263–273$ K) are generally unstable, given the relatively slow increase of the IR-cooling flux with surface temperature. The intermediate temperature solution in the case $a = 1$ AU is therefore unstable and the lower temperature solution in the marginal case $a = 0.696$ AU is stable only on one side (stable against reductions in temperature

but not against increases).

At small or large enough orbital distances, a unique stable solution exists, corresponding to either an ice-free or an ice-covered climate. At intermediate orbital distances (~ 1 AU), however, both the ice-free and the ice-covered climate solutions are valid and stable. This bi-stable property of Earth's climate in simple global radiative balance models is well-known and it may be related to the snowball (globally ice-covered) events for which there is some evidence in the past history of Earth's climate (see the review by Hoffman & Schrag (2002) and references therein). We also note that these ice-covered equilibrium solutions correspond to the "cold-start" scenarios briefly mentioned by Kasting et al. (1993) in their detailed study of habitability with global radiative balance models.

Globally ice-covered climate solutions are often omitted from studies of planetary habitability because they violate the standard surface liquid water requirement for habitability. Yet, Earth's climate itself may have experienced one or two such global glaciation events and recovered from them without catastrophic damage to life on the planet, or at least without eradicating life from the planet (Hoffman et al. 1998; Hoffman & Schrag 2000; Baum & Crowley 2003; Hoffman & Schrag 2002). The fact that the current Earth is partially ice–covered, depending on the seasons, reveals another obvious shortcoming of global radiative balance calculations for habitability: an Earth-like planet can be habitable even if it is regionally and/or transiently ice-covered. This important class of climate regime is simply not addressed by global radiative balance models admitting only fully ice-covered

or fully ice-free stable solutions, as illustrated in Fig. 5.1. The present work on seasonally habitable climates is motivated precisely by the possibility to explore these more complex climate regimes with a simple class of climate models.

5.2.2 1–D Energy Balance Model

Issues of seasonal climate dynamics and susceptibility to climate transitions are not addressed by steady–state radiative balance models that consider only annually averaged global mean conditions. In principle, one can solve for the instantaneous radiative equilibrium conditions either locally or globally on an Earth–like planet. A planet satisfying radiative equilibrium globally, however, does not necessarily satisfy it locally, for two main reasons: First, atmospheric motions tend to transport heat from the hotter regions to the colder regions, effectively coupling the thermal states of various locations on the planet beyond a simple state of radiative balance. Second, the atmosphere has thermal inertia and does not immediately adjust its temperature to the local forcing. As a result, local radiative equilibrium solutions will generally not be accurate representations of local surface temperature conditions. Furthermore, because of this thermal inertia, planets do not necessarily satisfy even global radiative equilibrium, especially when on eccentric orbits. These important physical processes can be incorporated in a time–dependent EBM that models the seasonal climate as a latitudinal diffusion of atmospheric heat under prescribed astronomical forcing.

A justification of the equivalent diffusion equation satisfied by an atmospheric

layer, obtained by averaging the second law of thermodynamics and the continuity equation, can be found in Lorenz (1979). Lorenz (1979) also argues that the diffusion approximation is justified for the Earth, on large enough scales for the atmosphere to be considered a forced system. Following WK97 and previous work in the geophysical literature, we adopt the following prognostic diffusion equation for the planetary surface temperature,

$$C\frac{\partial T[x,t]}{\partial t} - \frac{\partial}{\partial x}\left(D(1-x^2)\frac{\partial T[x,t]}{\partial x}\right) + I[T] = S(1 - A[T]), \tag{5.3}$$

where all quantities are local functions, $x \equiv \sin\lambda$ is the sine of the latitude λ, T is the surface temperature, C is the effective heat capacity of the atmosphere (different over land, ocean, and ice), D is the diffusion coefficient that determines the efficiency of latitudinal redistribution of heat, I is the same IR-cooling flux as in §5.2.1, S is the diurnally-averaged insolation flux and A is the same albedo function as in §5.2.1. In the above equation, $C, D, I,$ and A are functions of $T, x, t,$ and possibly other relevant planetary parameters. The $(1 - x^2)$ factor multiplying the diffusion coefficient in Eq. (5.3) is a metric factor that arises from representing the Laplacian in spherical coordinates (see, e.g., North et al. 1983).

Our prescriptions for C and D are borrowed from WK97 and the existing geophysical literature on 1D EBMs. For simplicity and flexibility, we use simple, physically motivated prescriptions. For the surface heat capacity C, we assume a uniform ocean fraction of 70% at every latitude and use the same partial heat capacities over land, ocean, and ice as WK97: $C_l = 5.25 \times 10^9 \, \text{erg cm}^{-2} \, \text{K}^{-1}$ over

land, $C_o = 40\,C_l$ above the wind-mixed surface layer of the ocean, $C_i = 9.2\,C_l$ over ice when $263 < T < 273$ K, and $C_i = 2.0\,C_l$ for $T \leq 263$ K (see WK97 for details). For the diffusion coefficient D, we adopt a fiducial value that is adjusted to reproduce the current climate of Earth and is 93% of that used by WK97: $D_{\text{fid}} = 5.394 \times 10^2$ erg cm^{-2} s^{-1} K^{-1}. For the diurnally-averaged insolation flux S, we use the same standard formalism as WK97 (see their Appendix). Throughout this work, we adopt an Earth-like planetary obliquity, $i = 23.5°$, and assume zero eccentricity of the planetary orbit for simplicity.

Equation (5.3) is solved on a grid uniformly spaced in latitude with a highly efficient time-implicit numerical scheme and an adaptive time–step, as described in Hameury et al. (1998). We impose the boundary conditions $dT/d\lambda = 0$ at $\lambda = \pm 90°$ for symmetrical solutions. Resolution tests, which are further described in Appendix I, show excellent numerical convergence at our fiducial resolution corresponding to 1.25° in latitude (145 grid points pole–to–pole). All models are initiated with a uniform temperature corresponding to an ice-free planet. To guarantee satisfactory relaxation to a periodically-forced, seasonal climate regime that is unaffected by details in initial conditions, all models are integrated for at least 130 years, well in excess of the thermal-inertial timescale of about a decade of the modeled atmospheric layer. As shown in Appendix II, specific choices of initial conditions can potentially influence long-term model outcomes by inducing a dynamical transition to a fully ice-covered (snowball) climate even when ice-free initial conditions are specified. We systematically initialize our models with temperatures well above

freezing (e.g., $T \geq 350$ K for Earth-like planets), which guarantees that partially ice-covered climates are favored over fully ice-covered snowball states. This is similar to excluding "cold start" scenarios in global radiative balance studies of habitability (e.g., Kasting et al. 1993).

To summarize, we make the following assumptions in the models presented here:

1. *Heating/Cooling* – The heating and cooling functions are given by the diurnally averaged insolation from a sun–like (1 $M_\odot$, 1 $L_\odot$) star, with albedo and insolation functions given in Table 5.1.

2. *Latitudinal Heat Transport* – We test the influence on climate of three different efficiencies of latitudinal heat transport, within the diffusion equation approximation: an Earth-like diffusion coefficient, and a diffusion coefficient scaled down and up by a factor of 9 (which correspond to 8-hour and 72-hour rotation according to the $D \propto \Omega_p{}^{-2}$ scaling described in § 5.3.1.1 below).

3. *Ocean Coverage* – Every model has the same land–fraction in every latitude band. For most of the models presented in this paper, each latitude band has an Earth–like 30%:70% land:ocean ratio. We also present a series of models that represent a desert-world, with a 90%:10% land:ocean ratio.

4. *Initial Conditions* – The models all have a hot-start initial condition, with the planet's temperature uniform and at least 350 K, to minimize the chances of ending up in a global snowball state owing solely to the choice of initial

conditions (see Appendix II for more details). Time begins at the northern winter solstice.

5.2.2.1 Model Limitations

Climate systems involve interactions over a wide range of time and length scales, so that any given model (even a modern General Circulation Model – GCM) only emphasizes a limited range of scales. This is particularly true of the simple 1D EBMs considered here. Before presenting results from EBM calculations, it is therefore important to clarify what the main limitations of these models are.

One type of limitation is related to the low spatial dimensionality of our 1D EBM. Our latitudinal EBM includes only a diurnally-averaged formulation of insolation. While this is acceptable for a planet like the Earth, with a relatively fast spin rate and significant overall thermal inertia from a large ocean fraction, this simplification will not be generally valid. In addition, since the models do not resolve planetary longitudes, the surface temperature at a given latitude must be interpreted as an average over the corresponding latitude circle. This can lead to some ambiguities, in the sense that a 2-dimensional model resolving the longitudes would generally find different physical conditions to be present over land vs. oceanic regions. This land/ocean distinction, which is generally a function of latitude, like on Earth, is not well addressed by our longitudinally-averaged 1D EBMs.

A second type of limitation is related to the restricted range of timescales

captured by our models. Our work focuses on the atmospheric response to a seasonal forcing and therefore emphasizes climate features emerging typically on orbital timescales. As a result, it omits many of the slower processes involved in determining the long-term climate on a planet like the Earth. For example, oceans are only treated as partial heat reservoirs in our models, thereby neglecting their circulatory effects on longer timescales. Similarly, the role of the carbonate–silicate cycle in regulating the atmospheric CO_2 composition (e.g., Kasting et al. 1993; WK97) and the resulting effects of atmospheric CO_2 on both cloud–formation and the greenhouse effect are entirely neglected in our EBMs. In particular, on timescales longer than considered in our models, it is expected that the type of CO_2 warming discussed by Kasting et al. (1993), Forget & Pierrehumbert (1997), and Mischna et al. (2000) would extend the outer boundary of the habitable zone. A massive release of CO_2 in Earth's atmosphere, via an asymmetric carbon-silicate cycle for millions of years, is also the leading candidate scenarios invoked to explain how the paleoclimate was ultimately able to escape a globally-frozen snowball state (Hoffman & Schrag 2002). On the much shorter timescales described by our models, however, snowball events are simply semi-permanent states of the climate.

Finally, we note that there are numerous additional atmospheric processes that are ignored from our modeling strategy. They include, for instance, feedback effects from clouds on the albedo and IR-cooling functions or the possibility of transient oceanic evaporation and partial atmospheric escape in the regionally hottest models considered. Despite these important shortcomings, our EBMs are

useful to explore many important dynamical climate features possibly relevant to planetary habitability.

5.2.2.2   Model Validation

In what follows, the model we refer to as the "fiducial" one is the Earth–like EBM built from the IR-cooling and albedo functions listed as model 2 in Table 5.1, using a uniform 70% ocean fraction, a latitudial heat diffusion constant $D = D_{\text{fid}}$ and "hot-start" (ice-free) initial conditions. We find very good agreement between the latitudinal surface temperature distribution predicted by this fiducial model and the one observed on Earth. The National Center for Environmental Prediction (NCEP), in conjunction with the National Center for Atmospheric Research (NCAR), has released 6–hourly global temperature maps for Earth that are the results of a detailed global climate model that is tightly constrained by observed temperatures (Kistler et al. 1999; Kalnay et al. 1996). For model validation, we use these data for model validation and the phenomenalogical fit to Earth's annual mean latitudinal temperature profile proposed by North & Coakley (1979):

$$T[\lambda] = 302.3 \, \text{K} - 45.3 \sin^2[\lambda]. \qquad (5.4)$$

Figure 5.2 shows the annual mean temperature distribution in our fiducial model (solid line) after 130 years of thermal relaxation. It is within 5 K of the phenomenalogical fit in Eq. (5.4), shown as a dotted line, at all latitudes. The model also reproduces quite well the zonally and temporally averaged surface

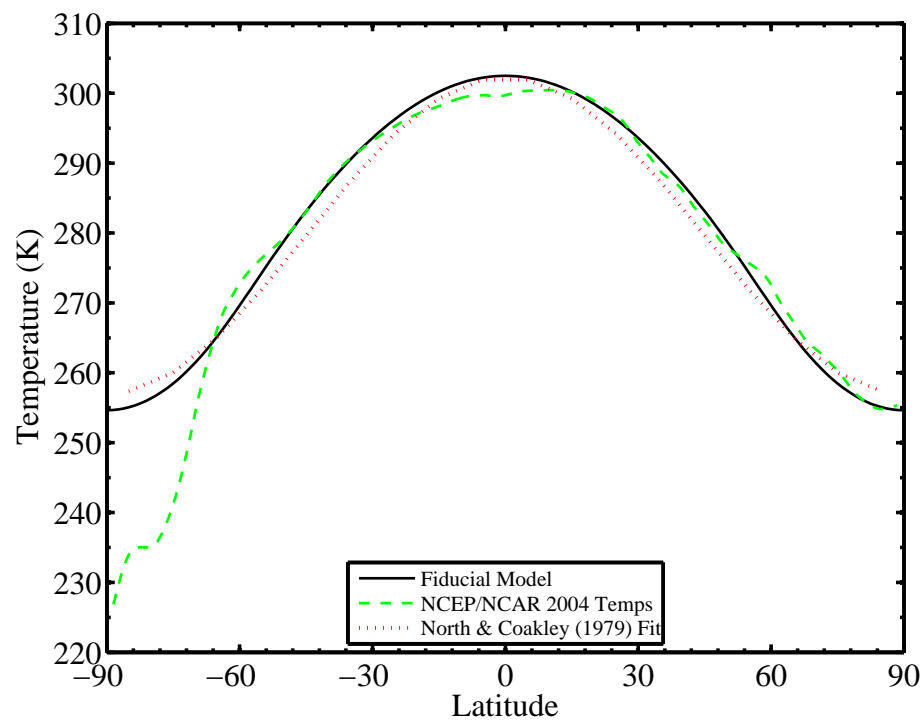Figure 5.2 Model validation based on the mean latitudinal temperature profile of the Earth. The solid line shows the annual mean temperature profile in our fiducial Earth-like model. The dashed line shows the average taken from the NCEP/NCAR global temperature data in 2004. The dotted line shows the fit to Earth's mean temperature profile proposed by North & Coakley (1979). The discrepancy at the South Pole is due to Antartica.

temperatures from NCEP/NCAR for the year 2004[5] at all latitudes north of 60° S. The North–South asymmetry of the land–distribution on Earth, due to the presence of a continental circumpolar region – Antarctica – in the south, breaks the symmetry of the temperature profile.[6] It is therefore not surprising that our simple, symmetrical EBM fails to reproduce the sharply colder temperatures near the South Pole.

In addition to reproducing annual mean properties, the fiducial model also captures important seasonal variations seen in temperature data from the Earth or in advanced GCMs of the Earth. In particular, the temperature response is found to lag the solar forcing in our model by about a quarter annual cycle: the warmest temperatures happen at the local autumnal equinox and the coolest temperatuers at the local vernal equinox. This well known phase lag is the result of the high heat capacity of the wind-mixed ocean layer (e.g., North & Coakley (1979)), which covers 70% of our model Earth-like planet. For detailed comparison, we obtained surface layer outputs from the full three–dimensional MIT Oceanic GCM (Wunsch & Hiembach 2007). Figure 5.3 shows the temperature as a function of time of year at latitudes from the equator to ±60°, in our model and the uppermost ocean layer of the MIT GCM (5 m depth) for the year 2004. Time of year is measured in fraction of a year from the northern winter solstice. Our model is symmetrical with respect to the equator, so there is no real distinction between North and South, but the MIT GCM is not. Both models show the same response–lag of ~ 90°. Towards the poles, our model gets significantly colder than the (ocean only) GCM, but at equatorial

---

[5]The match is good for other years as well, but only the 2004 data product is shown here.
[6]The figure illustrates the potential magnitude of effects due specific land/ocean configurations.
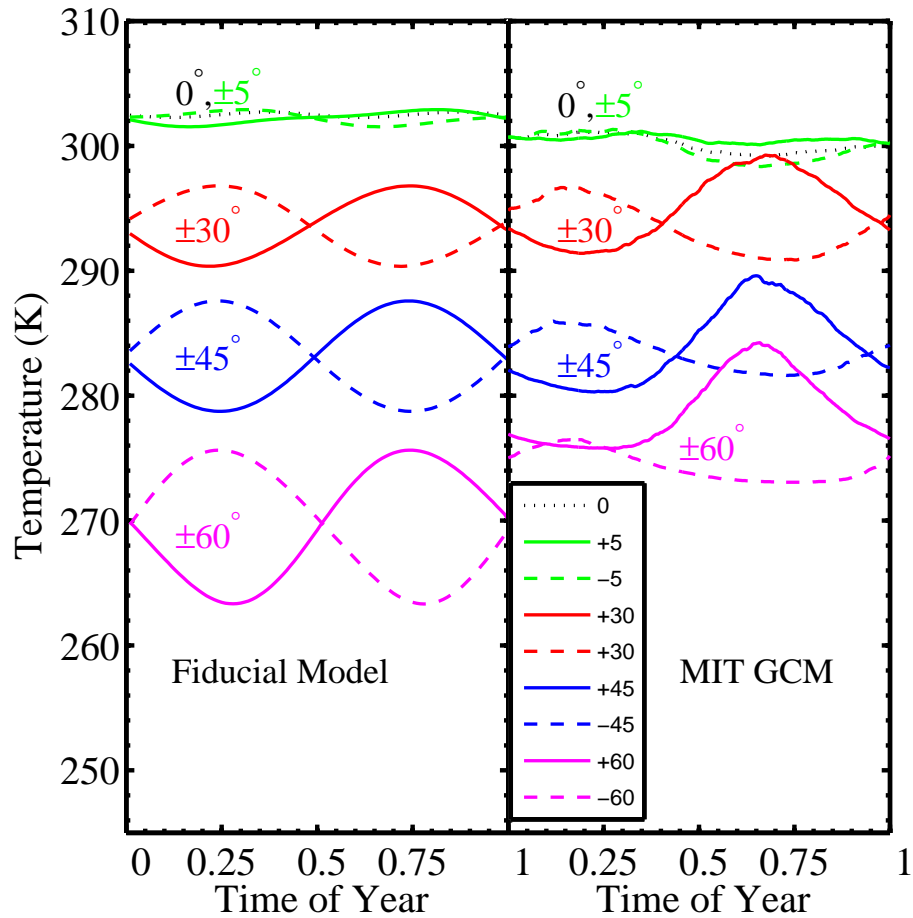
Figure 5.3 Model validation based on seasonality at different latitudes. The left panel shows temperatures as a function of time for several latitudes in our fiducial Earth-like model. The right panel shows the 2004 ocean–surface temperatures in the MIT Oceanic GCM (Wunsch & Hiembach 2007) for the same latitudes. In both panels, solid lines correspond to temperatures in the North and dashed lines to temperatures in the South. Time of year is measured in fraction of a year from the northern winter solstice. Note that our model is North–South symmetrical while the MIT Oceanic GCM is not. Extrema of temperature occur approximately one quarter cycle after the extrema in forcing, as expected for a seasonally-forced planet in the limit of large thermal inertia (e.g., North & Coakley 1979).

and mid–latitudes, remarkably, the deviations are in most places significantly less than 5 K.

Although our fiducial model is quite successful in reproducing various seasonal features of Earth's temperature distribution, we caution that this alone does not guarantee that predictions will remain accurate for parameters other than the fiducial ones (e.g., for changes in orbital distance leading to systematically hotter or colder climates). To achieve a reasonably good fit to the Earth, various parameters in the model were adjusted (e.g., constants in model 2 of Table 5.1 or value of D_{fid}). Since the model is based on parameterizations, rather than fundamental laws, it is possible that its behavior will deviate quantitatively from what Earth's response would be when new regimes of climate are explored. Nevertheless, the detailed model validation performed here with the Earth and the physics-based prescriptions adopted in our modeling strategy should guarantee that our EBM predictions are qualitatively robust. As we shall see below, this notion is supported by the generally good agreement of results obtained from the three different sets of atmospheric IR-cooling and albedo models listed in Table 5.1.

5.3   Study of Habitability

Equipped with a well-tested and validated EBM, we now reconsider several issues related to the habitability of Earth-like planets. In the present study, we emphasize differences between EBM and global radiative balance calculations, focusing on the concept of fractional habitability, and we highlight the potentially important role

played by climate stability against snowball events in determining the habitability of an Earth-like planet.

5.3.1 Climate Dynamics and Seasonality

5.3.1.1 Climate Stability

Even a simple EBM such as ours can exhibit a surprisingly rich climate phenomenology, as compared to the outcomes of global radiative balance models. A particularly striking example of this phenomenology is the susceptibility of the climate system to albedo feedback effects that can induce transitions from partially ice–covered states to globally-frozen snowball states. In the context of habitability studies, these transitions are important because they could render a planet essentially "uninhabitable," unless transitions back to partially ice–covered or ice–free states are possible. Since the seminal work of Budyko (1969) and Sellers (1969), it has been well known that Earth's partially ice-covered climate allows for transitions to globally-frozen snowball states in response to relatively minor changes in forcings (e.g., Hartmann (1994); Ghil (2002); Hoffman & Schrag (2002)).

Climate stability is a subtle issue, even in the context of a 1D diffusion model. Held et al. (1981) describe how it is ultimately an interplay between latitudinal heat transport and albedo feedback effects that determines climate stability. Here, we will not attempt a thorough exploration of climate stability with our EBM but instead will illustrate how this issue could be critical in determining the habitability of a seasonally-forced Earth-like planet.

168

Latitudinal heat transport by atmospheric motions is inhibited by Coriolis effects on a rotating planet (e.g., Pedlosky (1982); Holton (1992)). As a result, latitudinal heat transport should be typically reduced on a pseudo-Earth rotating faster than the current Earth. There is currently no first principle theory allowing reliable estimates of how much transport would be reduced on a faster rotating Earth-like planet. There have been, however, simple scaling arguments proposed in the geophysical literature. Here, in the interest of simplicity, we will use the "thermal Rossby number" scaling advocated by Farrell (1990), which suggests that D is $\propto \Omega_p^{-2}$, where Ω_p is the angular rotation rate of the planet (see also Stone 1973). We note that detailed GCM experiments for slow-rotating Earths do not necessarily support the extreme simplicity of the above scaling (del Genio et al. 1993; del Genio & Zhou 1996).

Figure 5.4 shows space-time diagrams of temperature in two Earth-like models differing only in the magnitude of latitudinal heat transport. The top two panels show temperatures for the full 130 years (upper left) and for the last 5 years (upper right) of our fiducial Earth-like model. The bottom two panels show temperatures, over the same time spans, in a model of a pseudo-Earth that is in all respects identical to our fiducial model, except for a diffusion coefficient D reduced by a factor of 9 (which, according to the above scaling, corresponds to a planet rotating three times as fast as Earth does[7]). Both models start with a uniform $T = 350\,\mathrm{K}$, but their evolutions follow strikingly different trajectories. The fiducial model settles to

[7]We note that it has been suggested that the young Earth may have spun four times as fast as it currently does (Canup 2004).
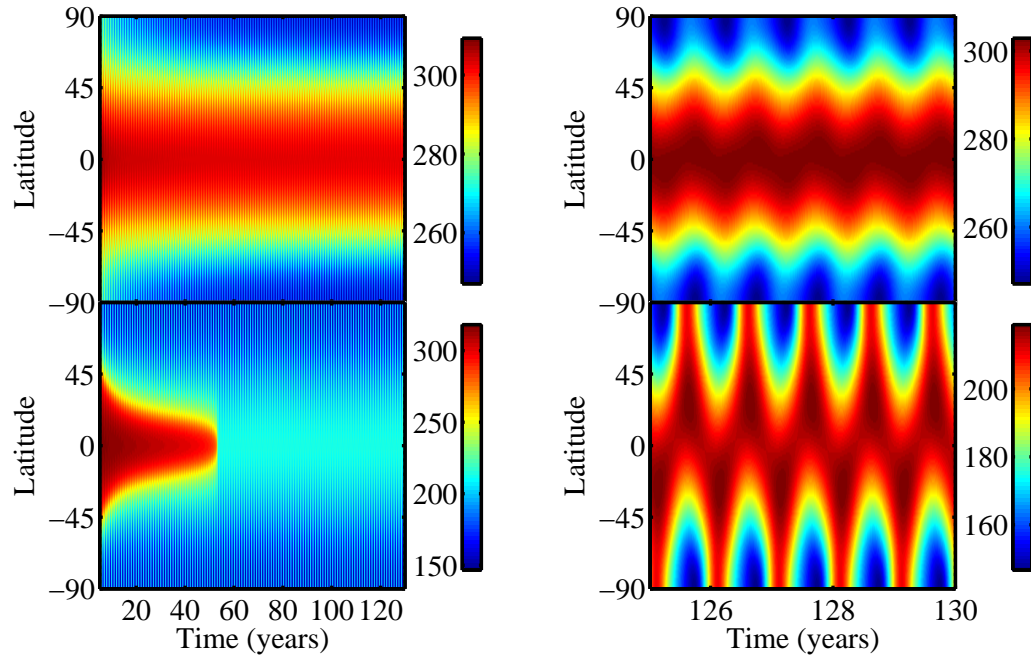
Figure 5.4 Space-time diagrams of temperature (in K) for two different pseudo-Earth models. Left panels show the entire 130 years of evolution while the right panels show only the last 5 years of seasonal climate. The top two panels show the evolution of temperatures in our fiducial Earth-like model. The bottom two panels show the evolution of temperatures in a pseudo-Earth model that is identical to the fiducial model, except for a coefficient of latitudinal heat transport reduced by a factor of 9 from the fiducial value ($D_{\mathrm{fid}}/9$). A reduced efficiency of latitudinal heat transport is expected on fast-rotating pseudo-Earths. The model with inefficient latitudinal heat transport makes a dynamical transition to a globally-frozen snowball state after approximately 50 years.

the standard, partially ice-covered solution. In contrast, the model with inefficient latitudinal heat transport undergoes rapid evolution through the first ~ 50 years, and then abruptly, over the course of a single year, transitions to a permanent snowball state with a mean temperature around 200 K.

Note that in the Earth–like fiducial model (upper right panel), the poles are, unsurprisingly, the coldest parts of the planet. By contrast, in the fast-Earth model (lower right panel), the poles become nearly as warm during local summer as the warmest parts of the planet. This is because, with our albedo function (model 2 in Table 5.1), the entire planet has identical albedo in the snowball state, and so the poles, which actually experience greater diurnally averaged insolation during local summer than any other part of the planet, can warm a great deal in response to the high solar irradiance. In the fiducial Earth-like model, on the other hand, parts of the planet that are above freezing absorb significantly more solar radiation, owing to their lesser albedo, than parts that are well below freezing; this is what keeps the poles relatively cold year–round in the fiducial model.

The main conclusion to be drawn from this comparison between our fiducial Earth-like model and a fast-rotating pseudo-Earth model is that issues of climate dynamics, especially climate stability with respect to snowball transitions, can critically influence the habitability of Pseudo–Earth planets. While the planetary rotation rate is simply inconsequential for global radiative balance models of habitability, it becomes a potentially important planetary attribute for habitability in more elaborate dynamical climate models, even for Earth-like planets at 1 AU from

a Sun-like star, as was assumed in both models shown in Figure 5.4.

5.3.1.2   Comparison to Global Radiative Balance Results

We now relax our 1 AU assumption and calculate EBMs for Earth-like planets at a range of orbital distances from their Sun-like star. We place our model planets at 41 distances from the star, in the range 0.3 to 1.3 AU, with 0.025 AU steps. Each model is run for 130 local years (orbits), which is sufficient for adequate model relaxation from the hot-start to the final periodic climate solution. At every distance, we also calculate the radiative equilibrium temperature that obtains from the global average insolation, so that we can directly compare our time-dependent EBM results to those of an equivalent global radiative balance model.

As discussed in § 5.2.1, there is in general a range of orbital distances around ~ 1 AU for which multiple radiative equilibrium solutions exist. We focus exclusively on the two stable, low and high temperature solutions here. Each panel in Figure 5.5 compares the radiative equilibrium solutions (dotted and dashed-dotted curves) to results for the annual mean temperature weighted by surface area in our EBM calculations (solid line), for a range of orbital distances. From left to right, the three columns show results using models 1, 2, and 3 in Table 5.1, respectively, for the IR-cooling and albedo functions. In each column, EBM results are shown for three efficiencies of latitudinal heat transport, corresponding to $D = 9\,D_{\mathrm{fid}}$, D_{fid}, and $D_{\mathrm{fid}}/9$, from top to bottom. Our fiducial Earth-like model corresponds to the central panel.
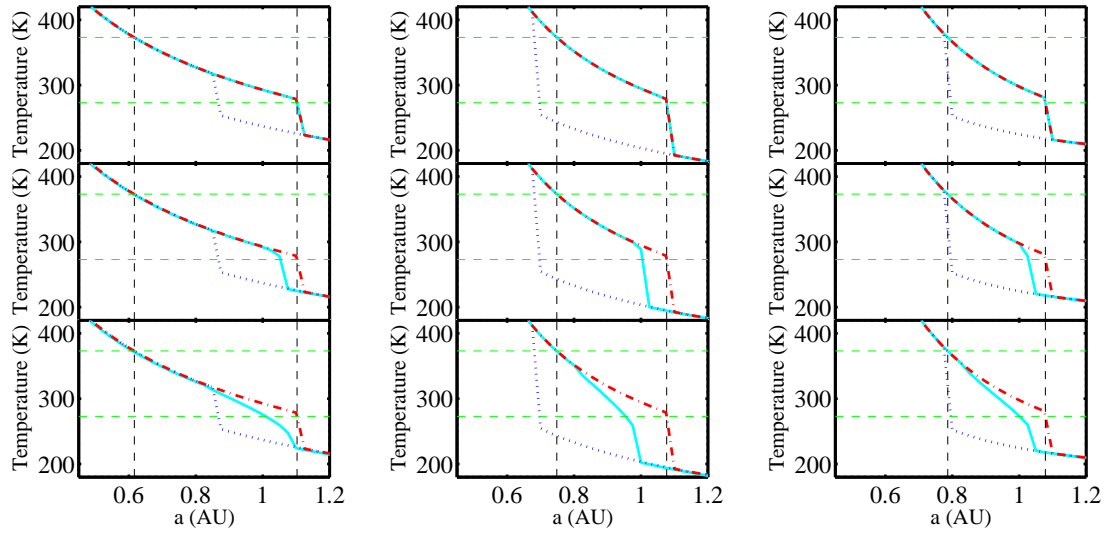
Figure 5.5 Comparison between global radiative equilibrium temperatures and averaged temperatures in our seasonal EBM, for nine different pseudo-Earth models. From left to right, the three columns show results obtained using models 1, 2 and 3 in Table 5.1 for the IR-cooling and albedo functions, respectively. In each column, from top to bottom, panels correspond to values $D = 9\,D_{\text{fid}}$, D_{fid}, and $D_{\text{fid}}/9$ for the latitudinal diffusion coefficient used in the EBM. In each panel, the solid curve shows the global and annual mean temperature in the EBM over a range of orbital distances. The dotted and dashed-dotted curves show the corresponding low- and high-temperature radiative equilibrium solutions, respectively. The mean temperature in the EBM deviates from the high-temperature radiative equilibrium solution for small and moderate values of the coefficient of latitudinal heat diffusion.

At small and large orbital separations, the EBM and radiative-balance curves nearly coincide with each other, indicating that the unique radiative equilibrium solution is very close to the averaged temperature in our seasonal EBM. At intermediate separations corresponding to partially ice-covered climates, however, the curves separate. The distinction among the three curves is particularly important for EBMs with weak latitudinal heat transport (bottom panels) because EBMs with efficient latitudinal heat transport (top panels) have generally more uniform temperature distributions, which is close to the implicit averaging assumption made in global radiative balance models.

Each panel in Figure 5.5 has horizontal dashed lines at 273 K (water freezing) and at 373 K (water boiling) and vertical dashed lines indicating where the high temperature radiative equilibrium solution crosses these 273 K and 373 K limits. In other words, the vertical dashed lines indicate the orbital extent of the radiative equilibrium habitable zone (hereafter, REHZ) for globally averaged conditions, analogous to the calculations in Kasting et al. (1993). The extent of the habitable zone defined on the basis of the globally and annually averaged temperature in our EBMs is the range of orbital radii, a, for which the solid curve remains between the two horizontal dashed lines (indicating surface liquid water conditions for an Earth-like planet). In all the models considered here, this EBM-average habitable zone tends to have nearly the same inner boundary as the REHZ, but its outer boundary can be significantly closer to the Sun-like star than that of the REHZ. Indeed, in EBMs with fiducial or reduced efficiencies of latitudinal heat transport,

the climate typically makes dynamical transitions to a fully ice-covered snowball state at orbital distances well inside of what global radiative balance models might indicate. Although quantitative discrepancies exist, this qualitative behavior is common to all three models with different IR-cooling and albedo functions shown in Fig. 5.5. Note that the dynamical transition to a snowball state shown in Fig. 5.4 for a fast-rotating pseudo-Earth at 1 AU is included in the lower middle panel of Fig. 5.5.

While useful for comparisons between EBM and global radiative balance results, the diagrams shown in Fig. 5.5 do not accurately capture the range of surface temperature conditions found in our seasonal EBMs. Indeed, for every globally averaged annual mean temperature from an EBM, there must exist a range of temperatures above and below this average, both in a regional and a temporal sense. In general, planets can therefore be habitable beyond what globally averaged studies might indicate. This leads us to give a formal definition of the notion of fractional habitability.

5.3.2 Fractional Habitability

For a fixed set of forcing and response parameters, a planet might be habitable (i.e., have a surface temperature between 273 K and 373 K) over only a portion of its total surface area or for only a fraction of its orbit. To quantify the various possible outcomes in our EBMs, we develop several metrics of fractional habitability.

Let $H[a, \lambda, t]$ be the "habitability function", equal to 1 if latitude λ, on a model

planet at a distance a from its star, has a habitable temperature at time t, and 0 otherwise:

$$H[a, \lambda, t] = \begin{cases} 1 & \text{if } 273 \le T[a, \lambda, t] \le 373, \\ 0 & \text{otherwise.} \end{cases} \qquad (5.5)$$

Using H, we define three habitability fractions.

At each latitude, we calculate the fraction of the year that that portion of the planet spends in the habitable temperature range,

$$f_{\text{time}}[a, \lambda] = \frac{\int_{t=0}^{P} H[a, \lambda, t]dt}{P}, \qquad (5.6)$$

where f_{time} is the habitable fraction as a function of latitude, and P is the length of the year at orbital separation a. At each time of year, we calculate the fraction of the planet's surface area that is habitable:

$$f_{\text{area}}[a, t] = \frac{\int_{\lambda=-\pi/2}^{\pi/2} H[a, \lambda, t] \cos[\lambda]d\lambda}{2}. \qquad (5.7)$$

Finally, at each orbital separation, we calculate the net fractional habitability, which is the area–weighted fraction of the $\lambda - t$ plane over which the planet is habitable:

$$f_{\text{hab}}[a] = \frac{\int_{\lambda=-\pi/2}^{\pi/2} \left\{ \int_{t=0}^{P} H[a, \lambda, t]dt \right\} \cos[\lambda]d\lambda}{2P}. \qquad (5.8)$$
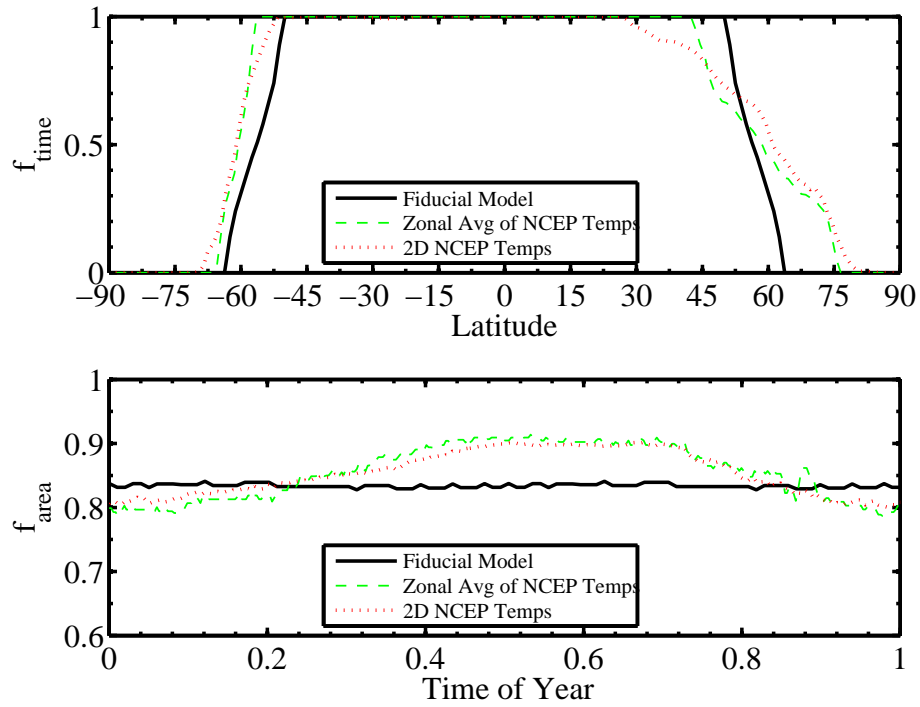
Figure 5.6 Temporal and regional habitability fractions of the Earth and in our fiducial Earth-like model. The top panel shows f_{time}, the fraction of the time that each latitude band spends between 273 K and 373 K. The bottom panel shows f_{area}, the fraction of the land area that is between 273 K and 373 K at different times of year, measured from the Northern winter solstice. In both panels, the solid line shows results for our fiducial model, the dashed line shows results for the Earth using the zonally-averaged NCEP/NCAR temperature data in 2004, and the dotted line shows results for the Earth using the full two–dimensional NCEP/NCAR 2004 temperature data.

5.3.2.1 The Case of Earth

Before analyzing the fractional habitability of model planets with different characteristics than Earth, it is instructive to consider the fractional habitability of Earth itself and of our fiducial Earth-like model. We return to the NCEP/NCAR temperature data in 2004 for this comparison. Figure 5.6 shows a remarkably good agreement between the temporal and regional habitability of our fiducial model (solid line) and that of Earth (dashed and dotted lines). We analyzed the 2004 Earth temperatures in two different ways. In order to create a variable that is directly comparable to our (one dimensional) model temperatures, we must perform zonal averages. In Figure 5.6, statistics derived from first zonally-averaged temperatures are represented by dashed lines. We also used the entire data set (latitude, longitude and time) to calculate habitability fractions before performing zonal averages. Statistics derived from this full set of temperature data are represented by dotted lines.

The top panel in Figure 5.6 shows the fraction of the time that each latitude strip spends at or above 273 K. There are some differences between the dashed and the dotted lines, most noticeably in the steepness of the descent from 100% habitable to 0% habitable moving from equator to North Pole. The crucial feature of this plot, from a model validation perspective, is that the f_{time} values derived from zonally averaged temperatures on Earth have a very steep descent in both hemispheres, similar to what is seen in our fiducial model.

The bottom panel in Figure 5.6 shows the fraction of the surface area that is

habitable as a function of time of year (measured from the Northern winter solstice). The fiducial model maintains nearly constant regional habitability throughout the year. Earth's regional habitability varies slightly over the year, increasing a bit in the Northern summer because the North Polar ice cap shrinks much more during its summer than the South Polar cap does during its.

Finally, the net fractional habitability (f_{hab}) of Earth and of our fiducial model are just the average values of the curves shown in the bottom panel of Figure 5.6. Our fiducial model is 83% habitable. Earth was 85% habitable in 2004 when calculated from two–dimensional temperature data (this fraction varies by less than 0.5% from year to year). The net fractional habitability of Earth was 86% in 2004 when calculated from a temperature field that is first zonally averaged.

5.3.2.2   Temporal Habitability of Pseudo-Earths

Figure 5.7 presents a striking view of how regionality depends on the efficiency of latitudinal heat transport with the same nine pseudo-Earth EBMs as shown in Fig. 5.5. In this figure, contours indicate the value of the temporal habitability, f_{time}, as a function of orbital separation (a) and latitude (λ). When latitudinal heat transport is efficient – in the top panels – temporal habitability depends little on latitude. When transport is inefficient – in the bottom panels – there is a great deal of latitudinal dependence. Since the poles tend to be colder at a given separation, they tend to be habitable when the planet is closer to the star. For the same reason, the equator tends to be habitable when the planet is further away from the star.
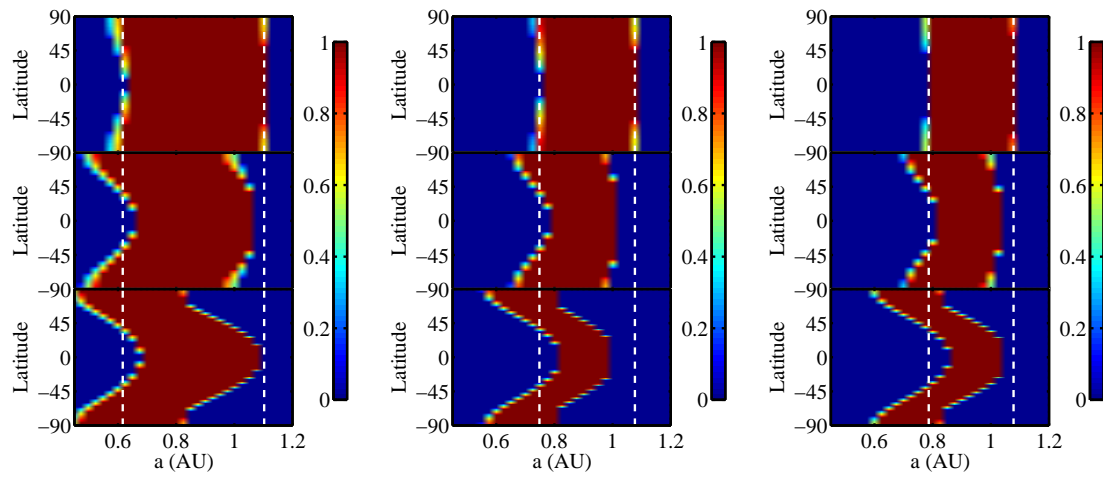
Figure 5.7 Temporal habitability fraction, f_{time}, as a function of orbital distance and latitude. Results are shown for the same nine pseudo-Earth models as in Fig. 5.5, with the same panel structure. Vertical dashed lines indicate the orbital extent of the radiative equilibrium habitable zone. Fractional habitability increasingly depends on latitude, and deviates from radiative equilibrium results, when the efficiency of latitudinal heat transport (D) is reduced (from top to bottom).

In each panel, the dashed vertical lines indicate the orbital extent of the radiative equilibrium habitable zone (REHZ), taken from Figure 5.5. In all models, allowing for regionality extends the inner boundary of the habitable zone, relative to the REHZ. Models with low efficiencies of latitudinal heat transport tend to have outer boundaries of the habitable zone that are closer to the star than models with high D values. In essentially all cases, this outer boundary is determined by a dynamical climate transition to a globally-frozen snowball state. As a result, the relationship between the outer boundary of the regionally habitable zone and D is not necessarily monotonic. In the left column of Figure 5.7, for example, the outer boundary of the regionally habitable zone is minimum for $D = D_{\text{fid}}$. In all models, at a given latitude, planets tend to be habitable either 100% of the time or 0% of the time, with little space for intermediate cases. As we shall see below, this is a consequence of the assumed large ocean fraction (70%) on the model planets. The large resulting thermal inertia minimizes the ability of a latitude band to swing below and above the freezing point during an annual cycle.

One may ask what is the physical process that actually drives a model planet to freeze over into a snowball state at a given orbital separation. For instance, in the central panel of Figure 5.7, which shows results for our fiducial Earth-like model, the planet has an Earth–like climate at $a = 1$ AU but is globally-frozen at $a = 1.025$ AU. What is different at 1.025 AU than at 1.000 AU to induce such a dynamical transition? Two possibilities spring to mind: the reduced insolation or the longer winters. In order to distinguish between these two possibilities, we

ran two additional models: one with Earth–like insolation and a year longer by a factor of $1.025^{3/2}$ (which did not freeze), and one with insolation reduced by a factor of 1.025^2 and a year the length of Earth's (which did freeze). This indicates that the reduced insolation is the dominant effect leading to snowball transitions in this model. As shown explicitly in Appendix II, however, longer winters can also contribute to dynamical snowball transitions.

5.3.2.3 Regional Habitability of Pseudo-Earths

Figure 5.8 shows temporal variation in fractional habitability for the same nine pseudo-Earth EBMs as before. In this figure, contours indicate the value of the regional habitability, f_{area}, as a function of orbital separation and time of year. As before, time of year is measured in fraction of a year from the northern winter solstice. There is little temporal variability in the nine models shown. This is a consequence of the North–South symmetry and the large thermal inertia of these models, with 70% ocean fraction. A model with North–South asymmetry in the distribution of land and sea, or even a symmetrical model with a much reduced ocean fraction, would show more significant variations in the value of f_{area} with time.

In models with efficient latitudinal heat transport (top panels), f_{area} tends to be either 0 or 1, taking on intermediate values only over a very small range of orbital separations. This is a consequence of the latitudinal isothermality of these models: essentiall the entire planetary surface is either above 273 K or below it.
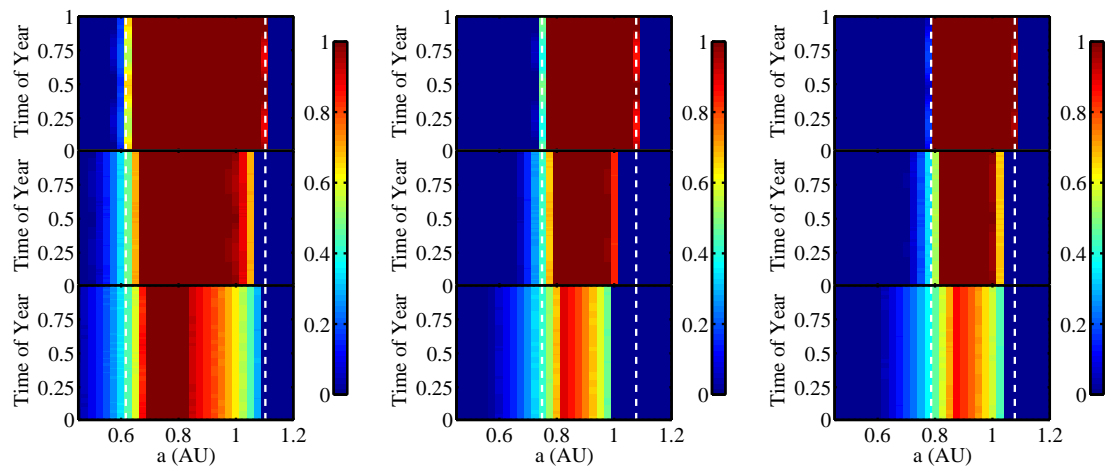
Figure 5.8 Regional habitability fraction, f_{area}, as function of orbital distance and time of year. Results are shown for the same nine pseudo-Earth models as in Fig. 5.5, with the same panel structure. Vertical dashed lines indicate the orbital extent of the radiative equilibrium habitable zone. In models with efficient latitudinal heat transport (top row), the regional habitability fraction tends to be either 0 or 1. In low transport efficiency models (bottom row), the regional habitability fraction takes on more intermediate values.

By contrast, models with inefficient latitudinal heat transport (bottom panels) are habitable over 100% of the planetary surface area only in a small portion of the regionally habitable zone. In fact, models shown in the lower middle and lower right panels are never habitable over the entire surface area of the planet.

5.3.2.4 Pseudo-Earth with a Small Ocean Fraction

The presence of abundant oceans on Earth plays a critical role in regulating its climate, but there is no fundamental reason why a terrestrial planet should have Earth–like ocean covering. The heat capacity C in equation (5.3) has units of [time]$\times$ [energy flux]/[temperature]. Multiplying C by a characteristic temperature and dividing by a characteristic flux, therefore, provides an estimate of the characteristic timescale for thermal relaxation:

$$t_{\text{relax}} \sim \frac{C \times \bar{T}}{(S/4) \times (1 - \bar{A})}. \tag{5.9}$$

For Earth–like conditions, using $T \sim 300$ K, $S \sim 1.4 \times 10^6$ erg cm^{-2} s^{-1}, and $\bar{A} \sim 0.5$, we find

$$t_{\text{relax}}^{\text{E}} \sim \frac{0.0017 \text{ K}}{1 \text{ erg cm}^{-2}} C \text{ s}. \tag{5.10}$$

In our EBM, we assign to the atmosphere a heat capacity that accounts for the type of (thermally coupled) surface that lies beneath it. Since the atmospheric heat capacity over land in our EBM is $C_l = 5.25 \times 10^9$ erg cm^{-2} K^{-1}, the corresponding relaxation timescale is roughly $t_{\text{relax}}^{\text{E},l} \sim 9 \times 10^6$ s, or about 0.3 years – between 3

and 4 months. On the other hand, the atmospheric heat capacity over ocean is 40 times larger because the atmosphere is able to tap into the larger thermal inertia of a mixed–wind layer at the ocean's surface. This results in a relaxation timescale $t_{\text{relax}}^{\text{E,o}}$ over ocean that is slightly longer than 10 years in our models. Consequently, in any EBM of a planet that is more than one part in 40 (2.5%) uniformly covered by ocean, the thermal relaxation timescale is approximately given by the ocean relaxation timescale times the ocean fraction.

The pseudo-Earth EBMs presented so far, all with a uniform 70% ocean fraction, have long thermal timescales of nearly a decade ($\sim$ 8 years). By contrast, a model planet with only a 10% ocean fraction would have a thermal timescale that is closer to a year. In such a model, one would expect significantly larger seasonal variations in temperature over the course of a year.

Figure 5.9 shows temporal and regional habitability diagrams for a pseudo-Earth EBM that is identical to our fiducial Earth-like model, except for a uniform ocean fraction reduced from 70% to 10%. The left panel shows the temporal habitability, f_{time}, as a function of orbital distance and latitude, while the right panel shows the regional habitability, f_{area}, as a function of orbital distance and time of year (measured in fraction of a year from the solstice). In contrast to the fiducial model with 70% ocean fraction shown in the central panels of Figures 5.7 and 5.8, there are now large swaths of parameter space, both in the left and right panels, that take on habitability fractions between 0 and 1. Seasonal variations of the regional habitability fraction are easily identified in the right panel, with
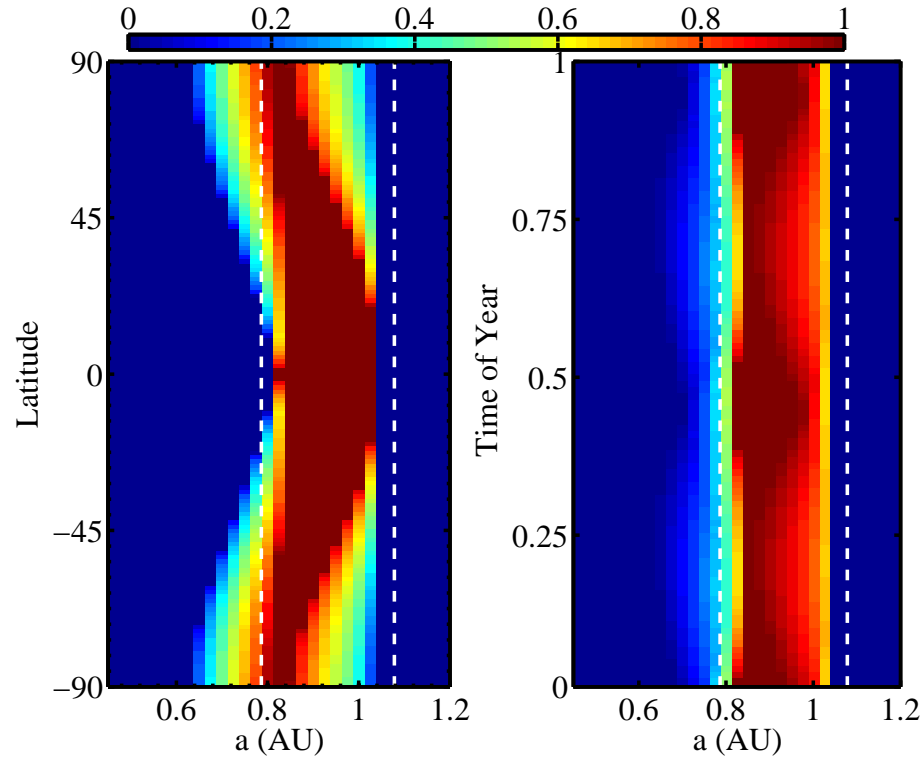
Figure 5.9 Temporal and regional habitability fractions in a pseudo-Earth model with a uniform 10% ocean fraction. The model is in every way similar to our fiducial Earth-like model, except for the reduced ocean fraction. The left panel shows the temporal habitability, f_{time}, as a function of orbital distance and latitude. The right panel shows the regional habitability, f_{area}, as a function of orbital distance and time of year. In each panel, vertical dashed lines indicate the orbital extent of the radiative equilibrium habitable zone. With this smaller ocean fraction, the temporal habitability (left panel) takes on more intermediate values (between 0 and 1) and the regional habitability (right panel) shows seasonal variations with time of year (two cycles per year because of the North–South model symmetry).

two "seasonal cycles" per year from the North–South symmetry of the model. Note that the range of orbital distances with nonzero fractional habitability, for this drier pseudo-Earth, differs only slightly from the corresponding range in the fiducial Earth–like model with 70% ocean fraction, with an outer edge somewhat extended.

Table 5.2 summarizes the different sizes of fractionally habitable zones for the various models considered in this study. Compared to global radiative balance calculations, fractionally habitable zones extend closer to the star, simply as a result of the range of regional conditions that obtain on the planet in our EBMs. On the other hand, the outer reach of fractionally habitable zones tends to be limited, before the outer radiative limit is reached, by dynamical climate transitions to globally-frozen snowball states. Planetary rotation rates and land-ocean fractions influence the orbital extent of fractionally habitable zones. Since these and other relevant planetary attributes will be unknown for terrestrial exoplanets, assessments of habitability for these newly discovered worlds will be complicated by these uncertainties.

5.4 On the Definition of Habitability

Determining the habitable zone of model or actual planets is an interdisciplinary endeavor that requires the contributions of biologists, climatologists, and astronomers. In light of our results on the rich phenomenology of habitable climates, we find it useful to mention here a few qualitative points that may prove to be important in refining the notion of habitability beyond the standard orbital ra-

Table 5.2 Habitable Zone Extents

Table 5.1 Model:	# 1	# 2	#3
Global Radiative Balance	0.616-1.103	0.748-1.077	0.786-1.078
EBM with $D = 9D_{\mathrm{fid}}$	0.550-1.125	0.700-1.100	0.750-1.100
EBM with $D = D_{\mathrm{fid}}$	0.450-1.075	0.625-1.025	0.675-1.050
EBM with $D = (1/9)D_{\mathrm{fid}}$	0.400-1.100	0.525-1.000	0.575-1.050
EBM with 10% ocean fraction	–	0.625-1.050	–

Note. — Habitable zones in each of the models considered. All orbital distances are in AU.

dius requirement for surface liquid–water generally discussed in the astronomical literature (e.g., Kasting et al. 1993).[8]

Habitability on Earth is not strictly tied to the freezing point of water. Carpenter et al. (2000) have discovered bacterial activity in South Pole snow, indicating that some organisms can reproduce at ambient temperatures below $-10°$ C and can live at temperatures down to $-85°$ C. Nor is habitability strictly tied to the sea–level boiling point of water on Earth. Kashefi & Lovley (2003) have cultured a strain of archaea that happily reproduces at $121°$ C, and remains viable at temperatures up to at least $130°$ C. The idea that life requires temperatures appropriate for liquid water at 1 Atm pressure is therefore at least partially flawed. These examples would indicate that the habitable temperature range ought to be extended to 263 K–394 K, or even possibly 188 K–403 K. Biologists will have the final say on three vital ranges of temperature, by determining when life can arise, survive and reproduce. The temperature regime in which organisms can reproduce may be the relevant one for the definition of habitability in the context of astronomical searches.

Even if life requires liquid water, one must still carefully consider the upper temperature limit to habitability. Water boils at 373 K only for a surface pressure of 1 Atm. There is so much water in the Earth's oceans that the surface temperature could be well above 373 K and the atmospheric H_2O pressure change could allow liquid water to remain on the surface. On the other hand, since water is a greenhouse gas and humidity increases with temperature, a planet's ability to

---

[8]We focus on water, rather than other potential solvents upon which a biochemistry might be able to develop, because water is the only molecule that we know can form the basis of a biosphere.

cool is likely to become significantly reduced at some critical temperature. Unless additional feedbacks intervene at these high temperatures, such as higher albedo from increased cloud coverage, a regime of "runaway greenhouse effect" may be reached. Kasting (1988) and Sugiyama et al. (2005) have argued that this runaway regime would be reached for global surface temperatures between ~ 350 K and ~ 400 K on Earth. These important considerations, which have not been addressed in our modeling strategy, suggest that the high–temperature end of the habitable range may be the critical temperature for runaway greenhouse effect, at least for Earth-like planets.

We note, however, that all these notions should also be fully integrated with the regional and seasonal character of climate systems. Mean planetary conditions do not necessarily determine the ability of a planet to host life. Rather, the various conditions achieved on different parts of the planet, at different times of year, do. Knowning mean conditions is useful only insofar as they are a good proxy for actual local conditions, but in many cases they may not be. Allowing for seasonality renders concepts such as the runaway greenhouse effect significantly more complicated. For example, if parts of a planet (e.g., equatorial regions) are above the critical temperature for runaway greenhouse effect, while other parts (e.g., polar regions) are not, determining whether or not the planet's temperature does run away becomes a non-trivially coupled problem.

Finally, we note that in order for a planet to be considered habitable, it may need attributes such as its net fractional habitability, f_{hab}, to exceed some minimum

threshold values. For one thing, this may be required for life to arise and survive on the planet. But such a requirement may also be important in terms of observational searches for biomarkers and other signatures of life. If only a small fraction of the planet's surface area is habitable, or if it is habitable only for a small fraction of the year, the planet may host life but this life may be unable to generate large enough signatures – e.g., in spectral biomarkers (Kaltenegger & Selsis 2007; Kiang et al. 2007; Grenfell et al. 2007; Seager et al. 2005) – to permit reliable detections with specific remote sensing instruments. Understanding the minimum fractional habitability necessary to produce sufficient atmospheric concentrations of biomarkers detectable by instruments such as *TPF* and *Darwin* will be an interesting challenge for the emerging field of astrobiology.

5.5   Conclusion

We have reconsidered the notion of habitability for Earth-like planets with seasonal energy-balance climate models. These models show that the concept of regional and seasonal habitability is generally important to assess the ability of terrestrial exoplanets to host life. We find that previous evaluations of habitable zones may have omitted important climatic conditions by focusing on close Earth analogies. We illustrate this with two specific examples: pseudo–Earths rotating at different rates or possessing a smaller ocean fraction than Earth itself. These pseudo-Earths have quantitatively different climatic habitability properties than the Earth itself.

Comparisons to global radiative balance calculations show that seasonal hab-

itability generally extends the inner orbital range of Earth-like planet habitable zones. The outer orbital range is reduced relative to what global radiative balance calculations would indicate, however, because the climate generally makes a dynamical transition to a globally-frozen snowball state before the outer radiative limit is reached. The stability of a planet's climate against snowball events therefore has a strong impact on its habitability. Since this stability is partly determined by external forcing factors such as the magnitude of insolation and the length of winters, we expect issues of climate dynamics to be central in determining the habitability of terrestrial exoplanets, particularly if their forcing conditions are generally different from the moderate cases encountered in the Solar System.

APPENDIX

In this appendix, we describe several numerical tests developed in the process of building our EBMs. We discuss the importance of sufficient spatial resolution for converged results and the effects that initial conditions have in determining the final periodic climate solutions obtained.

I Numerical Resolution

We investigated numerical resolution issues with a hot–start version of our reference Earth-like model. It has uniform initial temperature $T = 350$ K, uses "model 2" formulations for the IR-cooling and albedo functions (see Table 5.1) and adopts the
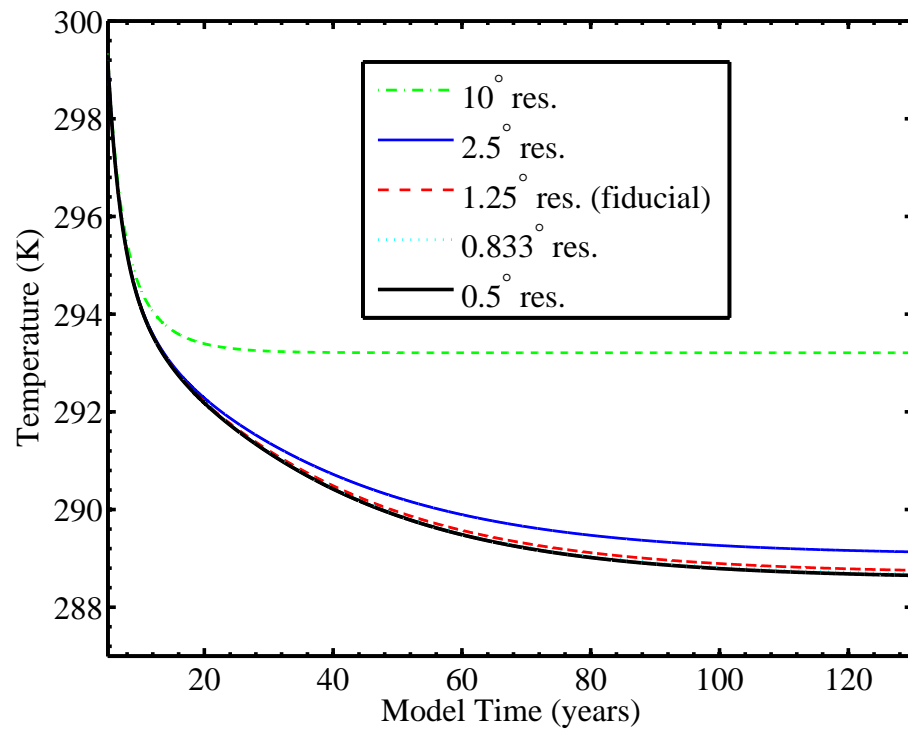
Figure 5.10 Numerical convergence tests. The global mean temperature in our fiducial Earth-like model is shown, convolved with a 1–year boxcar filter, for five different numerical resolutions. From top to bottom, the resolution is 10° on the sphere (19 latitudinal grid points), 2.5° (73 points), 1.25° (145 points), 0.833° (217 points), and 0.5° (361 points). The default resolution adopted in our work is 1.25°.

fiducial value $D = D_{\mathrm{fid}}$ for the magnitude of latitudinal heat transport. Here, we describe the results of a convergence test for five spatial resolutions: $10°$ on the sphere (19 grid points equally spaced in latitude, including the poles); $2.5°$ (73 points); $1.25°$ (145 points); $0.833°$ (217 points); and $0.5°$ (361 points; the "high resolution model"). Figure 5.10 shows the time evolution of the global mean surface temperature in the same Earth-like models at different numerical resolutions. Small-amplitude annual oscillations in the global mean temperature have been smoothed out by convolving each temperature time series with a 1–year boxcar filter.

As the numerical resolution is increased, differences between the global temperature relaxation curves diminish rapidly, as expected from the second-order spatial differencing scheme used. After 130 years, the model with $2.5°$ resolution achieves a periodic solution for the global mean temperature that is within ~ 0.5 K of the solution with the high resolution model. The $1.25°$ resolution model stays within 0.2 K of the high resolution model and the $0.833°$ resolution model is barely distinguishable from the high resolution model (within 0.1 K). To strike a good balance between numerical accuracy and integration speed, we adopt a $1.25°$ resolution in all our models. This guarantees that our results are not significantly affected by finite resolution effects. With this level of numerical accuracy, it is clear that the simplicity of various assumptions in our modeling strategy will be the key factor limiting the predictive power of our EBMs.

We note that the model with $10°$ resolution relaxes to a periodic solution much faster than the other models and that it settles to a global mean temperature that

deviates from the higher resolution results by more than 4 K. This consequence of poor numerical resolution may be important since several EBM studies have been carried out in the past at this standard 10° resolution (e.g., Sellers (1969), WK97). Traditionally, one of the main justifications for using such a low spatial resolution is that the diffusive model for latitudinal heat transport is not expected to provide a reliable description of much smaller spatial scales given an Earth-like circulation regime (e.g., Lorenz (1979); North & Coakley (1979)). It is our opinion, however, that it is preferable to reach satisfactory numerical convergence with the diffusive model even if the model itself is not be used to interpret phenomena below some intermediate scale effectively exceeding the numerical resolution. Otherwise, even global results like the mean surface temperature shown in Fig. 5.10 may be prone to finite resolution effects. This becomes particularly important in the context of climate dynamical studies where slight differences in model attributes, such as initial conditions, can have a profound influence on the ultimate outcome for the climate, as we shall now illustrate.

II Initial Conditions

Repeated experiments with our reference Earth-like EBM show that, as long as the uniform initial temperature is chosen to be far above 273 K, essentially all memory of the initial conditions is lost by 130 years of evolution. For initial temperatures approaching 273 K, however, there can be very sensitive dependence on the initial conditions, due to strong water-ice albedo feedback effects. Here we show that

even for the same initial conditions in temperature, different outcomes for the climate are possible depending on the time of year at which the model is started. Figure 5.11 shows space-time diagrams of temperature in two models which are identical in all respects, including a uniform initial $T = 275$ K, except for the initial time on the orbit. The model in the top panel begins with the 23.5°-obliquity planet at the northern winter solstice while the one in the bottom panel begins at the northern vernal equinox (a quarter year later). After 60 years, the model in the top panel is still evolving, but it is clearly making its way to the same periodic, partially ice-covered solution as in the top panels of Figure 5.4. The model in the bottom panel, on the other hand, transitions to a fully ice-covered snowball state after slightly less than 40 years. The North–South asymmetry evident in the bottom panel shows that the model never lost memory of its unbalanced start. The model's northern hemisphere started in the spring, and therefore experienced 6 months of greater insolation than the southern hemisphere, from the very beginning. As a result, the southern hemisphere remained significantly, and increasingly, colder than the northern hemisphere, until the cold temperatures in the south were able to drag the entire planet down to a globally-frozen snowball state.
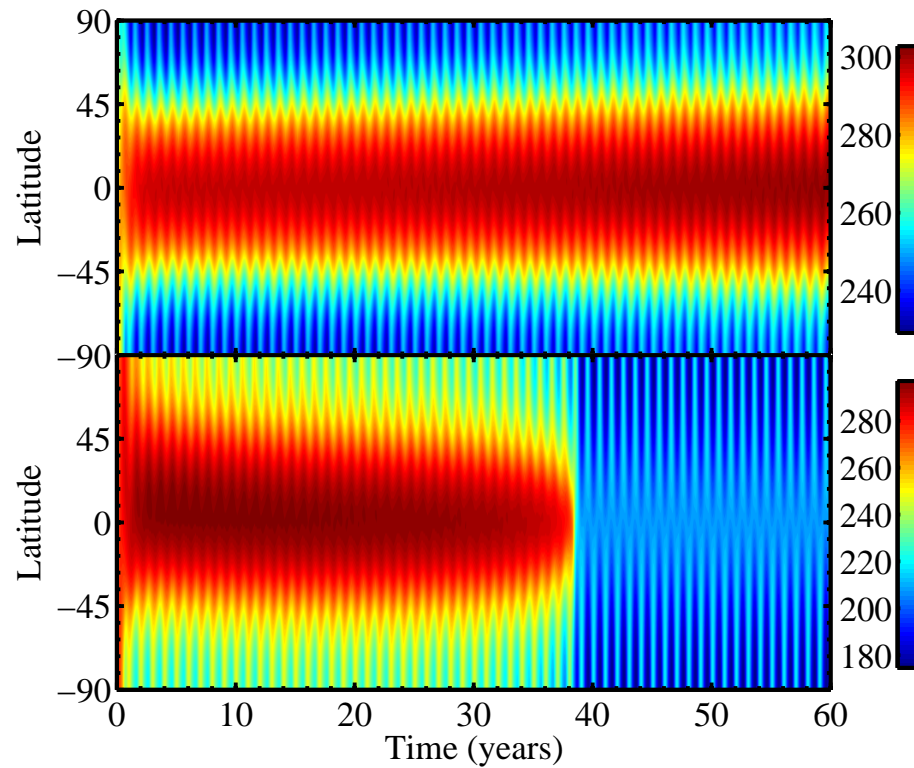
Figure 5.11 Dependence on initial conditions (orbital time). Both panels show space-time diagrams of temperature in our fiducial Earth-like model starting with a uniform temperature of 275 K. The model in the top panel begins at the Northern winter solstice and it recovers from the rather cool initial condition by asymptotically approaching the same partially ice-covered climate solution as for typical hot–start cases. The model in the bottom panel, however, begins at the Northern vernal equinox and makes a dynamical transition to a globally-frozen snowball state after approximately 38 years.

Chapter 6

Habitable Climates: The Influence of Obliquity

6.1   Introduction

The Earth's obliquity is remarkably stable: the angle between the spin–axis and the normal to the orbital plane varies by no more than a few degrees from its present value of $\sim 23.5°$. This stability is maintained by torque from the Moon (Laskar et al. 1993; Neron de Surgy & Laskar 1997). Even within our own Solar System, though, the obliquity of other terrestrial planets has varied significantly more; the analysis of Laskar & Robutel (1993) indicates that Mars' obliquity exhibits chaotic variations between $\sim 0°$ and $\sim 60°$.

How does climate depend on obliquity? How does the range of orbital radii around a star at which a planet could support water-based life depend on the

planet's obliquity? Has the stability of Earth's obliquity made it a more climatically hospitable home? The answers to these questions will be crucial to evaluate the fraction of stars that have potentially habitable planets. There are now more than 280 extrasolar planets known, several of which are less than 10 times the mass of the Earth.[1] *Corot*, which has already launched, and *Kepler*, scheduled to launch in less than one year, are dedicated space-based transit-detecting observatories that will monitor a large number of stars to detect the small decreases in stellar flux that occur when terrestrial planets cross in front of their host stars (Baglin 2003; Borucki et al. 2003, 2007). These missions are expected to multiply by perhaps a thousandfold or more the number of known terrestrial planets, depending on the distribution of such planets around solar-type stars (Borucki et al. 2007, 2003; Basri et al. 2005; although see revised predictions in Beatty & Gaudi 2008). NASA and ESA have plans for ambitious future missions to obtain spectra of nearby Earth-like planets in the hope that they would reveal the first unambiguous signatures of life on a remote world: NASA's *Terrestrial Planet Finder* and ESA's *Darwin* (Leger & Herbst 2007). The design of such observatories, and the urgency with which they will be built and deployed, will depend on the habitability analysis of terrestrial planets that will be found in the next 5-10 years.

Over the last 50 years, various authors have addressed how to predict the way in which terrestrial planet habitability depends on star-planet distance. Several of the important initial calculations predated the first discoveries of extrasolar planets, including Dole (1964), Hart (1979), and the seminal work of Kasting et al. (1993).

---

[1]See http://exoplanet.eu/.

More recently, Williams & Kasting (1997, hereafter WK97) and Williams & Pollard (2003) have tackled precisely the questions posed above, and have concluded that obliquity variations do not necessarily render a planet non-habitable. Here we seek to generalize these analyses to model planets that are less close analogs to Earth than have been considered previously. Chapter 5 (Spiegel et al. 2008, hereafter SMS08) examines how regionally and temporally habitable climates are affected by variations in the efficiency with which heat redistribes on a planet, and by variations in the ocean fraction. Following in the footsteps of that analysis, here we again consider the influence on habitability of several unobservable planetary attributes: we simultanously vary with obliquity both the efficiency of latitudinal heat transport, and the distribution of land and ocean.

The remainder of this chapter is structured as follow: In § 6.2 we describe the energy balance model we use; in § 6.3 discuss several validation tests in which our model performs well enough to give us some confidence in its behavior for conditions that differ from those found on Earth; in § 6.4 we examine the influence on regional and seasonal habitability of various excursions from Earth-like conditions; and in § 6.5 we conclude by considering why the Earth's South Pole is so cold and by examining the suitability of global radiative balance models for habitable zone analysis.

6.2 Model

In order to describe the temperature structure and evolution on a planet, we use a 1-dimensional time-dependent energy balance model based on a diffusion equation. This type of model has been used in previous investigations of habitable climates (WK97, SMS08), in modeling Martian climate under changes in forcing (Nakamura & Tajika 2002, 2003), and in studies of the Earth's climate (North et al. 1981 and references therein).

Our model is based on the following prognostic equation for the planetary surface temperature (as described in SMS08):

$$C\frac{\partial T[x,t]}{\partial t} - \frac{\partial}{\partial x}\left(D(1-x^2)\frac{\partial T[x,t]}{\partial x}\right) + I = S(1-A). \tag{6.1}$$

In this equation, $x \equiv \sin\lambda$ is the sine of latitude λ, T is the temperature, C is the effective heat capacity of the surface layer, D is the diffusion coefficient that determines the efficiency of latitudinal redistribution of heat, I is the infrared emission function (energy sink), S is the diurnally averaged insolation function (energy source) and A is the albedo. In the above equation, C, D, I, and A may be functions of T, x, t, and possibly other relevant parameters.

Our prescriptions for the functions C, D, I, and A follow SMS08 and are largely borrowed from WK97 or the existing geophysical literature on 1D EBMs. For simplicity and flexibility, many of our models use very simple, physically motivated prescriptions. As described in SMS08, we find that an infrared cooling

function of the form

$$I[T] = \sigma T^4/(1 + (3/4)\tau_{\mathrm{IR}}[T]) \tag{6.2}$$

(i.e., a one-zone model combined with a simple Eddington transfer approximation; e.g., Shu 1982), reproduces the greenhouse effect on Earth reasonably well. Here, τ_{IR} represents the opacity of the atmosphere to long wavelength infrared radiation. SMS08 describes three pairs of infrared radiation functions and albedo functions. In this analysis, we will use two of the three: (I_2, A_2), which gives the closest match to the Earth's temperature distribution, and (I_3, A_3), which uses the standard linearized cooling function of North & Coakley (1979). The albedo functions A_2 and A_3 are constant and low (~ 0.3) for high temperatures, constant and high (~ 0.7) for low temperatures (to represent the high albedo of snow and ice), and vary smoothly in between. The (I, A) functions are presented in Table 6.1. One other I function that we use is the one proposed in WK97, derived from full radiative-convective calculations, here denoted I_{WK97}. This function includes the influence of CO_2 on radiative cooling. For C, we assume various configurations of land and ocean, in each configuration using the same land, ocean, and ice partial C values as WK97. Finally, we adopt $D_{\mathrm{fid}} = 5.394 \times 10^2$ erg cm^{-2} s^{-1} K$^{-1} \times (\Omega_p/\Omega_\oplus)^{-2}$, as described in SMS08, where Ω_p is the angular spin frequency of the model planet and $\Omega_\oplus$ is that of the Earth.

Equation (6.1) is solved as described in SMS08, on a grid uniformly spaced in latitude (1.25° resolution, found to be sufficient from convergence tests). We again choose "hot start" ($T \geq 350$ K) initial conditions, to minimize the likelihood that

Table 6.1   Atmospheric Models.

Model	IR Cooling Function	Albedo Function
2[a]	$I_2[T] = \frac{\sigma T^4}{1+(3/4)\tau_{\mathrm{IR}}[T]}$	$A_2[T] = 0.525 - 0.245\,\tanh[\frac{(T-268\mathrm{K})}{5\mathrm{K}}]$
3[b]	$I_3[T] = A + BT$	$A_3[T] = 0.475 - 0.225\,\tanh[\frac{(T-268\mathrm{K})}{5\mathrm{K}}]$
3[c]	$I_{\mathrm{WK97}}[T, p\mathrm{CO}_2]$	$A_2[T]$

[a] Model with T–dependent optical thickness: $\tau_{\mathrm{IR}}[T] = 0.79(T/273\mathrm{K})^3$.

[b] Linearized model: $A = 2.033 \times 10^5$ erg cm^{-2} s^{-1}, $B = 2.094 \times 10^3$ erg cm^{-2} s^{-1} K^{-1}.

[c] WK97 cooling function; A_2 albedo: The functional form is quite complicated and is presented in the appendix of WK97.

Note. — σ is the Stefan-Boltzmann constant.

models will undergo a dynamical transition to a fully ice-covered (snowball) states from which they cannot recover because of ice-albedo feedback.

To summarize, we make the following assumptions in the models presented below:

(a) *Heating/Cooling* – The heating and cooling functions are given by the diurnally averaged insolation from a sun–like (1 $M_\odot$, 1 $L_\odot$) star, with albedo and insolation functions described above and in Table 6.1.

(b) *Latitudinal Heat Transport* – We test the influence on climate of three different efficiencies of latitudinal heat transport, within the diffusion equation approximation: an Earth-like diffusion coefficient, and diffusion coefficients scaled down and up by a factor of 9 (which correspond to 8-hour and 72-hour rotation according to the $D \propto \Omega_p^{-2}$ scaling).

(c) *Ocean Coverage* – We vary both the fraction and the distribution of ocean coverage. For ocean fraction, we present a series of models with Earth–like 30%:70% land:ocean ratio, and another series of models that represent a desert-world, with a 90%:10% land:ocean ratio. For ocean distribution, we present models in which there is a uniform distribution (in every latitude band) of land and ocean, and others in which the land-mass is a single continent centered on the North Pole, while the rest of the planet is covered with ocean.

(d) *Initial Conditions* – As described in SMS08, the models all have a hot-start

initial condition, with the planet's temperature uniform and at least 350 K, to minimize the chances of ending up in a global snowball state owing solely to the choice of initial conditions. Time begins at the Northern winter solstice.

6.3 Model Validation

In SMS08, we verified that our "fiducial" model at 1 AU – 70% ocean, I_2, A_2 cooling/heating – predicts temperatures that match the Earth's actual temperature distribution at all latitudes that are not significantly affected by Antarctica (north of 60° S or so). This indicates that the model accounts for the overall planetary energy balance reasonably well. Another obvious test is whether the model correctly predicts the individual energy fluxes that together go into the overall balance.

The diffusion equation model is essentially a statement of conservation of energy. By definition,

$$C\frac{\partial T}{\partial t} \equiv \frac{\partial \sigma}{\partial t},$$

(6.3)

where σ is the energy surface density (internal energy per unit surface area on the globe). The diffusion equation, therefore, says that the rate of change of internal energy at a given point equals the sources of energy (insolation), minus the sinks (infrared radiation), minus whatever energy flows away from the point under consideration.

Figure 6.1 presents a comparison between the annually averaged fluxes of incoming and outgoing radiative energy in the fiducial model with the corresponding

fluxes on Earth, taken from NASA's Earth Radiation Budget Experiment (ERBE) in the mid-1980s (Barkstrom et al. 1990). While our model does not capture the full shape of the Earth's cooling and heating functions – in particular, the annually averaged model heating function is a bit below the Earth's at the poles – still, both cooling and heating fluxes are within 10% of the Earth's over most of the planet's surface.

Figure 6.2 offers an even more compelling validation. In this figure, each of the 12 panels shows solar (i.e., heating), terrestrial (i.e., cooling), and net (solar minus terrestrial) radiative fluxes as functions of latitude for one month. We were heartened to see that not only are our annually averaged cooling and heating functions in reasonable gross agreement with Earth's, as per Figure 6.1, but furthermore the temporal variability of radiative fluxes in our model is similar to that of Earth.

Another way to consider the time variability of heating and cooling fluxes is to look at the global average of each with respect to time. Figure 6.3 presents precisely this comparison, for both our fiducial model and for actual Earth data. The bottom panel of this figure shows the net heating flux as a function of time of year, measured in fraction of a year from the Northern winter solstice. Earth's heating flux varies by about 5% with respect to the heating flux, while our model's varies somewhat less. The heating function for the Earth exceeds the cooling function during Northern winter for two main reasons: First, the nonzero eccentricity ($e \approx$ 0.0167) of the Earth's orbit places its perihelion – which occurs during Northern winter – approximately 3.4% closer to the Sun than its aphelion. This is responsible
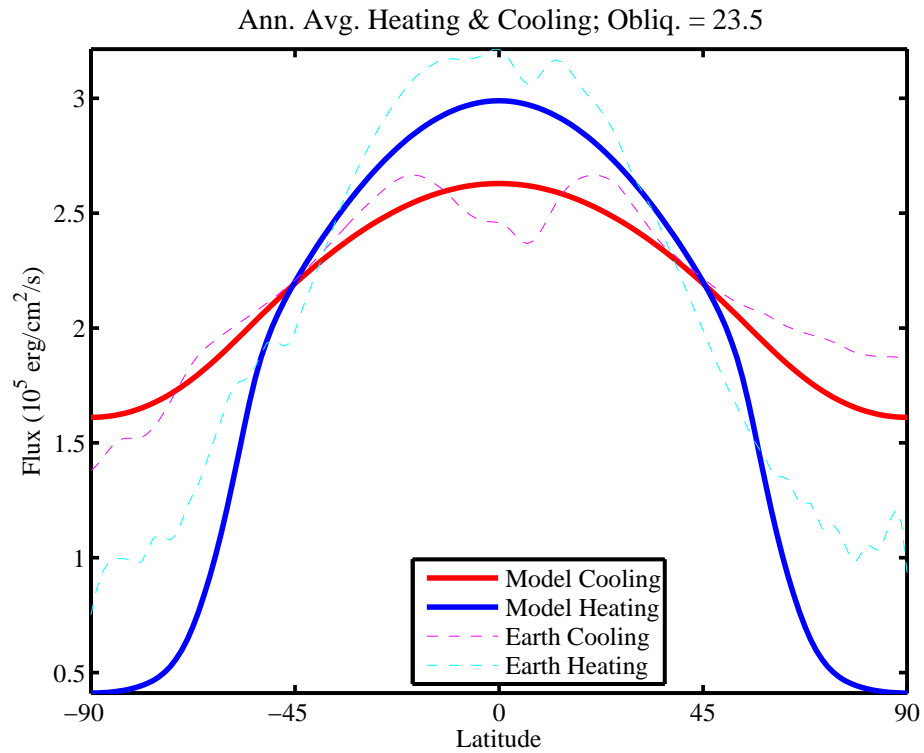
Figure 6.1 Annually averaged cooling and heating fluxes for fiducial model at 1 AU and Earth. The thick red line is infrared cooling in our fiducial model (70% ocean, I_2, A_2); the thick blue line is absorbed solar flux in our fiducial model. The thin dashed magenta line is the Earth's annually averaged long wavelength infrared radiation, and the thin dashed cyan line is the annually averaged absorbed solar flux on Earth.
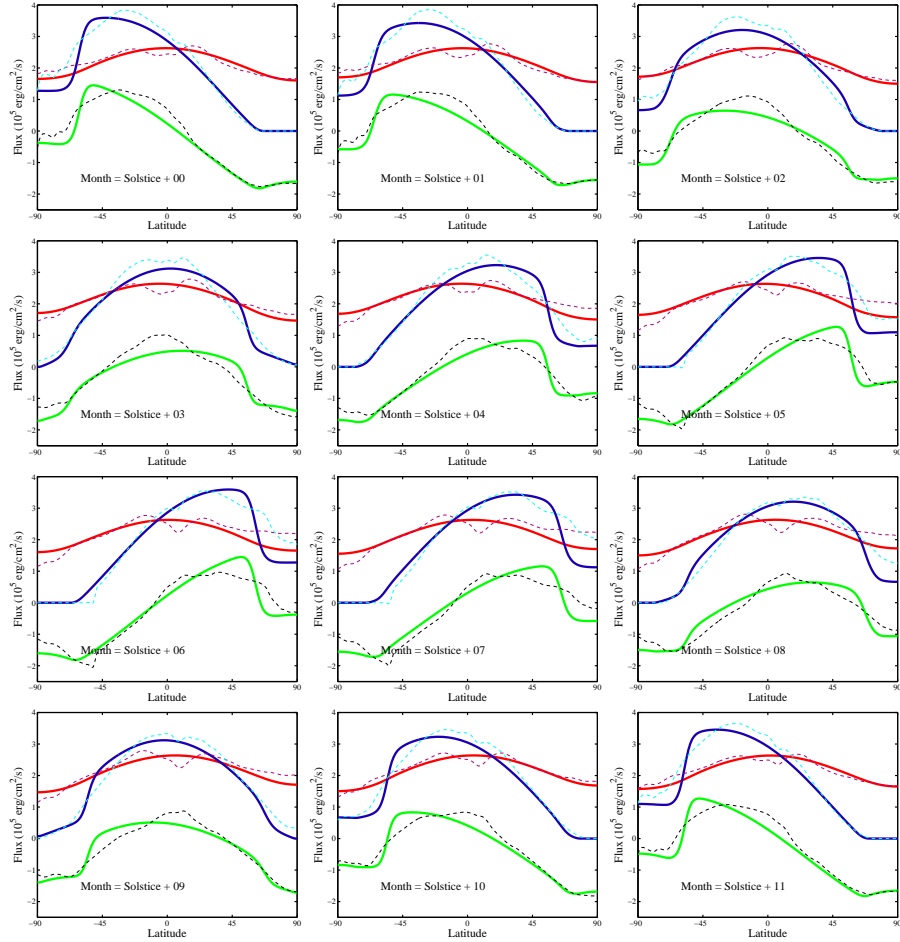
Figure 6.2 Monthly cooling, heating, and net fluxes for fiducial model at 23.5° obliquity, at 1 AU, and for Earth. Each panel presents the average cooling, heating, and net (heating minus cooling) radiative fluxes, as functions of latitude, for one month of the year, starting at the Northern winter solstice (upper left panel), and incrementing by one month with each panel to the right. These fluxes are presented for both the model (thick solid lines) and the Earth (thin dashed lines). Model heating is blue; model cooling is red; model net heating is green. Earth heating is cyan; Earth cooling is deep magenta; Earth net heating is black.

for ~ 7% of the annual ~ 10% annual variation in net heating flux. Another contributing factor is that the Earth's oceans on average absorb somewhat more insolation than the land, and the Southern hemisphere – which faces the Sun during Northern winter – has greater ocean coverage than the Northern hemisphere.

So far we have considered the radiative fluxes, but what about diffusive energy flux? We may combine equation (6.1) with equation (6.3) to produce:

$$\frac{\partial \sigma}{\partial t} - \frac{\partial}{\partial x}\left\{D\cos^2\lambda\frac{\partial}{\partial x}T\right\} = (\text{sources} - \text{sinks})_{\text{energy per area}} , \tag{6.4}$$

where we have substituted $\cos^2\lambda$ for $(1 - x^2)$. The λ part of the divergence in spherical coordinates is

$$\text{Div}\left[\vec{F}\right]_\lambda = \left(\frac{1}{R\cos\lambda}\right)\frac{\partial}{\partial\lambda}\{\cos\lambda F_\lambda\} , \tag{6.5}$$

where F_λ is the flux of some quantity in the $+\lambda$ direction. Since $x \equiv \sin\lambda$, $dx = \cos\lambda d\lambda$. Therefore, we may rewrite equation (6.5) as follows:

$$\text{Div}\left[\vec{F}\right]_\lambda = \left(\frac{1}{R}\right)\frac{\partial}{\partial x}\{\cos\lambda F_\lambda\} = \frac{\partial}{\partial x}\left\{\frac{\cos\lambda F_\lambda}{R}\right\} \tag{6.6}$$

By the form of the diffusion equation (eq 6.4), the term whose divergence is being taken is the rate of energy transport (per unit length). Referring back to equation (6.6), the term inside the curly braces equals $(\cos\lambda/R)F_\lambda$, where F_λ is the
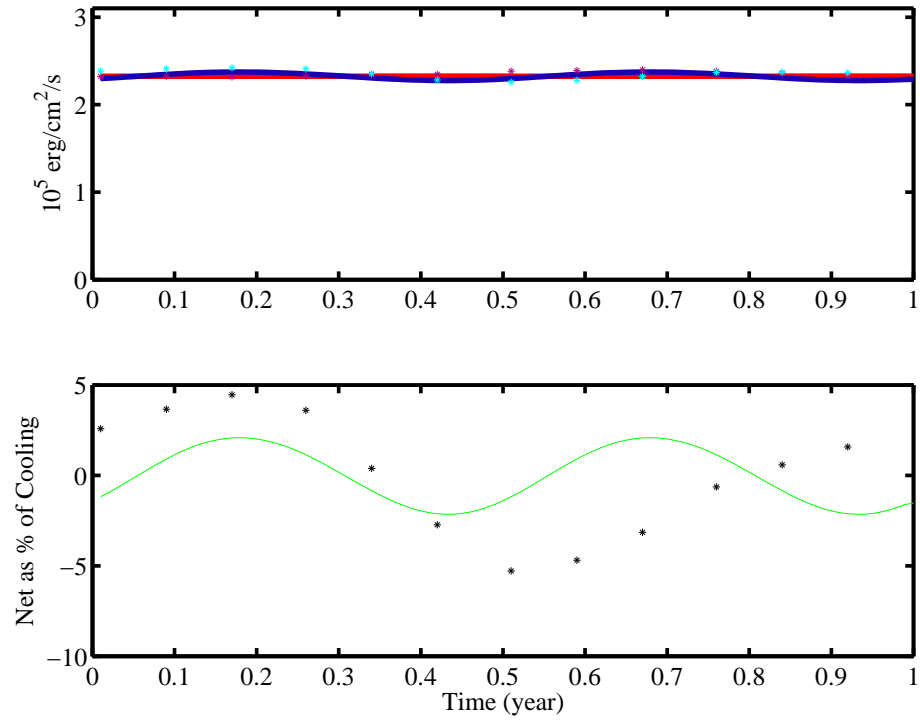
Figure 6.3 Global average cooling, heating, and net radiative flux, as functions of time in fiducial model at 23.5° obliquity. Model fluxes are solid lines; ERBE data are stars. *Top Panel:* Global average cooling (red curves and magenta stars) and heating (blue curves and cyan stars) fluxes as a function of time of year, measured in fraction of a year from the Northern winter solstice. *Bottom Panel:* Net heating flux (heating minus cooling) for the model (green curve) and ERBE data (black stars), plotted as percent of the cooling flux.

rate of energy transport per unit length. As a result,

$$F_\lambda = RD \cos \lambda \frac{\partial T}{\partial x}. \tag{6.7}$$

The total diffusive flux $\mathcal{F}_\lambda$ (taking into account pathlength), therefore, is $2\pi R \cos \lambda F_\lambda$:

$$\mathcal{F}_\lambda = 2\pi R \cos \lambda F_\lambda = 2\pi R^2 D \cos^2 \lambda \frac{\partial T}{\partial x} \tag{6.8}$$

Figure 6.4 shows the meridional diffusive heat flux in our fiducial model, at Earth-like 23.5° obliquity and at extreme 90° obliquity. In the Earth-like configuration, heat flows from the equator toward the poles. In the highly oblique configuration, heat flows in the other direction, from the poles to the equator (in an annually averaged sense). Williams & Pollard (2003) present a full general circulation model (GCM) of an Earth-like planet at Earth-like and higher obliquity. Figure 2 of that paper shows the meridional heat flux within their models for 23.5° obliquity and 85° obliquity, and the results are strikingly similar to ours.[2] At 23.5° obliquity, our model's diffusive flux is very close to that of the GCM. At higher obliquity, the flux of our model still appears to be within 30% of the GCM's at all latitudes.[3] This reasonable concordance indicates that the treatment of heat transport within our model, despite being very simple, is still likely to make plausible

[2] Trenberth & Solomon (1994) present the oceanic poleward heat transport for the Earth, which is roughly 50% the flux in the GCM. This is sensible, because the job of carrying heat to the poles is shared roughly evenly between the atmosphere and the oceans.

[3] We say "appears" because we are gauging the degree of correspondance from visual inspection of their figure, since we do not have access to the GCM data.

predictions of heat transport in less-Earth-like conditions. We emphasize that it is a nontrivial point that this entirely different regime of transport should remain well-captured by a diffusion approximation.

6.4 Study of Habitability

For model planets with 23.5° obliquity at 1 AU, both pairs of infrared cooling functions and albedo functions that we present in this paper are reasonably good matches for the Earth's current climate as measured by latitudinally averaged temperature, with $(I_2, A2)$ being the somewhat closer fit. This gives us some confidence that these functions are reliable guides to how the climate might respond under different forcing conditions. In this investigation, we consider how variations in intrinsic planetary characteristics combine with the changes in insolation and year-length that would result from moving changing a model planet's orbital radius from 1 AU, in an effort to map the zone of regionally habitable climates.

We follow SMS08 and say that, at a given time, a part of a planet is habitable if its surface temperature is between 273 K and 373 K, corresponding to the freezing and boiling points of pure water at 1 Atm pressure. This criterion may be criticized for reasons discussed in SMS08 and references therein, but it provides a reasonable starting point for making investigations concrete. We will frequently quantify habitability of pseudo-Earths with the temporal habitability fraction, $f_{\text{time}}[a, \lambda]$, defined in SMS08, where a is orbital semimajor axis, λ is latitude, and f_{time} is the fraction of the year that the point in parameter space specified by (a, λ) spends in
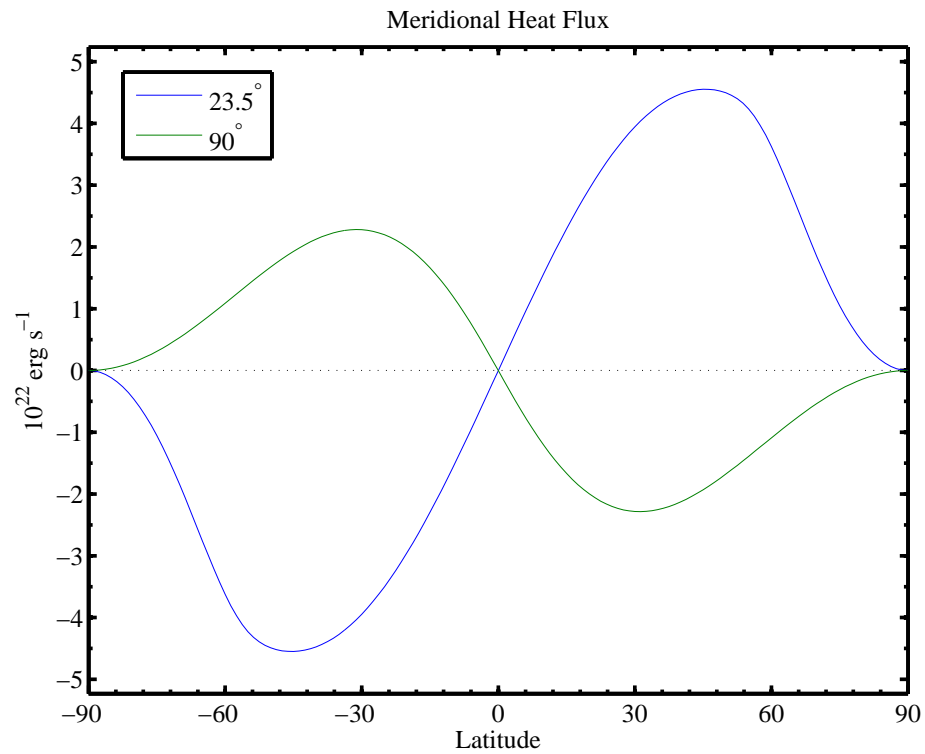
Figure 6.4 Annually averaged, longitudinally integrated, meridional heat transport at 23.5° and 90° obliquity. The curves indicate the longitudinally integrated, annually averaged, northward diffusive heat transport. In the Earth-like case (blue curve), heat flows from the equator to the poles; in the highly oblique case (green curve), the annually averaged heat flows in the opposite direction.

the habitable temperature range.

6.4.1 Efficiency of Heat Transport

Terrestrial planets with different rotation rates will redistribute heat from the substellar point (or, in a 1D model, the substellar latitude) differently. According to the scaling described above, wherein the effective diffusion coefficient varies with the inverse square of the planetary rotation rate, slower spinning planets will redistribute heat more efficiently, while faster spinning planets will do so less efficiently. But from where, and to where, is heat redistributed? How does this depend on obliquity and rotation rate? And what influence does this have on climatic habitability?

6.4.1.1 Direction of Heat Flow

For an Earth-like 23.5° obliquity, the substellar latitude does not vary very much over the course of the year: the tropics are fairly close to the equator (the tropical region is less than one third of the Earth's surface area). As a result, it is a reasonable approximation that heat is always being transported from the equator to the poles. In contrast, on a planet with significantly larger obliquity, the direction of heat flow changes over the course of the annual cycle. At the equinoxes, the equator is the most strongly insolated part of the planet (regardless of the obliquity), and so at the equinoxes heat builds up at the equator, eventually to be partially redistributed by atmospheric motions (including weather systems). But on a highly oblique

planet, polar summers are extremely intense, as measured by diurnally averaged insolation. As a result, heat builds up at the poles during their summers, and the flow of heat reverses direction.

Figure 6.5 demonstrates the effect of such strong polar summers on the global radiation budget of a model planet. In Figure 6.1 we see the annually averaged cooling and heating, as functions of latitude, for an Earth-like 23.5° model. As one would expect, over the annual cycle, the equator receives significantly more solar radiation than do the poles, and accordingly the annually averaged heating exceeds the cooling at the equator. This indicates that atmospheric motions transport heat poleward from the equator. In Figure 6.5, we present the analogous functions, in the case of high and extreme obliquity models. The left panel shows the heating and cooling functions for a model at 60° obliquity; the right panel shows the same functions for a model at 90° obliquity. In these models, the polar summers are so intense that, averaged over the year, the most strongly insolated parts of the planet are the North and South Poles! In an annually averaged sense, then, heat flows from the poles to the equator, although clearly the direction of flow changes with the seasons, as described above. The import of these plots is that our notion that the poles are the cold parts might have to be revised in the case of highly oblique worlds.

Figures 6.6 and 6.7 show in greater detail the extreme way that the insolation varies over the annual cycle in highly oblique models. Notice that the cooling remains much more consistent than the heating in these models. This is because of
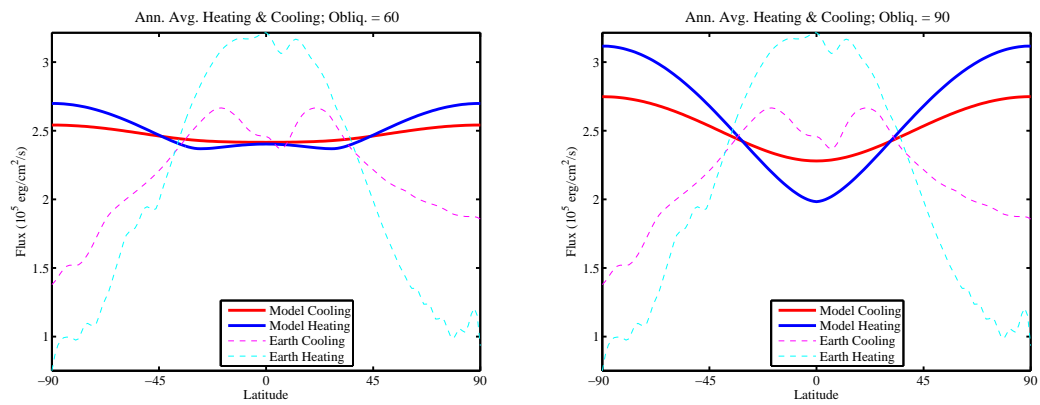
Figure 6.5 Annually averaged cooling and heating fluxes for high and extreme obliquity model planets model and for Earth. Analogous to Fig. 6.5, which shows the annually averaged heating and cooling as functions of latitude for 23.5° obliquity, here we present the same functions for Earth-like models at 60° obliquity (left panel) and at 90° obliquity (right panel). Thick lines are model results; thin dashed lines are ERBE data. For both high obliquity cases, unlike on Earth, there is net annually averaged heating at the poles (i.e., heating exceeds cooling), and, especially for the extreme obliquity case, net annually averaged cooling at the equator.

the high effective heat capacity of the atmosphere above ocean (recall that in these models, 70% of the surface area in every latitude band is ocean). In models with less ocean coverage, or oceans that are nonuniformly distributed, the cooling too can vary dramatically over the annual cycle.

6.4.1.2 Implications for Habitability

As SMS08 demonstrates, model planets with efficient heat transport (slowly spinning planets) are more nearly isothermal than models with Earth-like rotation, which themselves exhibit less latitudinal variation of temperature than those with inefficient heat transport (fast spinning planets). As a result, models corresponding to slow-spinning worlds tend to be either entirely habitable or entirely nonhabitable at any given time. In contrast, Earth-like and faster spinning models may be only partially habitable at a particular time. They may, for instance, be frozen at the poles and temperate at the equator, or vice versa in the case of a highly oblique world.

Figure 6.8 demonstrates the kind of complicated interplay that can go on between obliquity and efficiency of heat transport, in determining a planet's habitability. This figure shows the temporal habitability fraction, as a function of orbital semimajor axis and latitude, for each of 12 different combinations of obliquity ($0°$, $30°$, $60°$, $90°$) and diffusion coefficient ($D_{\mathrm{fid}}/9$, D_{fid}, and $9D_{\mathrm{fid}}$, corresponding respectively to 72-hour, 24-hour, 8-hour rotation). The top panel shows these plots for the (I_2, A_2) pair, and the bottom panel depicts the (I_3, A_3) pair. In both panels, the
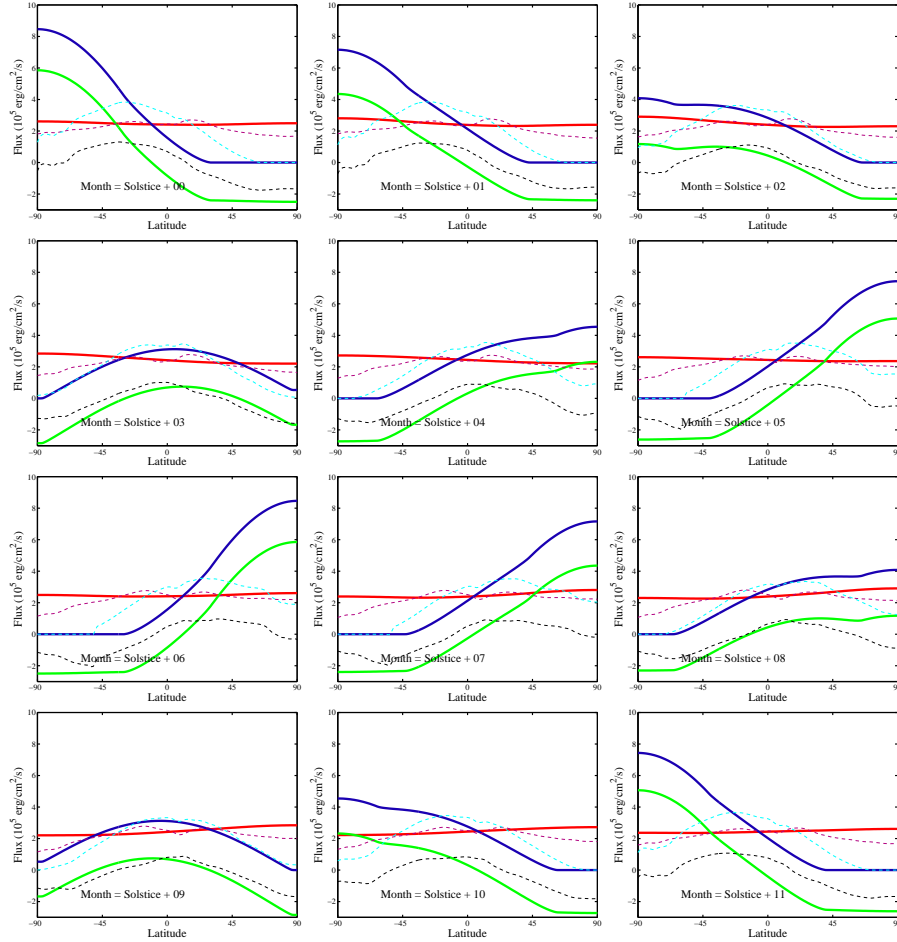
Figure 6.6 Monthly cooling, heating, and net fluxes for fiducial model at 60° obliquity, at 1 AU, and for Earth. Each panel presents the average cooling, heating, and net (heating minus cooling) radiative fluxes, as functions of latitude, for one month of the year, starting at the Northern winter solstice (upper left panel), and incrementing by one month with each panel to the right. These fluxes are presented for both the model (thick solid lines) and the Earth (thin dashed lines). Model heating is blue; model cooling is red; model net heating is green. Earth heating is cyan; Earth cooling is deep magenta; Earth net heating is black.
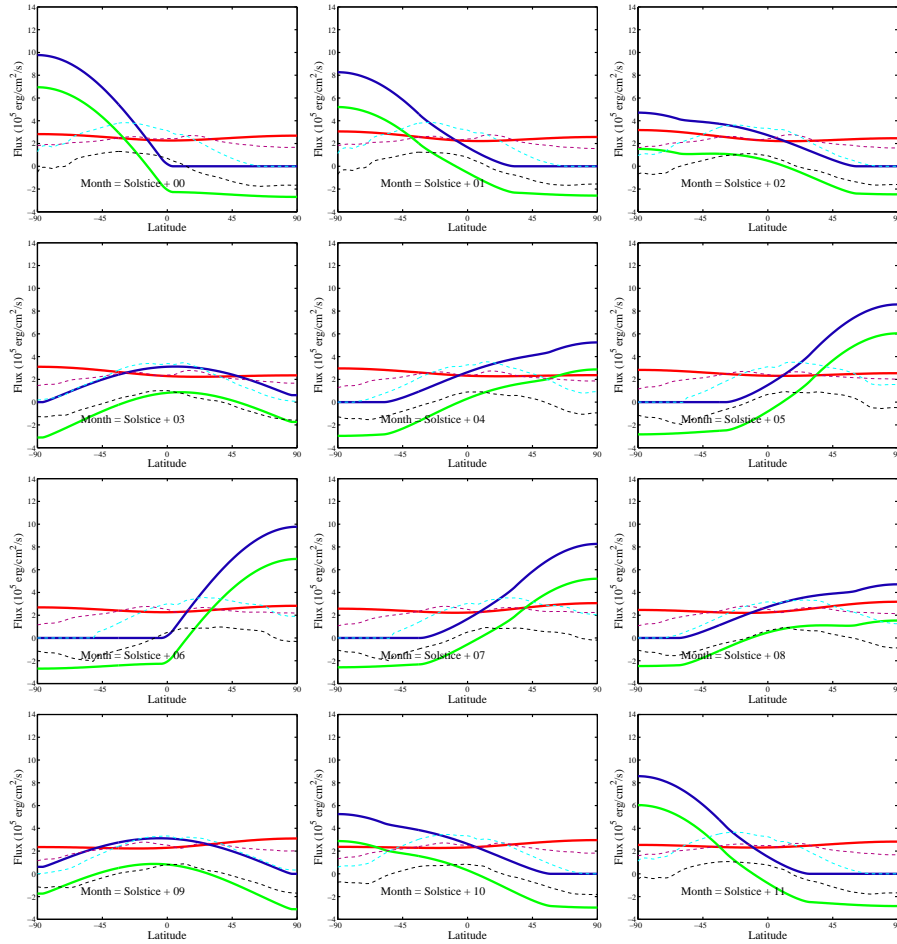
Figure 6.7 Monthly cooling, heating, and net fluxes for fiducial model at 90° obliquity, at 1 AU, and for Earth. Each panel presents the average cooling, heating, and net (heating minus cooling) radiative fluxes, as functions of latitude, for one month of the year, starting at the Northern winter solstice (upper left panel), and incrementing by one month with each panel to the right. These fluxes are presented for both the model (thick solid lines) and the Earth (thin dashed lines). Model heating is blue; model cooling is red; model net heating is green. Earth heating is cyan; Earth cooling is deep magenta; Earth net heating is black.

left column of plots represents efficient transport; the middle column represents Earth-like transport; and the right column represents inefficient transport. Each of the 24 plots in this figure shows the results of model runs from 0.45 AU to 1.25 AU in increments of 0.025 AU, with the color of each point indicating the fraction of the year that the latitude at that point spends in the habitable temperature range (273 K - 373 K) on a model planet at the specified orbital semimajor axis. In each plot, the white vertical dashed lines indicate the radiative equilibrium habitable zone, calculated (as discussed in SMS08) for a 0-dimensional model planet with annually averaged, globally averaged insolation.

There are a number of intriguing features of this figure. The most obvious is that, as expected, at every obliquity, less efficient transport results in more strongly latitudinally differentiated temporal habitability. In addition, at each transport-efficiency value, the > sign shape of the seasonally habitable ribbon at low obliquity reverses to something more like a < sign at high obliquity. In other words, at low obliquity, the relatively cold poles are habitable closer to the star and the relatively warm equator is habitable farther from the star, and at high obliquity, this reverses and the poles are relatively warm, while the equator is relatively colder.

Furthermore, notice that, in both panels, the plots in most cells show a very abrupt outer boundary to the seasonally habitable zone. This is because, as discussed in SMS08, the ice-albedo feedback renders these models quite sensitive to changes in forcing. Small reductions in insolation can become amplified, because the ice-coverage increases, which increases the global albedo and leads to further
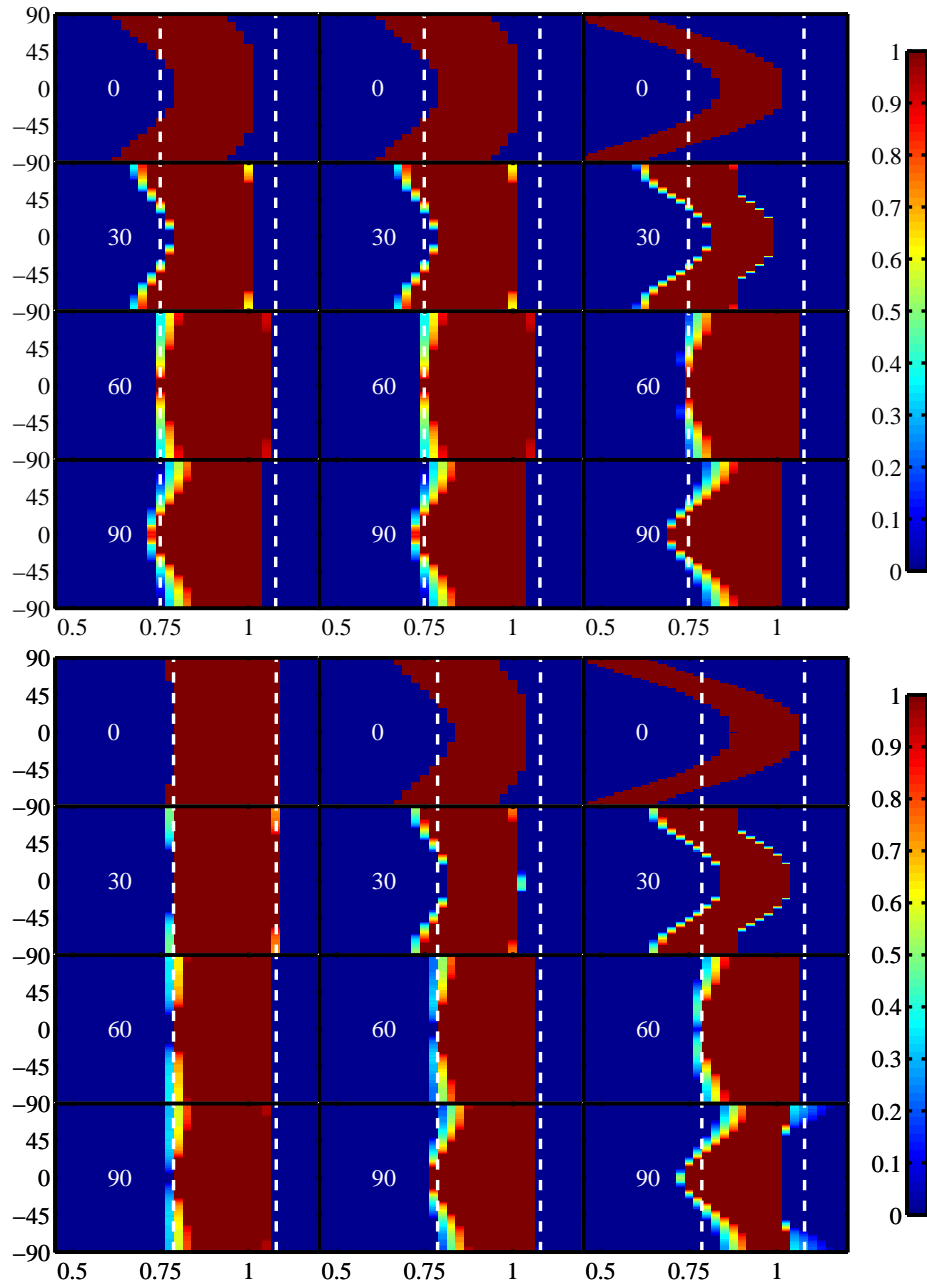
Figure 6.8 Model temporal habitability fraction under different obliquities, rotation rates, and cooling/heating functions. In both panels, obliquity varies from 0° (top row) to 90° (bottom row) and diffusion coefficient varies from $(1/9)D_0$ (left column) to $9D_0$ (right column). The x-axis of each cell is orbital radius, in AU, and the y-axis is latitude. The color of each point indicates the fraction of the year that that part of parameter space spends in the habitable temperature range (273 K - 373 K). *Top Panel: I_3, A_3 combination. Bottom Panel: I_2, A_2 combination.*

reduction in insolation. This feedback mechanism renders these models suscep-tible to succombing to a global snowball state, from which they cannot recover within our model framework. The main exceptions to this trend are the low obliq-uity, fast-spinning models in the upper right corners of both panels (although even these models drop to 0% habitability at orbital radii that are small relative to the outer boundary of the radiative equilibrium habitable zone, indicated by the white dashed lines). Interestingly, the fairly small change in cooling/heating functions from I_2, A_2 to I_3, A_3 allows the cell in the lower right corner of the bottom panel – extreme obliquity, inefficient transport – to avoid snapping to a snowball in a single step in orbital radius. In that model, the intense summer insolation at the poles, combined with the relative thermal isolation of different latitudes, allows the poles to heat up above the freezing point of water during their summers, even at orbital distances where other models would be entirely frozen. In sum, the susceptibility to snowball conditions is sensitive to details of parameterizations.

6.4.2 Land/Ocean Distribution

As described in SMS08, the large covering fraction of oceans on the Earth (roughly 70%) stabilizes our climate over an annual cycle, by virtue of the large effective heat capacity of atmosphere over ocean: over land, the thermal relaxation timescale is several months; over the 50 m mixing layer of the ocean, the thermal relaxation timescale is more than a decade. As a result, in a 1D model (such as ours) that does not resolve continents in longitude, any latitude band with significant ocean

fraction will have strongly suppressed annual temperature fluctuations relative to a latitude band with low ocean fraction. Because we do not know of any way to predict the distribution of continents and oceans on an extrasolar planet, it is important to consider the influence on climatic habitability of other possible distributions.

6.4.2.1 Nonuniform Ocean Coverage

We consider several planet models with distributions of land and ocean that are not uniform across different latitudes: one with 30% land coverage, with the landmass centered on the North Pole (extending down to ~ 24° North latitude), and the other (discussed in § 6.4.2.2) with 90% land, again centered on the North Pole.[4] Because of the relatively low thermal inertia of atmosphere over land, parts of a model planet that are dominated by land can freeze or boil during the course of the year and still return to temperate conditions at other times. In fact, at some orbital distances, and at high obliquity, the polar regions of some models freeze *and* boil within an annual cycle.

Figure 6.9 displays the tremendous swings of temperature that can occur over latitude bands that lack ocean, and also indicates that annually averaged calculations may miss a lot of the story of what conditions are like on planets. This figure shows the annually averaged temperature, and the temperature evolution, on a model planet with a North Polar continent that is 30% of the total surface area, at 1 AU. This model uses I_2, A_2 cooling/heating, and shows results with obliquity

---

[4]It is more natural to conceive of this model as having an ocean centered at the South Pole.

23.5°, 60°, and 90°. At all three obliquities, the left column – the annually averaged temperature profile – provides an impoverished view of the actual climatic conditions. Looking at just the left panels: the 23.5° obliquity model appears slightly asymmetrical in temperature distribution, with the continental North Pole 8 K warmer than the oceanic South Pole; the 60° obliquity model appears cooler at the continent pole; and the 90° obliquity model again appears warmer at the continent pole, but appears frozen over the whole globe. In truth, all three models reach significantly higher temperatures at the continent pole during its summer than at the other pole. At both 60° and 90° obliquity, North Pole summer temperatures exceed 410 K, as the Sun shines nearly straight down on the pole for months. An obvious puzzle is implicit in this figure: Although we may not have much intuition for what the polar summers should be like on high obliquity planets, it is surprising to see a prediction of summer polar continent temperatures in excess of 310 K in the 23.5° obliquity model, given that Antarctica – Earth's continental pole – is significantly *colder* than the non-continental pole, and for the most part neither pole ever reaches temperatures above freezing. Accounting for Antarctica in our model framework is not easy, and we address this issue in greater detail in § 6.5.1, but our sense is that the resolution of this issue may have to do with model initial conditions, or, in the case of the Earth, with the longterm history of temperature evolution.

Figure 6.10 presents plots of the temporal habitability fraction for the world just considered, at obliquities 0°, 30°, 60°, 90°. We see in Fig. 6.9 that the presence
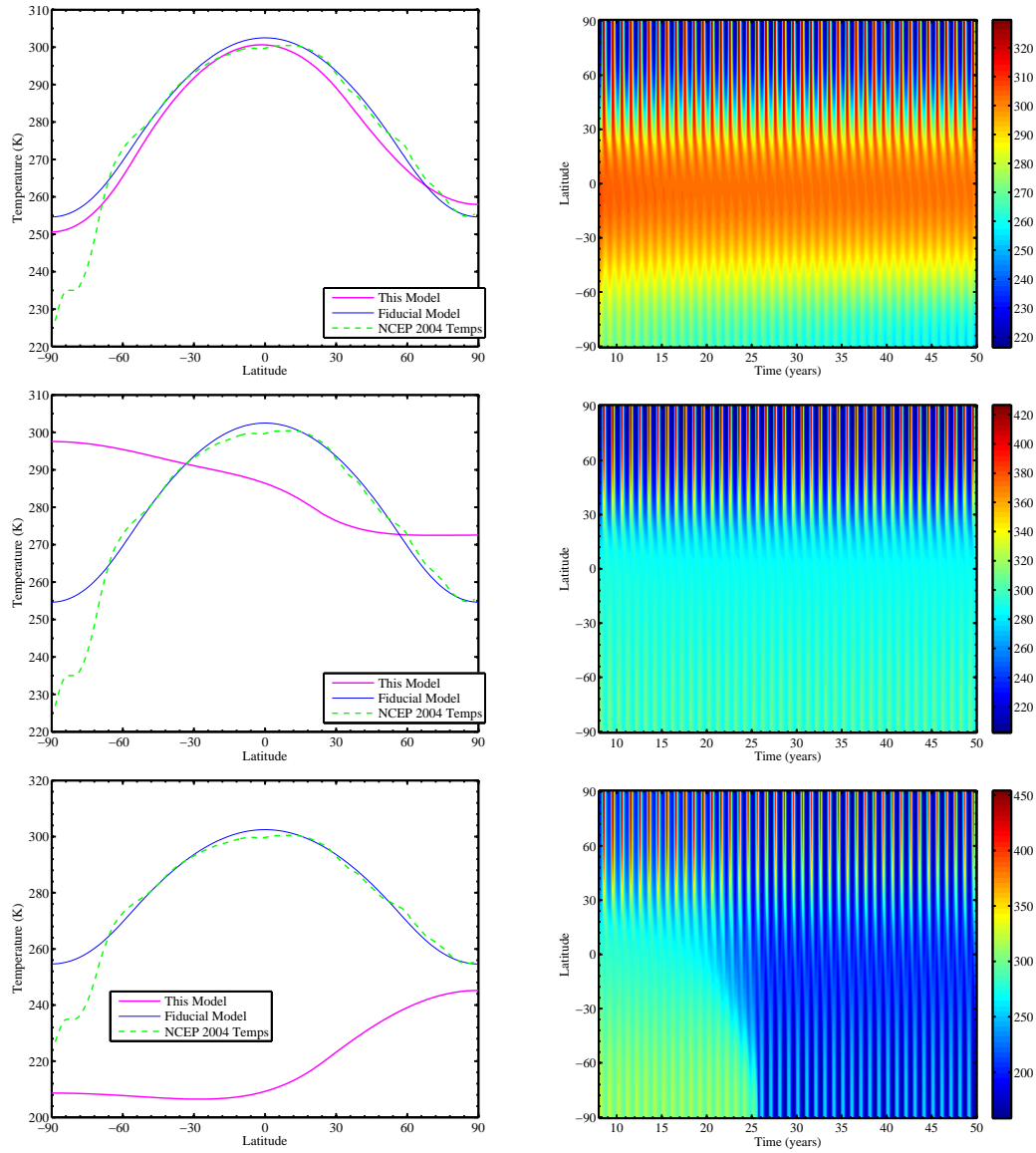
Figure 6.9 Annually averaged and space-time plot of temperatures on models with a North Polar continent that takes up 30% of surface area (the other 70% is ocean) at 1 AU, cooling/heating = I_2, A_2, obliquity = 23.5°, 60°, 90°. *Left:* The magenta curve shows the annually averaged temperature profile for models with North Polar continents that extend down to ~ 24° North latitude. The blue and dashed green curves are for reference – blue: the fiducial model (identical to this one, except the 70% ocean is uniformly distributed in every latitude band); dashed green: the Earth's actual temperature profile, measured by NCEP/NCAR in 2004 (Kistler et al. 1999; Kalnay et al. 1996). *Right:* Latitude/time plot of temperature evolution, from model-years 8 through 50, for the polar continent models. *Top Row:* 23.5°. *Middle Row:* 60°. *Bottom Row:* 90°.

of land at the North Pole causes tremendous swings in temperature there; we see in Fig. 6.10 that at nonzero obliquity this is the case at other values of semimajor axis, too. These large seasonal variations lead to exotic shapes in plots of temporal habitability. Compared with uniformly ocean-dominated worlds, much more of the parameter space at each obliquity except 0° is habitable neither 0% nor 100% of the year, but somewhere in between.

6.4.2.2 Desert Worlds

We now consider two model planets with just 10% ocean fraction. We examine the cases of both uniformly distributed ocean (10% in every latitude band) and ocean concentrated at the South Pole (extending northward to ∼ 53° South latitude). Figure 6.11 presents the temporal habitability for the uniform desert world, and Figure 6.12 presents the analogous plot for the South Polar ocean world. We again see regions of some model planets that swing from freezing to boiling temperatures over the course of the year. This is responsible for the butterfly shape of the temporal habitability plots in the 60° and 90° cells of Fig. 6.11: at $a \sim 0.9$ AU, the poles are habitable for less of the year than both the more equatorial regions at that orbital distance and than the poles at closer and more distant orbits.

These models, and those presented in § 6.4.2.1, suggest that at extreme obliquity the inner edge of the zone of regionally and seasonally habitable climates is extended dramatically inward, while the outer boundary is extended only mildly outward. A caveat that should accompany this observation is that assuming a
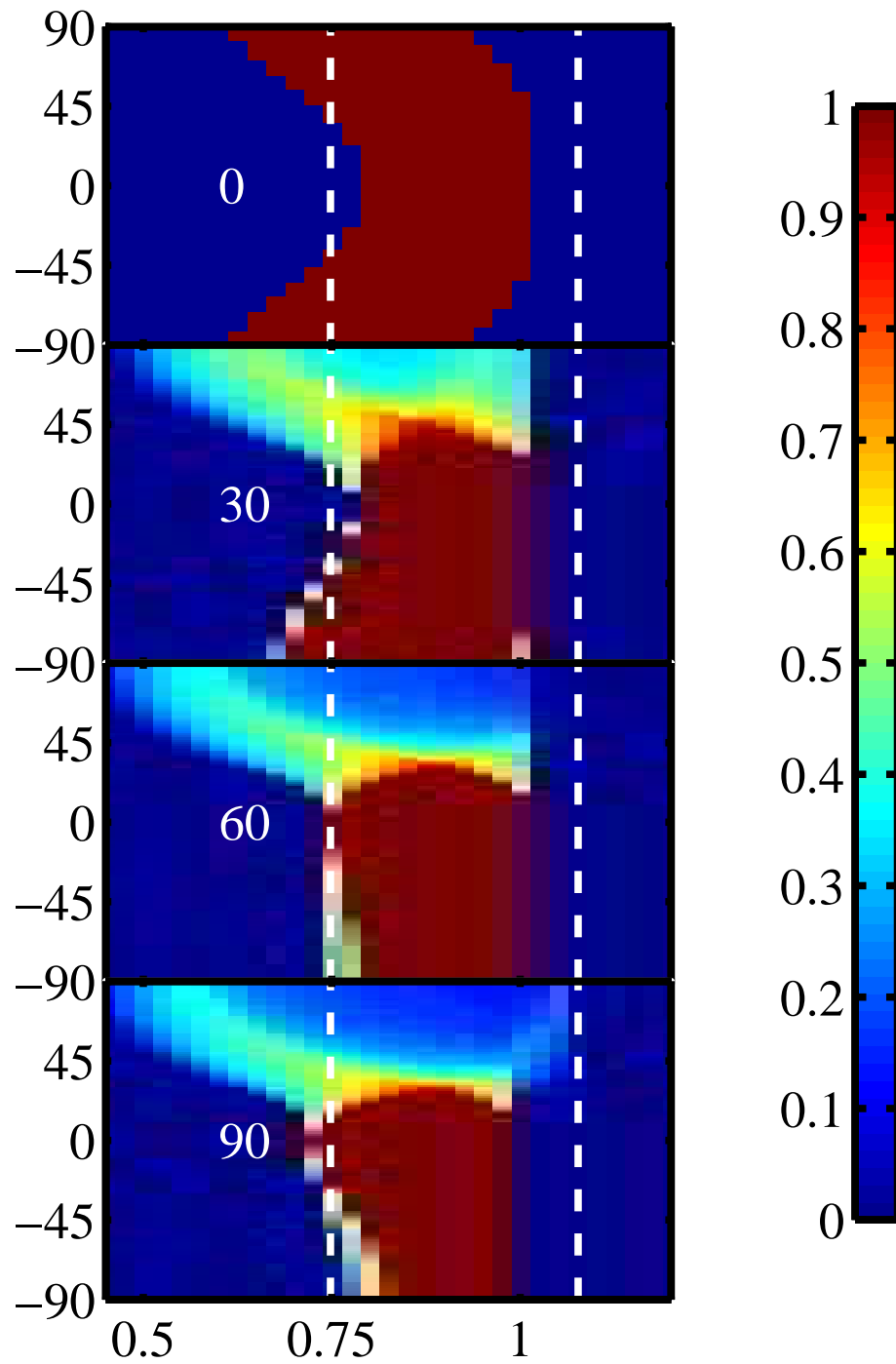
Figure 6.10 Model temporal habitability fraction under different obliquities, North Polar continent covering 30% of surface. The diffusion coefficient is $D = D_0$. Similar to Figure 6.11.
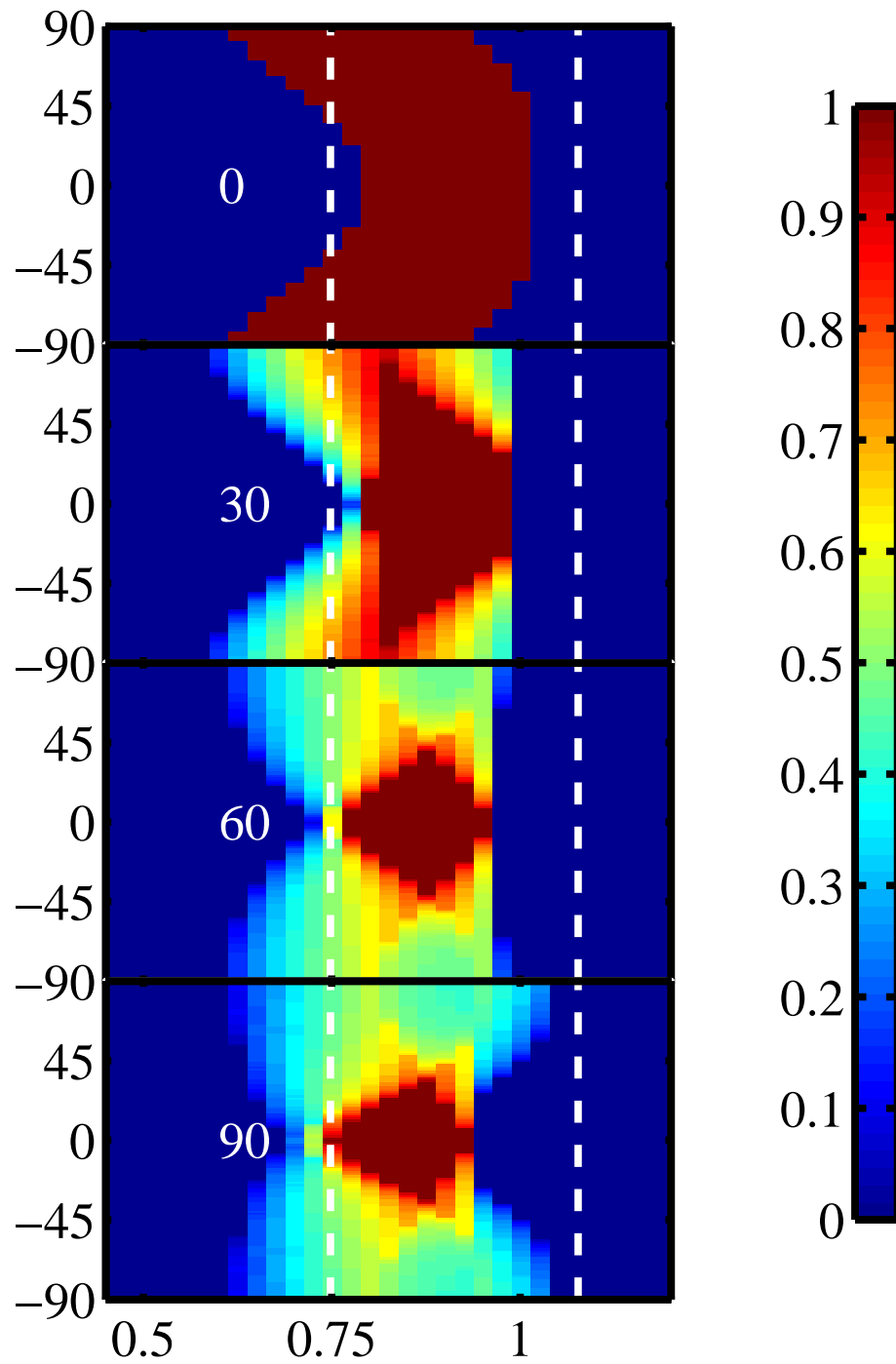
Figure 6.11 Model temporal habitability fraction under different obliquities, 10% ocean uniformly distributed. The diffusion coefficient is $D = D_0$. Similar to Figure 6.8. The color of each point indicates the fraction of the year that that part of parameter space spends in the habitable temperature range (273 K - 373 K).
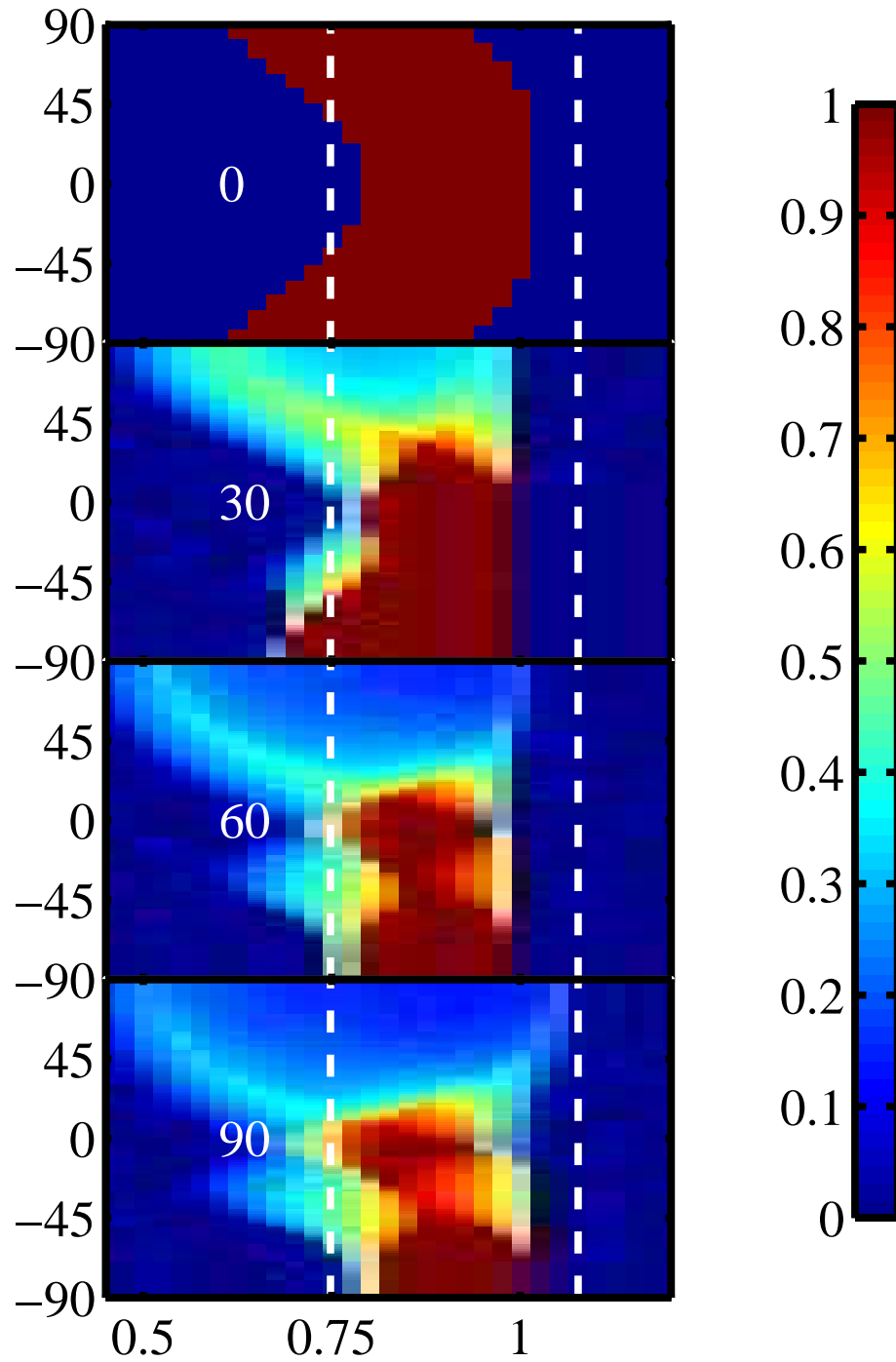
Figure 6.12 Model temporal habitability fraction under different obliquities, North Polar continent covering 90% of surface (South Polar ocean). The diffusion coefficient is $D = D_0$. Similar to Figure 6.11.

cooling function that is constant with orbital radius probably leads to a flawed treatment of both the high and low insolation limits of these models. At the inner edge of the habitable zone, large increases in atmospheric water content can cause a reduction in the cooling function that causes a runaway greenhouse effect; the eventual catastrophic water loss can result in Venus-like outcomes, as described by Kasting et al. (1993) and references therein (although this type of outcome might be mitigated by a reduction of the heating function due to increased cloud-albedo, as mentioned in SMS08). At the outer edge, reduced efficiency of the carbonate-silicate weathering cycle is likely to lead to a significant increase the partial pressure of CO_2 (Kasting et al. 1993), which could extend the habitable zone from ~ 1 AU in our models to ~ 1.4 AU or more in some cases. We discuss in greater detail the effect of a cooling function that varies with orbital distance in § 6.4.3.

Still, for most plausible cooling functions, the low thermal inertia of atmosphere over land might lead to severe polar climates in highly oblique models. What are we to make of partial "habitability" by our criterion in the case of a region of a planet that actually boils and freezes every year? There are some microbes on Earth that can reproduce at freezing temperatures, and others that can reproduce at boiling temperatures, although none of which we are aware that can do both. If a part of a planet regularly swings through these wild extremes of climate, is it fair to call it habitable? This is an open question, but it is worthwhile to keep in mind that microbes on Earth appear to be as hardy as they need to be: nearly everywhere that biologists have searched, they have found some microbes thriving. Perhaps

most significant from the perspective of habitability, the reduced thermal inertia of these models appears to render them less susceptible to global snowball events.

6.4.3   Modeling the Far Reaches of the Habitable Zone

Walker et al. (1981) propose that a planet's temperature is regulated on long timescales by a feedback mechanism involving weathering of silicate rocks through carbonic acid from CO_2 disolved in water. They argue that since the rate of weathering (and hence of removal of CO_2 from the atmosphere) increases with temperature, this process is an important negative feedback on climate that acts to keep temperatures near the freezing point of water. Kasting et al. (1993) point out that this negative feedback can significantly offset the extreme sensitivity of climate to changes in orbital distance away from 1 AU seen in models such as those of Hart (1979) and the models presented in SMS08 and thus far in this paper. These models are so sensitive because they contain a significant positive feedback of the Earth-climate – the ice-albedo feedback whereby at lower temperatures the absorbed insolation is dramatically reduced because of the high albedo of ice – but they ignore the counterbalancing negative feedback of the carbonate-silicate weathering cycle.

In order to probe the combined influence on climate of rotation rate and obliquity in the context of the expected CO_2-rich atmosphere that a pseudo Earth would have at 1.4 AU, we switch to the cooling function I_{WK97} used in WK97, with CO_2 partial pressure (pCO_2) set to 1 bar and 2 bars; we maintain the simple

albedo function A_2. We find that at both levels of CO_2, model planets maintain globally temperate conditions at all obliquities for both $D = 9D_0$ and $D = D_0$ – corresponding to slow and Earth-like rotation. Perhaps counterintuitively, at reduced transport efficiency, corresponding to fast rotation, these models are more susceptible to global glaciation. One might expect that the thermal isolation of different latitudes in these models, relative to slower spinning models, might allow parts of the models that receive less insolation to cool off without having much effect on other latitudes. Indeed, in the limit as $D \rightarrow 0$ this must happen. But with a small but nonvanishing diffusion coefficient, it turns out that the thermal near-isolation allows cold regions to become very cold, but this information is still communicated to other regions, dragging down their temperatures. Sometimes, this process can drag down global temperatures enough to plunge the models into global snowball state.

Figure 6.13 shows the global average temperature and the climate evolution for fast-spinning model planets at 1.4 AU with 1 bar atmospheric CO_2, with I_{WK97}, A_2 cooling/heating. At 23.5° obliquity, the cold temperatures at the poles drag the model into a snowball state; at 90° obliquity, the cold equator drags the model to the same fate; and at 60° obliquity, no part of the planet receives consistently low enough insolation to end up in this trap. Note that the snowball effect seen in the 23.5° and 90° obliquity models might be particularly calamitous because of the possibility of atmospheric collapse on a much shorter timescale than the vulcanism can replenish CO_2. At 1 bar, the freezing point of CO_2 is ~ 195 K. Both the

23.5° and 90° obliquity models reach temperatures below this threshold over large enough regions of their surfaces that significant amounts of the atmospheric CO_2 might condense out as dry ice, thereby reducing the greenhouse effect. The risk of atmospheric collapse would be somewhat lessened because of the latent heat of condensation, which would tend to prevent too much CO_2 from freezing out during any winter. Realistically treating this possibility would require incorporating a latent heat term in the energy balance equation, as Nakamura & Tajika (2002) do in their Mars EBM.

Figure 6.14 shows similar plots, but for a thicker atmosphere, with $pCO_2 = 2$ bars. In this case, all three obliquities avoid a snowball fate. Still, poles in the 23.5° model are below the freezing point of CO_2 at 2 bars (~ 203 K) for more than a third of the year, indicating that it is conceivable that this model might also risk losing its CO_2 greenhouse on a shorter timescale than it can be replenished.

6.5   Discussion and Conclusions

6.5.1   Whence Antarctica's Extreme Cold?

As discussed in § 6.4.3, many of the models presented here suffer from several limitations, including overly simplified treatments of both the inner and outer edges of the habitable zone. An additional puzzle, suggested in § 6.4.2.1, is to figure out why our Earth-like obliquity models with polar continents have such different climate from the Earth's Solar Polar continent of Antarctica. There are

Figure 6.13 Annually averaged and space-time plot of temperatures on fast spinning world at 1.4 AU, WK97 cooling function, $pCO_2 = 1$ bars, obliquity = 23.5°, 60°, 90°. Same as Figure 6.9 except cooling function is I_{WK97} with $pCO_2 = 1$ bars, and $D = (1/9)D_0$.
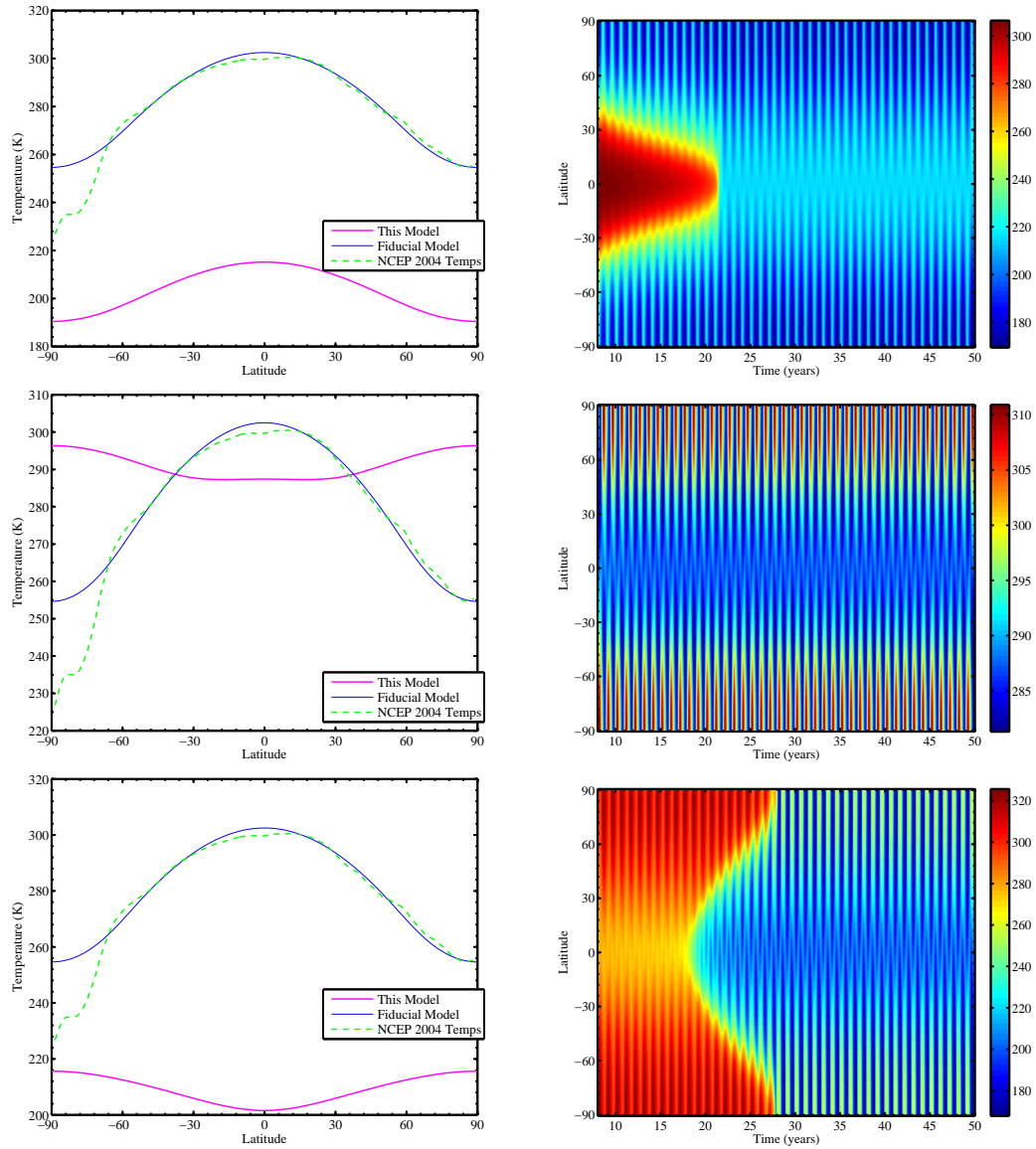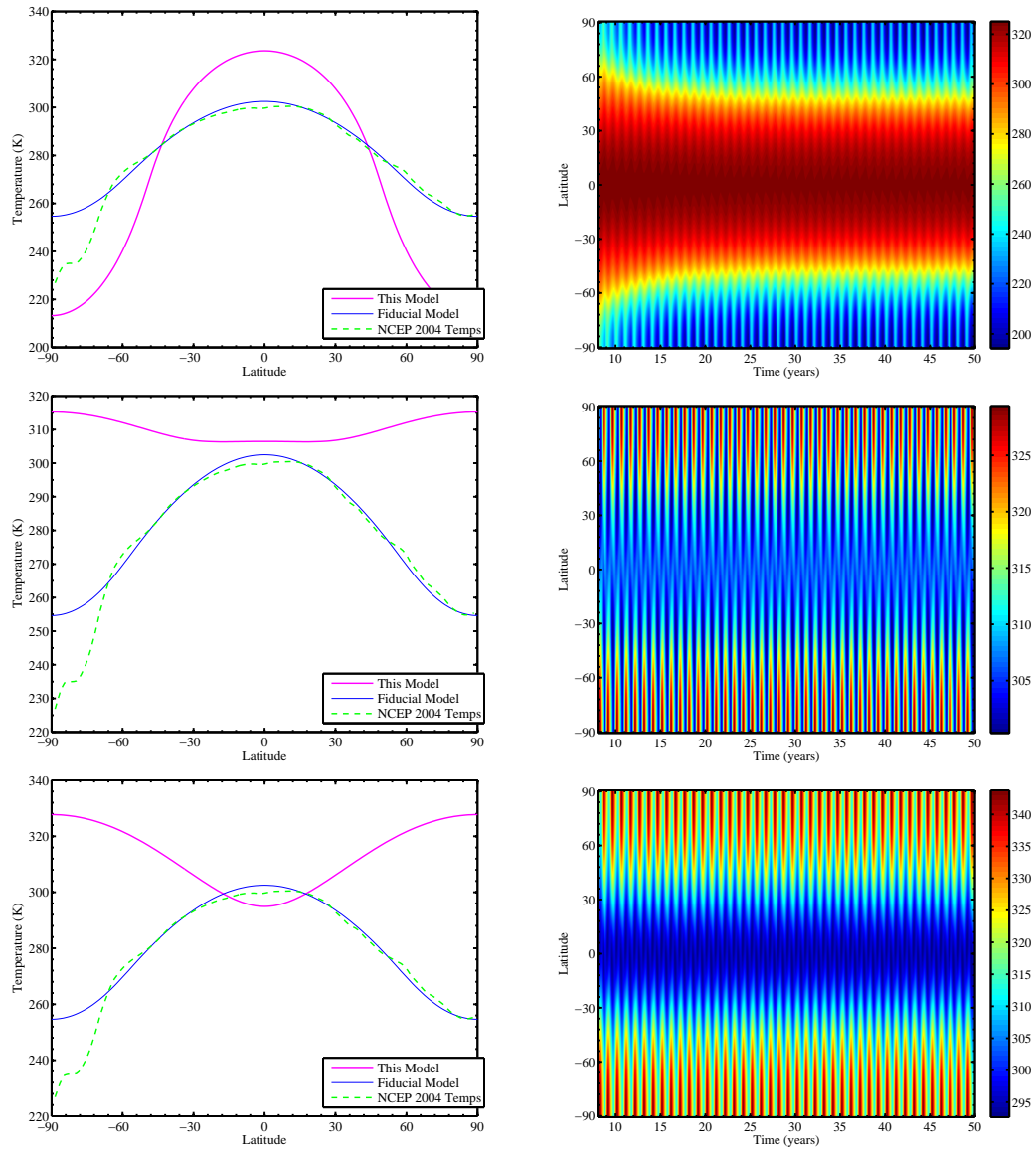
Figure 6.14 Annually averaged and space-time plot of temperatures on fast spinning world at 1.4 AU, WK97 cooling function, pCO_2 = 2 bars, obliquity = 23.5°, 60°, 90°. Same as Figure 6.14 except pCO_2 = 2 bars.

various reasons why Antarctica is so cold. One obvious and important factor is that it is covered with over a mile of ice. This guarantees that on short timescales anyway, regardless of the temperature the albedo is always high. In actuality, the enforced high albedo of course helps to prevent the temperature from exceeding the freezing point of water. In an attempt to simulate the Earth's sharply colder temperatures at the South Pole, we therefore set up several models in which we impose ice at the South Pole, regardless of temperature. This prescription affects the albedo, forcing it to be constanly high, and the heat capacity, increasing it above the heat capacity for atmosphere above pure land by a factor of 2.

Figure 6.15 shows the results of two attempts to duplicate the temperature at Antarctica within our models. In both models, there is a South Polar ice continent that extendes to 70° South latitude (~ 3% the total surface area), and the rest of the planet is $\approx$ 72.2% covered with ocean, which combine to make an overall ocean fraction of 70%. The top panels of this figure use the standard A_2 albedo function everywhere. Clearly, the high albedo (constantly 0.77) at the ice-continent has only a very mild effect on that continent's temperature. The bottom panels use a modified albedo function: $A_2[T]$ everywhere except the simulated Antarctica, and "shiny ice"(A = 0.88) on the South Polar continent. This prescription forces the temperature at the ice continent somewhat colder than the other continent, but through the diffusion term the cooler South Polar temperature is communicated to the rest of the planet, too, eventually supressing the North Polar temperature by more than 10 K. Using even shinier ice at the South Pole can, when using the I_2

function, drive the planet to a snowball state.

6.5.2 Applicability of Global Radiative Balance Calculations

Historically, calculations of habitable zones have often assumed global radiative balance conditions. Although these calculations by definition cannot account for the regional character of habitability, one might hope that they would still provide a decent proxy for the global average of actual conditions. Indeed, as seen in Fig. 6.3, the Earth itself is within ~ 5% of radiative equilibrium throughout the year. Accordingly, global radiative balance models have provided a very useful starting point for considerations of how habitability depends on orbital radius.

Figure 6.16 presents the globally averaged cooling, heating, and net radiative fluxes for a model with a North Polar continent that covers 30% of the planet's surface, at 60° and 90° obliquity. The 60° obliquity model gets up to nearly 40% out of global radiative balance at some times during the year, and the 90° obliquity model reaches nearly 60% out of global radiative balance! It has always been recognized that planets on highly eccentric orbits experience forcings that are significantly different from annually and globally averaged conditions, but our here results show that, even on circular orbits, planets can experience conditions that are far from radiative equilibrium. This further underscores the importance of regional, time-dependent models for addressing the habitability of extrasolar terrestrial planets.
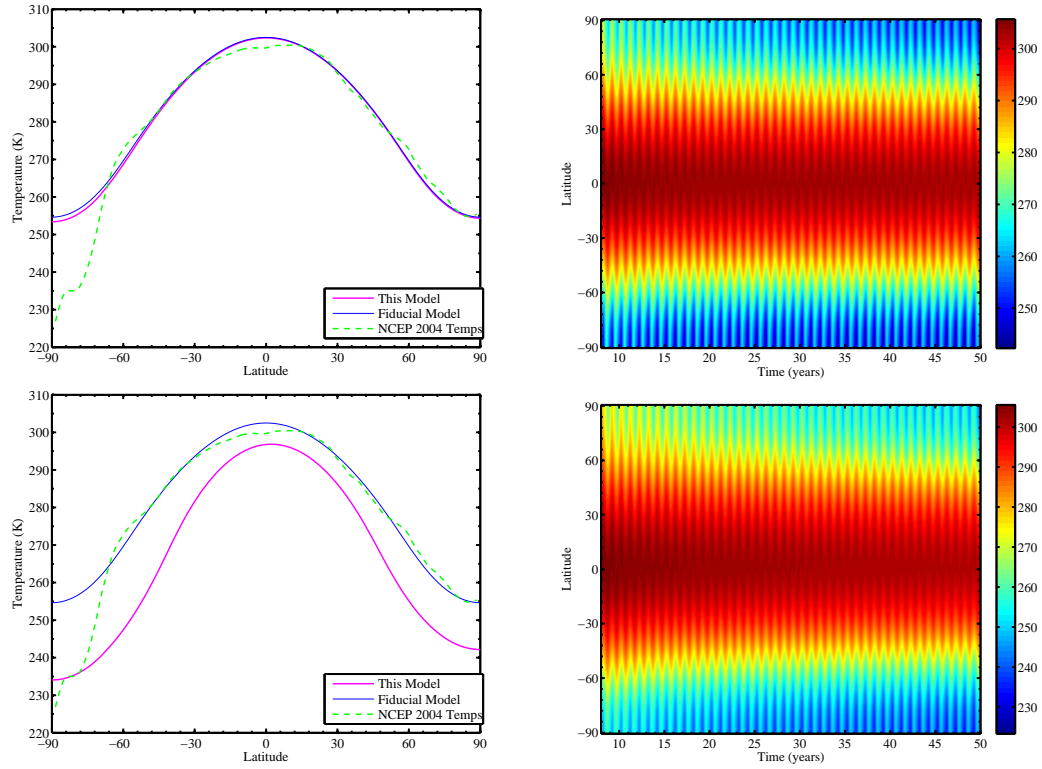
Figure 6.15 Annually averaged and space-time plot of temperatures on models with Antarctica analogs, at 1 AU, cooling/heating = I_2, A_2, obliquity = 23.5°. These models have 70% ocean coverage overall, with a South Polar ice continent extending up to 70° South latitude, and $\approx$ 72.2% ocean fraction over the rest of the planet. *Top Row:* Normal ice (A = 0.77) at South Pole. *Bottom Row:* "Shiny ice" (A = 0.88) at South Pole. Similar to Figure 6.9.

Figure 6.16 Global average cooling, heating, and net radiative flux, as functions of time in model with North Polar continent at 60° and 90° obliquity. Similar to Figure 6.3. *Top two panels:* 60° obliquity; *Bottom two panels:* 90° obliquity.
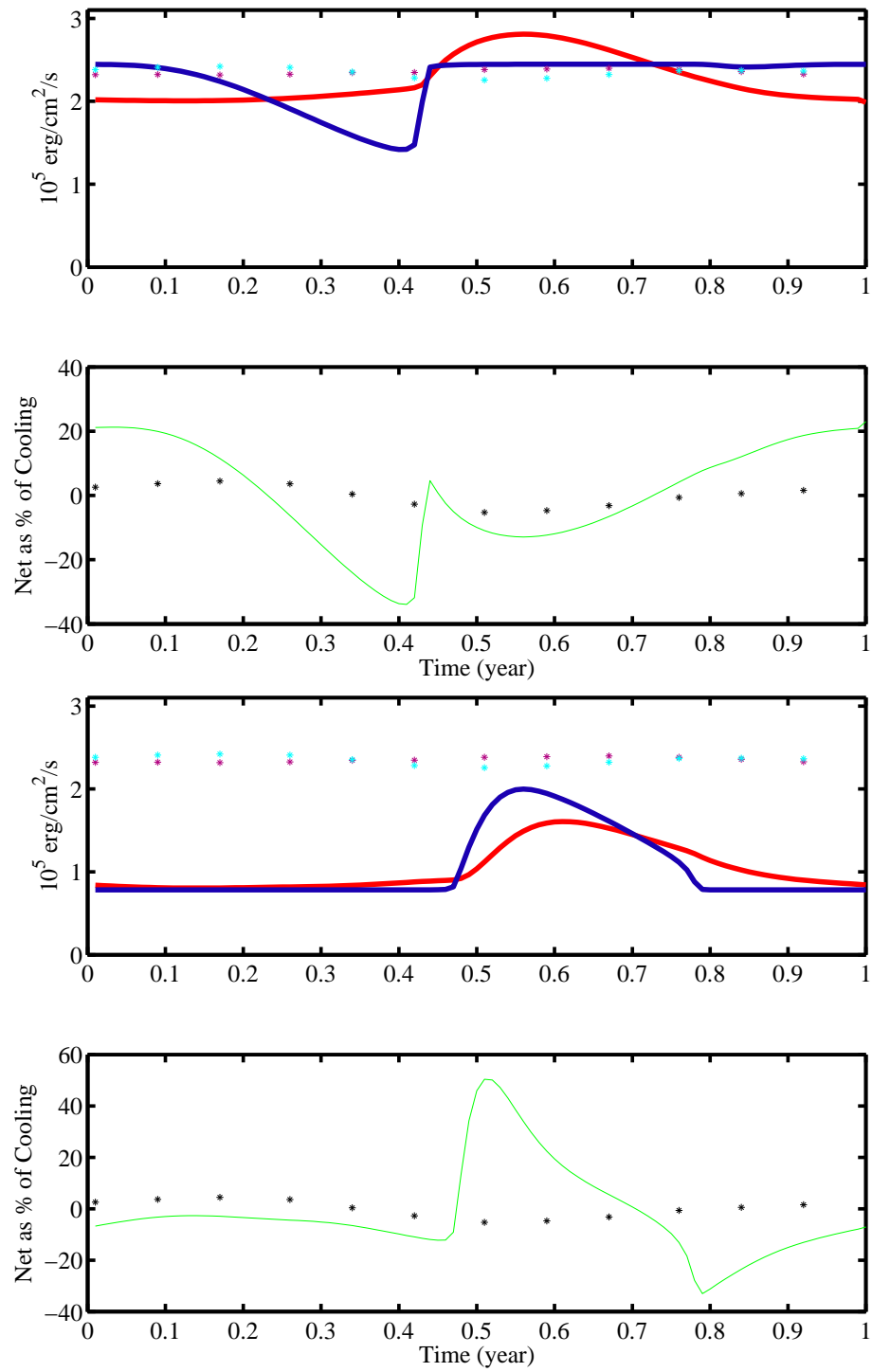
6.5.3   Summary

We have presented a series of 1D EBMs to address the variety of possible climatic conditions that might result on extrasolar terrestrial planets. We considered forcings determined by a variety of unobservable planetary attributes, including obliquity, rotation rate, the distribution of land/ocean coverage, and the detailed nature of the radiative cooling and heating functions. Previous investigations have found that high obliquity does not necessarily render planets more susceptible to global glatiation. Our results provide preliminary indications of an even more promising result: models at high obliquity often appear to be less prone to snowball states, which is a result that appears to hold for a wide variety of other properties. Counterintuitively, faster rotating Earth-like planets may fall victim to snowball glaciation events at closer orbital radii than slower rotating planets; this result appears to hold for various cooling functions, including both Earth-like infrared cooling and the expected radiative cooling for the kind of massive CO_2 atmosphere that Kasting et al. (1993) predict to obtain in the outer parts of habitable zones. We find that planets with small ocean fractions or polar continents can experience very severe seasonal climatic variations, but that these planets also might maintain seasonally and regionally habitable conditions over a larger range of orbital radii than more Earth-like planets. We furthermore find that high obliquity models with nonuniform distributions of land and ocean experience forcing conditions that are very far from global radiative balance relative to what the Earth experiences. This last finding provides a substantial justification of the present time-dependent,

regional model.

Chapter 7

Conclusion: Looking Forward

This thesis addresses several properties of extrasolar planets that are beyond present observational capacities. It is interesting to think about what attributes will be considered cutting edge or nontrivial in the future, especially since, although nearly 300 planets are now known around other stars, just two dozen years ago none were. At that time, the mere *existence* of exoplanets was an extremely nontrivial property. The next several decades will see many more, increasingly sophisticated, discoveries. In concluding, I will speculate about a few of these.

7.1   Magnetic Planets and Moons?

As more and more terrestrial planets are discovered, improving our understanding of what planetary attributes are conducive to habitability will be an ever more important goal. Climatic habitability will surely continue to draw attention, and

deservedly so, because of the clear influence on climate (temperature) of orbital distance, one of the main observable properties of exoplanets. But what about other aspects of habitability? The high energy particles in the solar wind might prevent the formation of complex biological molecules if it were not for the Earth's magnetic field, which acts as a shield. Indeed, Dehant et al. (2007) and others have argued that a planet might require a magnetic field in order to develop and maintain life as we know it. But how to measure the magnetic field of an extrasolar planet?

The gas giant planets of our Solar System all exhibit strong, low frequency (kilometric to decametric) radio emission (Zarka 1992), which has prompted researchers to search for radio emission from extrasolar giant planets. Although Lazio & Farrell (2007) reported that observations of the τ Bootis system with the Very Large Array yielded no detection, Farrell et al. (2004) and Grießmeier et al. (2005) have suggested that finding evidence of a magnetic field around an exo-Jupiter might be within the reach of the Low Frequency Array (*LOFAR*). Such a detection would be exciting but would bear no direct relevance to astrobiology, at least not to the habitability of terrestrial planets.

More germane from an astrobiological perspective: it might not be beyond the capability of future astronomers to detect radio signals from even an Earth-like planet. A simple analysis indicates that, with the right instrument, radio wavelength radiation may be among the easiest types of radiation to detect from an extrasolar planet.[1] It is in general very difficult to detect light from an extrasolar

[1]The right instrument – a space-based radio array, as described below – is stunningly simple

planet: a solar–type star is brighter than a hot–Jupiter by a factor of at least $\sim 10^4$ at optical wavelengths. The ratio of starlight to planet–light is somewhat smaller (a few times 10^3) at infrared wavelengths, but still only a small fraction of the light that we receive is from the planet. Although emergent planetary radiation has been detected in several transiting systems, the technical challenge of detecting emitted or reflected radiation from even a hot–Jupiter remains great. Optical and infrared radiation from an Earth-clone orbiting a Sun-like star at 1 AU would be $\sim 10^4 - 10^5$ times dimmer than that from a hot Jupiter. In parts of the radio spectrum, however, the star–planet contrast is less daunting. More than three decades ago, Gurnett (1975) discovered that the Earth is a very strong radio–emitter at long wavelengths. There are several mechanisms responsible for the Earth's kilometric radiation, the strongest of which is the cyclotron maser interaction of the solar wind with the magnetosphere at the poles, the same mechanism that is responsible for optical polar aurorae. Strong radio bursts are observed to be associated with bright auroral arcs, and so these flares are called Auroral Kilometric Radiation (AKR). During particularly strong flaring, the Earth shines at wavelengths from 1 to 30 km with integrated power of several billion watts (10^{16} erg s^{-1}), which makes Earth brighter than the combined output of the Sun and Jupiter by a factor of 10 or more. The bulk of this radiation originates at a height above the Earth's surface of between $0.75 R_{\oplus}$ and $5 R_{\oplus}$, with some emission detected out to $40 R_{\oplus}$ (Mutel et al. 2004).

from a physics perspective, though building it might pose significant technical and engineering challenges.

Because AKR is well below the plasma frequency of the ionosphere ($\sim$ 10 MHz), it does not penetrate to the ground, so an instrument to detect extraterrestrial AKR would need to be in orbit, above the ionosphere. An Earth–like planet around a Sun–like star could be expected to have similar emission properties to Earth's, if it had a similar magnetic field. Searching for such emission poses several difficulties that may turn out to be prohibitive, but the promise of such a favorable contrast ratio, together with the important complementary information that a detection of AKR from an exo-Earth would give us, indicate that a study of the feasibility of such an observation warrants further investigation. Janhunen et al. (2003) have proposed a space-based interferometric array for exactly this purpose. The good news is that the detectors would not need to be anything terribly sophisticated. Simply an array of kilometer-long wire antennas in space might allow synthesis-imaging of an exo-Earth. But even with an ideal radio array, could we hope to see Earth at 10 pc? The answer is unclear. Novaco & Brown (1978) provide a rudimentary characterization of nonthermal Galactic emission at very long wavelengths, but not much is actually known about the strength and spatial structure of this radiation. An immediate concern is that scattering by the interstellar medium might severely degrade the quality of an image, similar to how the Earth's atmosphere reduces the quality of seeing at optical wavelengths. In fact, Cohen & Cronyn (1974) and others characterize the wavelength dependence of interstellar scattering, and find that the apparent diameter of a source scales with square of the wavelength, at long wavelengths. With $\lambda \sim 1$ km, the formula of Cohen & Cronyn (1974) predicts that a point

source would be blurred to a ~ 1° disc, hopelessly spread out. Still, existing characterizations of interstellar scattering tend to be based on assuming a very-distant source. Within our local solar neighborhood, the scattering might be significantly less intense, and it might still be reasonable to hope to see extraterrestrial AKR from nearby Earth-like planets.

In the near future, it will probably be impossible to detect an Earth-like magnetic field around a terrestrial planet, but it is conceivable that detecting the magnetic field of an Earth-like moon of a giant planet will be possible with instruments that will come online in the near future. Williams et al. (1997) and Scharf (2006) have proposed that moons of giant planets might be hospitable locations for life, both in our own Solar System and in others. An icy moon far away from the star, such as Europa, might be able to protect sub-ice life from high energy stellar radiation without the help of a magnetic field; on the other hand, life beneath kilometers of ice in another solar system will probably be practically impossible to detect from Earth. But a moon in the star's habitable zone would presumably require a magnetic field to shelter its atmosphere and its inhabitants from its stellar wind. If a moon orbiting a magnetized giant planet possesses its own magnetic field, the interaction of the moon's and the planet's magnetic fields must influence the planetary decametric radiation in a similar way to how Io's magnetic field influences Jupiter's radio emission: it enhances power at the high-frequency (~ 20 MHz) end of the spectrum (Zarka 1992). Observations with *LOFAR*, and in the future with the Square Kilometer Array, may allow astronomers to detect the radio emission from

a magnetized moon orbiting a magnetized exo-Jupiter, if the radiometric signature of the moon would be a discernible variant of the emission from a planet that lacks a strongly magnetized moon.

7.2   Predicting Spectra with EBMs

The study of hot Jupiter planets is rapidly becoming a data-driven field. As *Spitzer* obtains more phase curves and secondary eclipse measurements, studies such as those of Burrows et al. (2007a) are becoming crucial to the interpretation of the emergent radiation of these planets. We are still several years away from having similar photometric or spectroscopic data for Earth-like extrasolar planets, but it is not too early to begin to think about setting up a framework for similarly deriving information about planetary structure from those future data.

The type of simple energy balance models discussed in Chapters 5 and 6 might prove to be an invaluable resource in interpreting future observational studies of terrestrial planets. The 1D (latitude only) nature of the EBM discussed earlier in this thesis is reasonable only to the extent that a planet may be said to have temperature that varies primarily with latitude, which for the most part is a justifiable assumption only for planets that spin much faster than they orbit. This assumption is therefore likely to fail for planets in the habitable zones of lower mass stars, which may be locked into synchronous or near-synchronous rotation; as a result, the 1D EBM I have already developed will not be appropriate for modeling them. It might be the case that the majority of habitable planets are orbiting stars significantly

less massive than the Sun; so it is imperative to develop realistic climate models of these planets. Joshi et al. (1997) and Joshi (2003) have run impressively detailed 3D global circulation models (GCMs) of tidally locked planets around M-dwarfs. As a critical complement to the work of Joshi (2003), I hope to modify my current 1D model to include a second spatial dimension. Although neglecting explicit calculations of vertical structure and of motions in the vertical direction will be a limitation of this approach, the advantages are twofold: I will both be able to (*i*) run the model at higher spatial resolution, and (*ii*) still run it efficiently enough to test the influence of both observable and unobservable planetary properties on regional habitability. Charbonneau & Deming (2007) have proposed monitoring all 10,000 M-dwarfs within 35 pc of Earth for transits. As this monitoring campaign proceeds and more such planets are discovered, the 2D EBMs that I will develop will be an essential tool for determining the likelihood that regions of these planets will be habitable.

Furthermore, there is an entire regime of habitability that has so far been largely ignored, between solar mass stars and the lowest mass stars. Around a K-dwarf, it is probable that planets will have habitable temperatures at distances at which they are not quite synchronous rotators, but instead might be rotating slowly, similar to Mercury. A database of 2D climate models, structured similarly to the investigations carried out so far in this thesis, could provide an important resource when considering possible climate structures on the terrestrial planets that will be discovered orbiting K-dwarf stars.

But how will a 1D or 2D EBM help with interpretative studies? In order to do so, it will be important to produce models that not only predict scarcely observable features such as habitability, but also predict features that are more easily observable (e.g., albedo, infrared luminosity, and strong spectral features). Tying the output of energy balance models to data that might become available with new observatories in the next decade will allow these models to be helpful tools not only for choosing targets of biosignatures-measuring telescopes of the distant future, but also for doing useful science in the more short-term future.

Bibliography

Adams, E. R., Seager, S., & Elkins-Tanton, L. 2007, ArXiv e-prints

Akahori, T. & Masai, K. 2005, PASJ, 57, 419

Akritas, M. G. & Bershady, M. A. 1996, ApJ, 470, 706

Albrow, M. D., Beaulieu, J.-P., Caldwell, J. A. R., Depoy, D. L., Dominik, M., Gaudi, B. S., Gould, A., Greenhill, J., Hill, K., Kane, S., Martin, R., Menzies, J., Naber, R. M., Pogge, R. W., Pollard, K. R., Sackett, P. D., Sahu, K. C., Vermaak, P., Watson, R., Williams, A., & The PLANET Collaboration. 1999, ApJ, 522, 1022

Alonso, R., Brown, T. M., Torres, G., Latham, D. W., Sozzetti, A., Mandushev, G., Belmonte, J. A., Charbonneau, D., Deeg, H. J., Dunham, E. W., O'Donovan, F. T., & Stefanik, R. P. 2004, ApJ, 613, L153

Andreon, S., de Propris, R., Puddu, E., Giordano, L., & Quintana, H. 2008, MNRAS, 383, 102

Arnaud, M., Pointecouteau, E., & Pratt, G. W. 2005, A&A, 441, 893

Ashton, C. E. & Lewis, G. F. 2001, MNRAS, 325, 305

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

Bakos, G. Á., Knutson, H., Pont, F., Moutou, C., Charbonneau, D., Shporer, A., Bouchy, F., Everett, M., Hergenrother, C., Latham, D. W., Mayor, M., Mazeh, T., Noyes, R. W., Queloz, D., Pál, A., & Udry, S. 2006, ApJ, 650, 1160

Bakos, G. A., Kovacs, G., Torres, G., Fischer, D. A., Latham, D. W., Noyes, R. W., Sasselov, D. D., Mazeh, T., Shporer, A., Butler, R. P., Stefanik, R. P., Fernandez,

J. M., Sozzetti, A., Pal, A., Johnson, J., Marcy, G. W., Winn, J., Sipocz, B., Lazar, J., Papp, I., & Sari, P. 2007a, ArXiv e-prints, 705

Bakos, G. Á., Noyes, R. W., Kovács, G., Latham, D. W., Sasselov, D. D., Torres, G., Fischer, D. A., Stefanik, R. P., Sato, B., Johnson, J. A., Pál, A., Marcy, G. W., Butler, R. P., Esquerdo, G. A., Stanek, K. Z., Lázár, J., Papp, I., Sári, P., & Sipőcz, B. 2007b, ApJ, 656, 552

Barkstrom, B. R., Harrison, E. F., & Lee, III, R. B. 1990, EOS Transactions, 71, 279

Barman, T. 2007, ApJ, 661, L191

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

Barnes, J. W. & Fortney, J. J. 2003, ApJ, 588, 545

Basri, G., Borucki, W. J., & Koch, D. 2005, New Astronomy Review, 49, 478

Batten, A. H., Fletcher, J. M., & Campbell, B. 1984, PASP, 96, 903

Baum, S. K. & Crowley, T. J. 2003, Geophys. Res. Lett., 30, 1

Beatty, T. G. & Gaudi, B. S. 2008, ApJ, 686, 1302

Beaulieu, J.-P., Bennett, D. P., Fouqué, P., Williams, A., Dominik, M., Jorgensen, U. G., Kubas, D., Cassan, A., Coutures, C., Greenhill, J., Hill, K., Menzies, J., Sackett, P. D., Albrow, M., Brillant, S., Caldwell, J. A. R., Calitz, J. J., Cook, K. H., Corrales, E., Desort, M., Dieters, S., Dominis, D., Donatowicz, J., Hoffman, M., Kane, S., Marquette, J.-B., Martin, R., Meintjes, P., Pollard, K., Sahu, K., Vinter, C., Wambsganss, J., Woller, K., Horne, K., Steele, I., Bramich, D. M., Burgdorf, M., Snodgrass, C., Bode, M., Udalski, A., Szymański, M. K., Kubiak, M., Więckowski, T., Pietrzyński, G., Soszyński, I., Szewczyk, O., Wyrzykowski, Ł., Paczyński, B., Abe, F., Bond, I. A., Britton, T. R., Gilmore, A. C., Hearnshaw, J. B., Itow, Y., Kamiya, K., Kilmartin, P. M., Korpela, A. V., Masuda, K., Matsubara, Y., Motomura, M., Muraki, Y., Nakamura, S., Okada, C., Ohnishi, K., Rattenbury, N. J., Sako, T., Sato, S., Sasaki, M., Sekiguchi, T., Sullivan, D. J., Tristram, P. J., Yock, P. C. M., & Yoshioka, T. 2006, Nature, 439, 437

Bodenheimer, P., Laughlin, G., & Lin, D. N. C. 2003, ApJ, 592, 555

Bond, I. A., Udalski, A., Jaroszyński, M., Rattenbury, N. J., Paczyński, B., Soszyński, I., Wyrzykowski, L., Szymański, M. K., Kubiak, M., Szewczyk, O., Żebruń, K., Pietrzyński, G., Abe, F., Bennett, D. P., Eguchi, S., Furuta, Y., Hearnshaw, J. B., Kamiya, K., Kilmartin, P. M., Kurata, Y., Masuda, K., Matsubara, Y., Muraki, Y., Noda, S., Okajima, K., Sako, T., Sekiguchi, T., Sullivan, D. J., Sumi, T., Tristram, P. J., Yanagisawa, T., & Yock, P. C. M. 2004, ApJ, 606, L155

Borucki, W. J., Koch, D., Basri, G., Brown, T., Caldwell, D., Devore, E., Dunham, E., Gautier, T., Geary, J., Gilliland, R., Gould, A., Howell, S., & Jenkins, J. 2003, in ESA Special Publication, Vol. 539, Earths: DARWIN/TPF and the Search for Extrasolar Terrestrial Planets, ed. M. Fridlund, T. Henning, & H. Lacoste, 69–81

Borucki, W. J., Koch, D. G., Lissauer, J., Basri, G., Brown, T., Caldwell, D. A., Jenkins, J. M., Caldwell, J. J., Christensen-Dalsgaard, J., Cochran, W. D., Dunham, E. W., Gautier, T. N., Geary, J. C., Latham, D., Sasselov, D., Gilliland, R. L., Howell, S., Monet, D. G., & Batalha, N. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 366, Transiting Extrapolar Planets Workshop, ed. C. Afonso, D. Weldrake, & T. Henning, 309–+

Bouchy, F., Bazot, M., Santos, N. C., Vauclair, S., & Sosnowska, D. 2005a, A&A, 440, 609

Bouchy, F., Pont, F., Santos, N. C., Melo, C., Mayor, M., Queloz, D., & Udry, S. 2004, A&A, 421, L13

Bouchy, F., Udry, S., Mayor, M., Moutou, C., Pont, F., Iribarne, N., da Silva, R., Ilovaisky, S., Queloz, D., Santos, N. C., Ségransan, D., & Zucker, S. 2005b, A&A, 444, L15

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

Budyko, M. I. 1969, Tellus, 21, 611

Burke, C. J., McCullough, P. R., Valenti, J. A., Johns-Krull, C. M., Janes, K. A., Heasley, J. N., Summers, F. J., Stys, J. E., Bissinger, R., Fleenor, M. L., Foote, C. N., Garcia-Melendo, E., Gary, B. L., Howell, P. J., Mallia, F., Masi, G., Taylor, B., & Vanmunster, T. 2007, ArXiv e-prints, 705

Burrows, A., Budaj, J., & Hubeny, I. 2007a, ArXiv e-prints, 709

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

Burrows, A., Hubeny, I., Hubbard, W. B., Sudarsky, D., & Fortney, J. J. 2004, ApJ, 610, L53

Burrows, A., Sudarsky, D., & Hubbard, W. B. 2003, ApJ, 594, 545

Butler, R. P., Marcy, G. W., Williams, E., McCarthy, C., Dosanjh, P., & Vogt, S. S. 1996, PASP, 108, 500

Campbell, B., Walker, G. A. H., & Yang, S. 1988, ApJ, 331, 902

Canup, R. M. 2004, Icarus, 168, 433

Carlberg, R. G., Yee, H. K. C., Morris, S. L., Lin, H., Hall, P. B., Patton, D. R., Sawicki, M., & Shepherd, C. W. 2001a, ApJ, 563, 736

—. 2001b, ApJ, 552, 427

Carpenter, E. J., Lin, S., & Capone, D. G. 2000, Applied and Environmental Microbiology, 66, 4514

Castro, S., Pogge, R. W., Rich, R. M., DePoy, D. L., & Gould, A. 2001, ApJ, 548, L197

Chabrier, G. & Baraffe, I. 2007, ApJ, 661, L81

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45

Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, ApJ, 568, 377

Charbonneau, D. & Deming, D. 2007, ArXiv e-prints, 706

Cho, J. Y.-K., Menou, K., Hansen, B. M. S., & Seager, S. 2003, ApJ, 587, L117

Cohen, M. H. & Cronyn, W. M. 1974, ApJ, 192, 193

Collier Cameron, A., Pollacco, D., Street, R. A., Lister, T. A., West, R. G., Wilson, D. M., Pont, F., Christian, D. J., Clarkson, W. I., Enoch, B., Evans, A., Fitzsimmons, A., Haswell, C. A., Hellier, C., Hodgkin, S. T., Horne, K., Irwin, J., Kane, S. R., Keenan, F. P., Norton, A. J., Parley, N. R., Osborne, J., Ryans, R., Skillen, I., & Wheatley, P. J. 2006, MNRAS, 373, 799

Cooper, C. S. & Showman, A. P. 2005, ApJ, 629, L45

Davé, R., Cen, R., Ostriker, J. P., Bryan, G. L., Hernquist, L., Katz, N., Weinberg, D. H., Norman, M. L., & O'Shea, B. 2001, ApJ, 552, 473

Dehant, V., Lammer, H., Kulikov, Y. N., Grießmeier, J.-M., Breuer, D., Verhoeven, O., Karatekin, Ö., van Hoolst, T., Korablev, O., & Lognonné, P. 2007, Space Science Reviews, 129, 279

del Genio, A. D. & Zhou, W. 1996, Icarus, 120, 332

del Genio, A. D., Zhou, W., & Eichler, T. P. 1993, Icarus, 101, 1

Dickey, J. M. & Lockman, F. J. 1990, ARA&A, 28, 215

Diego, F., Fish, A. C., Barlow, M. J., Crawford, I. A., Spyromilio, J., Dryburgh, M., Brooks, D., Howarth, I. D., & Walker, D. D. 1995, MNRAS, 272, 323

Dole, S. H. 1964, Habitable planets for man (New York, Blaisdell Pub. Co. [1964] [1st ed.].)

Dravins, D. 1985, in IAU Colloq. 88: Stellar Radial Velocities, ed. A. G. D. Philip & D. W. Latham, 311–+

Efford, N. D. 1991, Earth Moon and Planets, 54, 19

Ettori, S., Tozzi, P., & Rosati, P. 2003, A&A, 398, 879

Fabrycky, D. C., Johnson, E. T., & Goodman, J. 2007, ApJ, 665, 754

Fang, T., Gerke, B. F., Davis, D. S., Newman, J. A., Davis, M., Nandra, K., Laird, E. S., Koo, D. C., Coil, A. L., Cooper, M. C., Croton, D. J., & Yan, R. 2007, ApJ, 660, L27

Farrell, B. F. 1990, Journal of Atmospheric Sciences, 47, 2986

Farrell, W. M., Lazio, T. J. W., Zarka, P., Bastian, T. J., Desch, M. D., & Ryabov, B. P. 2004, Planet. Space Sci., 52, 1469

Fassnacht, C. D., Kocevski, D. D., Auger, M. W., Lubin, L. M., Neureuther, J. L., Jeltema, T. E., Mulchaey, J. S., & McKean, J. P. 2007, ArXiv e-prints, 711

Firsoff, V. A. 1963, Life beyond the earth; a study in exobiology. (New York, Basic Books [1964, c1963])

Ford, E. B., Quinn, S. N., & Veras, D. 2008, ApJ, 678, 1407

Forget, F. & Pierrehumbert, R. T. 1997, Science, 278, 1273

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

Fortney, J. J. & Marley, M. S. 2007, ApJ, 666, L45

Franck, S., von Bloh, W., Bounama, C., Steffen, M., Schönberner, D., & Schellnhuber, H.-J. 2000, J. Geophys. Res., 105, 1651

Fukugita, M. & Peebles, P. J. E. 2004, ApJ, 616, 643

Gaidos, E., Deschenes, B., Dundon, L., Fagan, K., Menviel-Hessler, L., Moskovitz, N., & Workman, M. 2005, Astrobiology, 5, 100

Gaidos, E. & Williams, D. M. 2004, New Astronomy, 10, 67

Gaudi, B. S. 2005, ApJ, 628, L73

Gaudi, B. S., Chang, H.-Y., & Han, C. 2003, ApJ, 586, 527

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

Ge, J., Angel, J. R. P., Jacobsen, B., Woolf, N., Fugate, R. Q., Black, J. H., & Lloyd-Hart, M. 2002, PASP, 114, 879

Ge, J., van Eyken, J., Mahadevan, S., DeWitt, C., Kane, S. R., Cohen, R., Vanden Heuvel, A., Fleming, S. W., Guo, P., Henry, G. W., Schneider, D. P., Ramsey, L. W., Wittenmyer, R. A., Endl, M., Cochran, W. D., Ford, E. B., Martín, E. L., Israelian, G., Valenti, J., & Montes, D. 2006, ApJ, 648, 683

Ghil, M. 2002, Encyclopedia of Atmospheric Sciences, J. R. Holton, J. Pyle, and J. A. Curry (Eds.), 432

Giacconi, R., Murray, S., Gursky, H., Kellogg, E., Schreier, E., Matilsky, T., Koch, D., & Tananbaum, H. 1974, ApJS, 27, 37

Gillon, M., Demory, B.-O., Barman, T., Bonfils, X., Mazeh, T., Pont, F., Udry, S., Mayor, M., & Queloz, D. 2007, A&A, 471, L51

Giménez, A. 2006, ApJ, 650, 408

Goldreich, P. & Peale, S. 1966, AJ, 71, 425

Goldsmith, D. & Owen, T. 2002, The Search for Life in the Universe (Sausalito, CA, University Science Books [2002] [3rd ed.].)

Gould, A. & Loeb, A. 1992, ApJ, 396, 104

Graff, D. S. & Gaudi, B. S. 2000, ApJ, 538, L133

Grenfell, J. L., Stracke, B., von Paris, P., Patzer, B., Titz, R., Segura, A., & Rauer, H. 2007, Planet. Space Sci., 55, 661

Grießmeier, J.-M., Motschmann, U., Mann, G., & Rucker, H. O. 2005, A&A, 437, 717

Grillmair, C. J., Charbonneau, D., Burrows, A., Armus, L., Stauffer, J., Meadows, V., Van Cleve, J., & Levine, D. 2007, ApJ, 658, L115

Guillot, T. 2005, Annual Review of Earth and Planetary Sciences, 33, 493

Guillot, T., Burrows, A., Hubbard, W. B., Lunine, J. I., & Saumon, D. 1996, ApJ, 459, L35+

Guillot, T. & Showman, A. P. 2002, A&A, 385, 156

Gurnett, D. A. 1975, J. Geophys. Res., 80, 2751

Haldene, J. B. S. 1954, New Biology, 16

Hameury, J.-M., Menou, K., Dubus, G., Lasota, J.-P., & Hure, J.-M. 1998, MNRAS, 298, 1048

Harrington, J., Hansen, B. M., Luszcz, S. H., Seager, S., Deming, D., Menou, K., Cho, J. Y.-K., & Richardson, L. J. 2006, Science, 314, 623

Hart, M. H. 1979, Icarus, 37, 351

Hartmann, D. L. 1994, Global Physical Climatology (Academic Press, New York), 411

Hatzes, A. P., Cochran, W. D., Endl, M., McArthur, B., Paulson, D. B., Walker, G. A. H., Campbell, B., & Yang, S. 2003, ApJ, 599, 1383

Heintz, W. D. 1988, JRASC, 82, 140

Held, I. M., Linder, D. I., & Suarez, M. J. 1981, Journal of Atmospheric Sciences, 38, 1911

Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, ApJ, 529, L41

Heyrovsky, D. & Loeb, A. 1997, ApJ, 490, 38

Hoekstra, H., Franx, M., Kuijken, K., Carlberg, R. G., & Yee, H. K. C. 2003, MNRAS, 340, 609

Hoekstra, H., Franx, M., Kuijken, K., Carlberg, R. G., Yee, H. K. C., Lin, H., Morris, S. L., Hall, P. B., Patton, D. R., Sawicki, M., & Wirth, G. D. 2001, ApJ, 548, L5

Hoffman, P. F., Kaufman, A. J., Halverson, G. P., & Schrag, D. P. 1998, Science, 281, 1342

Hoffman, P. F. & Schrag, D. P. 2000, Scientific American, 282, 68

—. 2002, Terra Nova, 114, 129

Holton, J. R. 1992, An introduction to dynamic meteorology (International geophysics series, San Diego, New York: Academic Press, —c1992, 3rd ed.)

Horne, K. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 294, Scientific Frontiers in Research on Extrasolar Planets, ed. D. Deming & S. Seager, 361–370

Howard, R., Gilman, P. I., & Gilman, P. A. 1984, ApJ, 283, 373

Huang, S.-S. 1959, PASP, 71, 421

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

Jacob, W. S. 1855, MNRAS, 15, 228

Janhunen, P., Olsson, A., Karlsson, R., & Griessmeier, J. . 2003, ArXiv Astrophysics e-prints

Johns-Krull, C. M., McCullough, P. M., Burke, C. J., Valenti, J. A., Janes, K. A., Heasley, J. N., Bissinger, R., Fleenor, M., Foote, C. N., Garcia-Melendo, E., Gary, B. L., Howell, P. J., Mallia, F., Masi, G., Prato, L. A., & Vanmunster, T. 2007, in American Astronomical Society Meeting Abstracts, Vol. 210, American Astronomical Society Meeting Abstracts, 96.05–+

Joshi, M. 2003, Astrobiology, 3, 415

Joshi, M. M., Haberle, R. M., & Reynolds, R. T. 1997, Icarus, 129, 450

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., & Joselph, D. 1996, Bull. Amer. Meteor. Soc., 77, 437

Kaltenegger, L. & Selsis, F. 2007, ArXiv e-prints 0710.0881

Karkoschka, E. 1994, Icarus, 111, 174

Kashefi, K. & Lovley, D. 2003, Science, 301, 934

Kasting, J. F. 1988, Icarus, 74, 472

Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. 1993, Icarus, 101, 108

Kayser, R. & Witt, H. J. 1989, A&A, 221, 1

Kiang, N. Y., Segura, A., Tinetti, G., Govindjee, Blankenship, R. E., Cohen, M., Siefert, J., Crisp, D., & Meadows, V. S. 2007, Astrobiology, 7, 252

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van del Dool, H., Jenne, R., & Fiorino, M. 1999, Bull. Amer. Meteor. Soc., 82, 247

Knutson, H. A., Charbonneau, D., Allen, L. E., Fortney, J. J., Agol, E., Cowan, N. B., Showman, A. P., Cooper, C. S., & Megeath, S. T. 2007a, Nature, 447, 183

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

Konacki, M., Torres, G., Jha, S., & Sasselov, D. D. 2003, Nature, 421, 507

Konacki, M., Torres, G., Sasselov, D. D., & Jha, S. 2005, ApJ, 624, 372

Konacki, M., Torres, G., Sasselov, D. D., Pietrzyński, G., Udalski, A., Jha, S., Ruiz, M. T., Gieren, W., & Minniti, D. 2004, ApJ, 609, L37

Kuchner, M. J. & Traub, W. A. 2002, ApJ, 570, 900

Laskar, J., Joutel, F., & Robutel, P. 1993, Nature, 361, 615

Laskar, J. & Robutel, P. 1993, Nature, 361, 608

Laughlin, G., Wolf, A., Vanmunster, T., Bodenheimer, P., Fischer, D., Marcy, G., Butler, P., & Vogt, S. 2005, ApJ, 621, 1072

Lazio, T. J. W. & Farrell, W. M. 2007, ApJ, 668, 1182

Leger, A. & Herbst, T. 2007, ArXiv e-prints 0707.3385

Leinert, C., Bowyer, S., Haikala, L. K., Hanner, M. S., Hauser, M. G., Levasseur-Regourd, A.-C., Mann, I., Mattila, K., Reach, W. T., Schlosser, W., Staude, H. J., Toller, G. N., Weiland, J. L., Weinberg, J. L., & Witt, A. N. 1998, A&AS, 127, 1

Levrard, B., Correia, A. C. M., Chabrier, G., Baraffe, I., Selsis, F., & Laskar, J. 2007, A&A, 462, L5

Lewis, G. F. & Belle, K. E. 1998, MNRAS, 297, 69

Lewis, G. F. & Ibata, R. A. 2000, ApJ, 539, L63

Loeb, A. 2005, ApJ, 623, L45

Lorenz, E. N. 1979, Journal of Atmospheric Sciences, 36, 1367

Mahdavi, A. & Geller, M. J. 2004, ApJ, 607, 202

Mao, S. & Paczynski, B. 1991, ApJ, 374, L37

Marcy, G., Butler, R. P., Fischer, D., Vogt, S., Wright, J. T., Tinney, C. G., & Jones, H. R. A. 2005, Progress of Theoretical Physics Supplement, 158, 24

Marcy, G. W. & Butler, R. P. 1996, ApJ, 464, L147+

Mayor, M. & Queloz, D. 1995, Nature, 378, 355

McCammon, D., Almy, R., Apodaca, E., Bergmann Tiest, W., Cui, W., Deiker, S., Galeazzi, M., Juda, M., Lesser, A., Mihara, T., Morgenthaler, J. P., Sanders, W. T., Zhang, J., Figueroa-Feliciano, E., Kelley, R. L., Moseley, S. H., Mushotzky, R. F., Porter, F. S., Stahle, C. K., & Szymkowiak, A. E. 2002, ApJ, 576, 188

McCammon, D. & Sanders, W. T. 1990, ARA&A, 28, 657

McCullough, P. R., Stys, J. E., Valenti, J. A., Johns-Krull, C. M., Janes, K. A., Heasley, J. N., Bye, B. A., Dodd, C., Fleming, S. W., Pinnick, A., Bissinger, R., Gary, B. L., Howell, P. J., & Vanmunster, T. 2006, ApJ, 648, 1228

McLaughlin, D. B. 1924, ApJ, 60, 22

Mischna, M. A., Kasting, J. F., Pavlov, A., & Freedman, R. 2000, Icarus, 145, 546

Mulchaey, J. S. 2000, ARA&A, 38, 289

Mulchaey, J. S., Davis, D. S., Mushotzky, R. F., & Burstein, D. 2003, ApJS, 145, 39

Mutel, R., Gurnett, D., & Christopher, I. 2004, Annales Geophysicae, 22, 2625

Nakamura, T. & Tajika, E. 2002, Journal of Geophysical Research (Planets), 107, 5094

—. 2003, Geophys. Res. Lett., 30, 18

Narita, N., Enya, K., Sato, B., Ohta, Y., Winn, J. N., Suto, Y., Taruya, A., Turner, E. L., Aoki, W., Tamura, M., Yamada, T., & Yoshi, Y. 2007, ArXiv Astrophysics e-prints

Neron de Surgy, O. & Laskar, J. 1997, A&A, 318, 975

North, G. R., Cahalan, R. F., & Coakley, Jr., J. A. 1981, Reviews of Geophysics and Space Physics, 19, 91

North, G. R. & Coakley, J. A. 1979, in Evolution of Planetary Atmospheres and Climatology of the Earth, 249–+

North, G. R., Short, D. A., & Mengel, J. G. 1983, J. Geophys. Res., 88, 6576

Novaco, J. C. & Brown, L. W. 1978, ApJ, 221, 114

O'Donovan, F. T., Charbonneau, D., Mandushev, G., Dunham, E. W., Latham, D. W., Torres, G., Sozzetti, A., Brown, T. M., Trauger, J. T., Belmonte, J. A., Rabus, M., Almenara, J. M., Alonso, R., Deeg, H. J., Esquerdo, G. A., Falco, E. E., Hillenbrand, L. A., Roussanova, A., Stefanik, R. P., & Winn, J. N. 2006, ApJ, 651, L61

Ohta, Y., Taruya, A., & Suto, Y. 2005, ApJ, 622, 1118

—. 2006, ArXiv Astrophysics e-prints

Oppenheimer, B. R., Golimowski, D. A., Kulkarni, S. R., Matthews, K., Nakajima, T., Creech-Eakman, M., & Durrance, S. T. 2001, AJ, 121, 2189

Osmond, J. P. F. & Ponman, T. J. 2004, MNRAS, 350, 1511

Ota, N. & Mitsuda, K. 2004, A&A, 428, 757

Pearson, K. 1901, Philosophical Magazine, 2, 609

Pedlosky, J. 1982, Geophysical fluid dynamics (New York and Berlin, Springer-Verlag, 1982. 636 p.)

Plionis, M. & Tovmassian, H. M. 2004, A&A, 416, 441

Pont, F., Bouchy, F., Queloz, D., Santos, N. C., Melo, C., Mayor, M., & Udry, S. 2004, A&A, 426, L15

Poynting, J. H. 1903, MNRAS, 64, A1+

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1992, Numerical Recipes: The Art of Scientific Computing, 2nd edn. (Cambridge (UK) and New York: Cambridge University Press)

Press, W. H. & Rybicki, G. B. 1993, ApJ, 418, 585

Press, W. H. & Schechter, P. 1974, ApJ, 187, 425

Reipurth, B., Jewitt, D., & Keil, K., eds. 2007, Protostars and Planets V

Richardson, L. J., Deming, D., Horning, K., Seager, S., & Harrington, J. 2007, Nature, 445, 892

Richardson, L. J., Deming, D., & Seager, S. 2003, ApJ, 597, 581

Robertson, H. P. 1937, MNRAS, 97, 423

Rosati, P., Borgani, S., & Norman, C. 2002, ARA&A, 40, 539

Rossiter, R. A. 1924, ApJ, 60, 15

Rowan-Robinson, M. & Fabian, A. C. 1975, MNRAS, 170, 199

Rubincam, D. P. 2004, Theoretical and Applied Climatology, 79, 111

Scharf, C. 2002, ApJ, 572, 157

Scharf, C. A. 2006, ApJ, 648, 1196

Schwartz, D. A. 1978, ApJ, 220, 8

Seager, S. & Hui, L. 2002, ApJ, 574, 1004

Seager, S., Kuchner, M., Hier-Majumder, C. A., & Militzer, B. 2007, ApJ, 669, 1279

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

Seager, S., Turner, E. L., Schafer, J., & Ford, E. B. 2005, Astrobiology, 5, 372

Seager, S., Whitney, B. A., & Sasselov, D. D. 2000, ApJ, 540, 504

Sellers, W. D. 1969, Journal of Applied Meteorology, 8, 392

Selsis, F., Kasting, J. F., Levrard, B., Paillet, J., Ribas, I., & Delfosse, X. 2007, A&A, 476, 1373

Showman, A. P. & Guillot, T. 2002, A&A, 385, 166

Shu, F. H. 1982, The physical universe. an introduction to astronomy (A Series of Books in Astronomy, Mill Valley, CA: University Science Books, 1982)

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

Spiegel, D. S., Haiman, Z., & Gaudi, B. S. 2007a, ApJ, 669, 1324

—. 2007b, ArXiv e-prints, 705

Spiegel, D. S., Menou, K., & Scharf, C. A. 2008, ApJ, 681, 1609

Spiegel, D. S., Paerels, F., & Scharf, C. A. 2007c, ApJ, 658, 288

Spiegel, D. S., Zamojski, M., Gersch, A., Donovan, J., & Haiman, Z. 2005, ApJ, 628, 478

Stone, P. H. 1973, J. Atmos. Sci, 30, 521

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

Sudarsky, D., Burrows, A., & Hubeny, I. 2003, ApJ, 588, 1121

Sudarsky, D., Burrows, A., & Pinto, P. 2000, ApJ, 538, 885

Sugiyama, M., Stone, P., & Emanuel, K. A. 2005, Journal of the Atmospheric Sciences, 62, 2001

Swain, M. R., Bouwman, J., Akeson, R., Lawler, S., & Beichman, C. 2007, ArXiv Astrophysics e-prints

Tokunaga, A. T., Bond, T., Elias, J., Chun, M., Richter, M., Liang, M., Lacy, J., Daggert, L., Tollestrup, E., Ressler, M., Warren, D., Fisher, S., & Carr, J. 2006, in Ground-based and Airborne Instrumentation for Astronomy. Edited by McLean, Ian S.; Iye, Masanori. Proceedings of the SPIE, Volume 6269, pp. 62693Y (2006).

Trenberth, K. E. & Solomon, A. 1994, Climate Dynamics, 10, 107

Trilling, D. E., Lunine, J. I., & Benz, W. 2002, A&A, 394, 241

Udalski, A., Pietrzynski, G., Szymanski, M., Kubiak, M., Zebrun, K., Soszynski, I., Szewczyk, O., & Wyrzykowski, L. 2003, Acta Astronomica, 53, 133

Udalski, A., Szewczyk, O., Zebrun, K., Pietrzynski, G., Szymanski, M., Kubiak, M., Soszynski, I., & Wyrzykowski, L. 2002a, Acta Astronomica, 52, 317

Udalski, A., Szymanski, M., Kubiak, M., Pietrzynski, G., Soszynski, I., Wozniak, P., Zebrun, K., Szewczyk, O., & Wyrzykowski, L. 2002b, Acta Astronomica, 52, 217

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

Udalski, A., Zebrun, K., Szymanski, M., Kubiak, M., Soszynski, I., Szewczyk, O., Wyrzykowski, L., & Pietrzynski, G. 2002c, Acta Astronomica, 52, 115

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

Ulrich, R. K. 1991, Advances in Space Research, 11, 217

Unwin, S. C., Shao, M., Tanner, A. M., Allen, R. J., Beichman, C. A., Boboltz, D., Catanzarite, J. H., Chaboyer, B. C., Ciardi, D. R., Edberg, S. J., Fey, A. L., Fischer, D. A., Gelino, C. R., Gould, A. P., Grillmair, C., Henry, T. J., Johnston, K. V., Johnston, K. J., Jones, D. L., Kulkarni, S. R., Law, N. M., Majewski, S. R., Makarov, V. V., Marcy, G. W., Meier, D. L., Olling, R. P., Pan, X., Patterson, R. J., Pitesky, J. E., Quirrenbach, A., Shaklan, S. B., Shaya, E. J., Strigari, L. E., Tomsick, J. A., Wehrle, A. E., & Worthey, G. 2007, ArXiv e-prints 0708.3953

Valencia, D., O'Connell, R. J., & Sasselov, D. 2006, Icarus, 181, 545

Valencia, D., Sasselov, D. D., & O'Connell, R. J. 2007, ApJ, 665, 1413

Vidal-Madjar, A., Désert, J.-M., Lecavelier des Etangs, A., Hébrard, G., Ballester, G. E., Ehrenreich, D., Ferlet, R., McConnell, J. C., Mayor, M., & Parkinson, C. D. 2004, ApJ, 604, L69

Vidal-Madjar, A., Lecavelier des Etangs, A., Désert, J.-M., Ballester, G. E., Ferlet, R., Hébrard, G., & Mayor, M. 2003, Nature, 422, 143

von Bloh, W., Bounama, C., Cuntz, M., & Franck, S. 2008, in IAU Symposium, Vol. 249, IAU Symposium, 503–506

Walker, J. C. G., Hays, P. B., & Kasting, J. F. 1981, J. Geophys. Res., 86, 9776

Williams, D. M. & Kasting, J. F. 1997, Icarus, 129, 254

Williams, D. M., Kasting, J. F., & Wade, R. A. 1997, Nature, 385, 234

Williams, D. M. & Pollard, D. 2002, International Journal of Astrobiology, 1, 61

—. 2003, International Journal of Astrobiology, 2, 1

Willis, J. P., Pacaud, F., Valtchanov, I., Pierre, M., Ponman, T., Read, A., Andreon, S., Altieri, B., Quintana, H., Dos Santos, S., Birkinshaw, M., Bremer, M., Duc, P.-A., Galaz, G., Gosset, E., Jones, L., & Surdej, J. 2005, MNRAS, 363, 675

Winn, J. N. & Holman, M. J. 2005, ApJ, 628, L159

Winn, J. N., Johnson, J. A., Marcy, G. W., Butler, R. P., Vogt, S. S., Henry, G. W., Roussanova, A., Holman, M. J., Enya, K., Narita, N., Suto, Y., & Turner, E. L. 2006, ApJ, 653, L69

Winn, J. N., Noyes, R. W., Holman, M. J., Charbonneau, D., Ohta, Y., Taruya, A., Suto, Y., Narita, N., Turner, E. L., Johnson, J. A., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2005, ApJ, 631, 1215

Wolszczan, A. & Frail, D. A. 1992, Nature, 355, 145

Wunsch, C. & Hiembach, P. 2007, Physica D, 230, 197

Xue, Y.-J. & Wu, X.-P. 2000, ApJ, 538, 65

Zarka, P. 1992, Advances in Space Research, 12, 99

Zwicky, F. 1937, ApJ, 86, 217

—. 1938, PASP, 50, 218