The Formation and Evolution of Planetary Systems

by

Niraj K. Inamdar

B.S.E., University of Pennsylvania (2008)
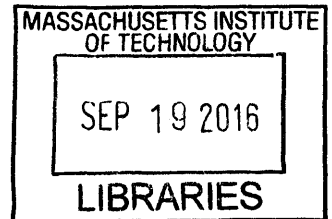S.M., Massachusetts Institute of Technology (2011)

Submitted to the Department of Earth, Atmospheric and Planetary Sciences in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Planetary Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2016

Author........ Signature redacted ....
Department of Earth, Atmospheric and Planetary Sciences

Certified by. Signature redacted August 2, 2016
....
Hilke E. Schlichting
Assistant Professor of Planetary Sciences
Thesis Supervisor

Certified by. Signature redacted ....
Sara Seager
Professor of Planetary Sciences and Physics
Thesis Supervisor

Accepted by . Signature redacted ....
Robert D. van der Hilst
Schlumberger Professor of Earth and Planetary Sciences
Head, Department of Earth, Atmospheric and Planetary Sciences

The Formation and Evolution of Planetary Systems

by

Niraj K. Inamdar

Submitted to the Department of Earth, Atmospheric and Planetary
Sciences
on August 2, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Planetary Science

Abstract

The *Kepler* space observatory and other surveys have revealed thousands of planets and planetary systems that look significantly different from our own. In particular, the preponderance of super-Earths and mini-Neptunes (planets with radii smaller than Neptune's but larger than Earth's) at short orbital radii has challenged planet formation theories developed in the context of our own Solar System. How and where these planets form remains an outstanding question. Given the large frequency with which such planets occur around other stars, understanding the formation pathways of these planets has the potential to reveal dominant mechanisms for planet formation and evolution and to place our own Solar System within a broader context. The purpose of this thesis is to address and develop frameworks by which these questions can be answered. The thesis is comprised of two parts.

In Part I, we consider the question of how and where close-in exoplanets formed. We do this in two ways. First, we use that fact that many close-in super-Earths and mini-Neptunes possess gaseous envelopes that comprise several percent or more of the total planet mass in order to construct a self-consistent planet formation history that accounts for core accretion, thermal evolution, and dynamical interactions during the core assembly process via giant impacts. We find that envelope masses accreted from the gas disc by planetary embryos are typically much smaller than those inferred for many exoplanets, and that the envelope mass fraction is further reduced substantially during the assembly phase when embryos merge. Fully assembled planets can accrete ob-

served envelope masses from the residual disc only if energy exchange between the envelope and the underlying core is totally inhibited. It is therefore very unlikely that such planets formed at their observed semi-major axes, but instead formed further out and migrated inwards.

Second, we consider the late-stage evolution of exoplanets after gas disc dissipation in order to explain the bulk structural diversity of observed super-Earths and mini-Neptunes. Whereas naïve application of core accretion models suggests a narrow mass-radius relationship for these exoplanets, the population possesses a great deal of diversity in mean density. While photoevaporative mass loss from the host star is the most-commonly invoked explanation for this diversity, we use the fact that many exoplanetary systems are in tightly packed orbital configurations to propose instead that late-stage collisions are at least in part responsible for the observed diversity. We infer envelope mass fractions for planets in the literature whose masses and radii have been measured, and on the basis of this, we identify multiplanet systems whose bulk structural diversity favors late-stage impacts as opposed to photoevaporative devolatilization.

In Part II, we turn our attention to the formation and evolution of our own Solar System. We do this within the context of NASA's OSIRIS-REx asteroid sample return mission. OSIRIS-REx, which launches in September 2016, will arrive at the near-Earth asteroid 101955 Bennu in 2019 with the objective of constraining its composition, orbit, and other bulk properties. In order to better understand the composition of Bennu, an instrument designed to measure its elemental abundances via X-ray fluorescence spectroscopy called REXIS was developed and built at MIT. In this part of the thesis, we investigate the ability of REXIS to constrain the composition of Bennu via fluorescence spectroscopy, as well as its potential to place Bennu within an analogue meteorite class. We carry out our analysis by modeling Solar X-ray activity and the X-ray fluorescence from Bennu, as well as by simulating data product and analysis from the instrument in order to predict REXIS's ability to carry out its goal of contextualizing Bennu within the asteroid and meteorite population.

Thesis Supervisor: Hilke E. Schlichting
Title: Assistant Professor of Planetary Sciences

Thesis Supervisor: Sara Seager
Title: Professor of Planetary Sciences and Physics

Acknowledgments

the team's work at design reviews multiple times. To the rest of the REXIS team, I thank you for more than two years of enjoyable science and engineering.

On a personal note, I would like to thank the friends I have had in my time at MIT: Neesha Regmi Schnepf, David Hernandez, my office mates Ben Corbin and Alex Gonzales, Leo de Castro, Calvin Zhong, Amanda Wang, and others whose names escape me for the moment. I would like to thank the residents and house team of East Campus, and the students of Fifth East in particular, where I had the pleasure of serving as graduate tutor for nearly two years. The students who were in the five classes for which I was teaching assistant reminded me how smart and talented our students are, and made teaching a genuinely fun experience. I thank you as well.

John Biersteker, I thank you for many hours of great banter and friendship. I know Skyharbour may cost quadrillions of dollars and be 200 years ahead of its time technologically, but I know that won't stop us from doing all we can from turning this amazing piece of science fiction into science fact. And Robbie Romero and Sarah Tortorici, you two have been like brother and sister to me.

Lastly, I would like to thank my dad. You're my best friend and have supported me and my interests—whether scientific, artistic, or otherwise—unconditionally throughout all these years in spite of many difficulties and hardships. For that, I am (and will always be) truly grateful.

"REVOCATE ANIMOS, MAESTUMQUE TIMOREM MITTITE:
FORSAN ET HAEC OLIM MEMINISSE IUVABIT."
AENEIS, LIBER I

Contents

List of Figures

14

List of Tables

15

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

For centuries, our own Solar System was the only means by which we could constrain our understanding of how planets and planetary systems form and evolve. Many of the theories applied to explain the nuances intrinsic to Solar Systems bodies—tidal theory and the theory of resonances, for instance—have met with considerable success when applied to our own Solar System. There are of course many outstanding questions remaining regarding our own Solar System: the remarkable isotopic similarity between the Earth and the Moon, whether Jupiter possesses a core or not, and the importance of giant planet migration in the early Solar System. Nevertheless, many of the basic models that were developed were able to reproduce a Solar System that appeared consistent with itself in a broad sense—the terrestrial planets, gas giants, and the icy giants are more or less where we "expect" to find them. (Where we expect to find these planets and why is expanded upon in Section 1.2).
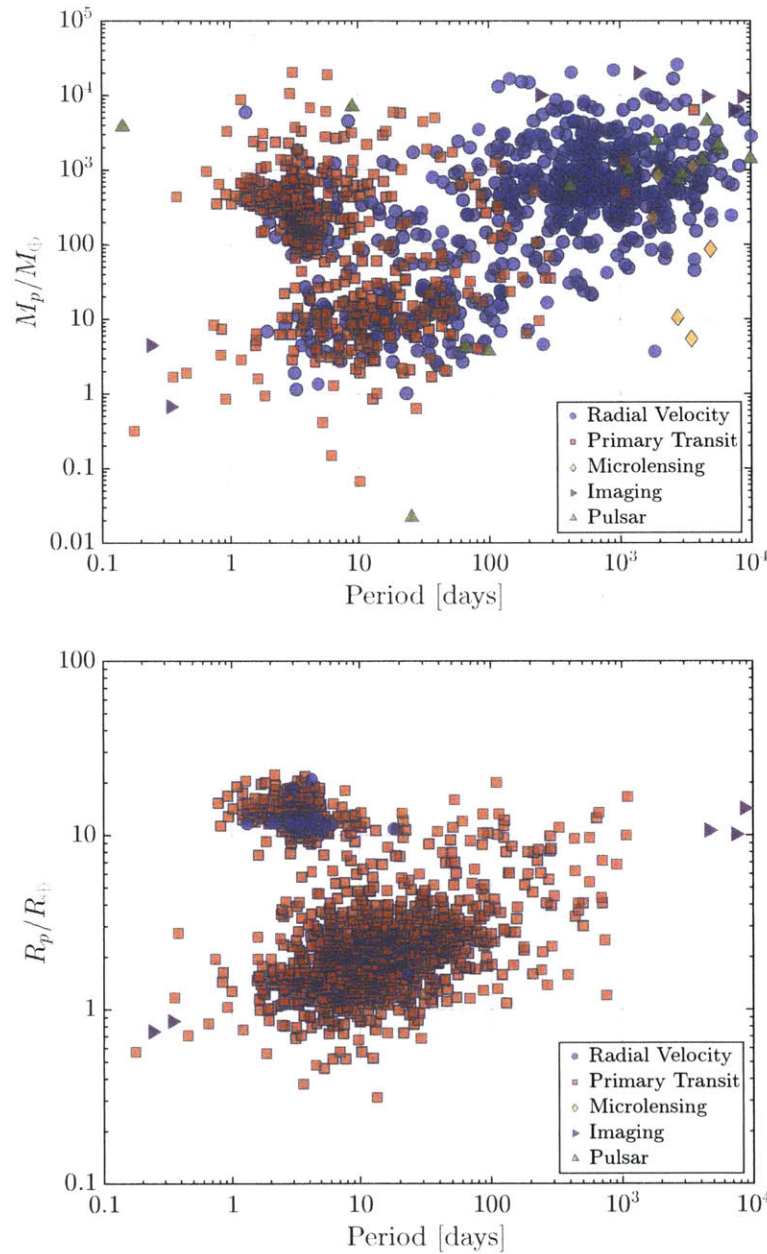
Figure 1-1: Observed exoplanets by detection technique. In the top panel, we show planet mass M_p in units of Earth mass $M_\oplus$ as a function of orbital period. In the bottom panel, we show planet radius R_p in units of Earth radius $R_\oplus$ as a function of orbital period. In both panels, we indicate the observational technique used to detect the planet. Due to the success of the *Kepler* mission, most of the data are from transits. Data from `exoplanet.eu` database (Schneider, Dedieu, Le Sidaner et al. 2011).

1.1 Observed Exoplanets

Twenty years ago, the discovery of the first exoplanets radically challenged many basic precepts of planet formation theory developed to that point (Mayor & Queloz 1995, Butler, Marcy, Williams et al. 1997). To date, there are over 2,000 confirmed exoplanets detected by the transit, radial velocity, microlensing, direct imaging, and astrometry methods (Schneider, Dedieu, Le Sidaner et al. 2011).

The first population of exoplanets discovered were the so-called "hot Jupiters": gas giants orbiting their host stars on extremely short periods on the order of days. While the population of hot Jupiters is striking and possesses many characteristics that hold tantalizing clues to its origin (Dawson & Murray-Clay 2013), their early discovery was largely a result of observational bias. Due to their high mass M_p and short orbital periods P, they possess a relatively high radial velocity signature, making detection via the Doppler effect relatively easy (top panel of Fig. 1-1). In fact, we now know that hot Jupiters comprise a relatively small percentage of the exoplanet population (on the order of a few percent). In recent years, however, thanks largely to the *Kepler* space telescope, thousands of additional exoplanets have been found via the transit method (see Fig. 1-3; Mullally, Coughlin, Thompson et al. 2015). The *Kepler* survey has revealed a population of planets radically different from our own. In particular the preponderance of short-period exoplanets is especially perplexing, since such planets are without precedent in our own Solar System. A number of these systems exhibit closely packed archi-

tectures, such as Kepler-11 (Lissauer, Fabrycky, Ford et al. 2011), which may hold a clue to their dynamical evolution (Deck, Payne, & Holman 2013). Furthermore, a number of the observed multiplanet systems have planet pairs that are preferentially near resonance but not in resonance (see, e.g., Goldreich & Schlichting 2014), again pointing to underlying dynamical effects that sculpt exoplanetary architectures. We note that the probability of detecting a transiting exoplanet is $R_\star/a$, where $R_\star$ is the radius of the host star and a is the semimajor axis of the planet's orbit (e.g., Winn 2010). Hence, while it may be argued that there is an observational bias for detecting such close-in planets, even correcting for this bias shows that this population of close-in planets is substantial (Fig. 1-2).

1.1.1 Super-Earths and Mini-Neptunes

Pending our ability to detect true Solar System analogues, the aforementioned population of exoplanets—which typically have radii greater than that of the Earth but less than that of Neptune, and most of which orbit their host stars at semimajor axes smaller than that of Mercury's—is perhaps the most intriguing population of exoplanets: the so-called super-Earths and mini-Neptunes (Fig. 1-3).*

---

*In the literature, the terms "super-Earths", "mini-Neptunes", and "sub-Neptunes" appear to have been used somewhat interchangeably to describe this entire population of planets. In this thesis, we will use the term "super-Earth" to refer to those that possess a negligible amount of mass in a gaseous envelope (as do the terrestrial planets in our Solar System) and "mini-Neptunes" to refer to those planet with non-negligible gaseous envelopes comprising up to several percent of their total mass, similar to Uranus and Neptune.
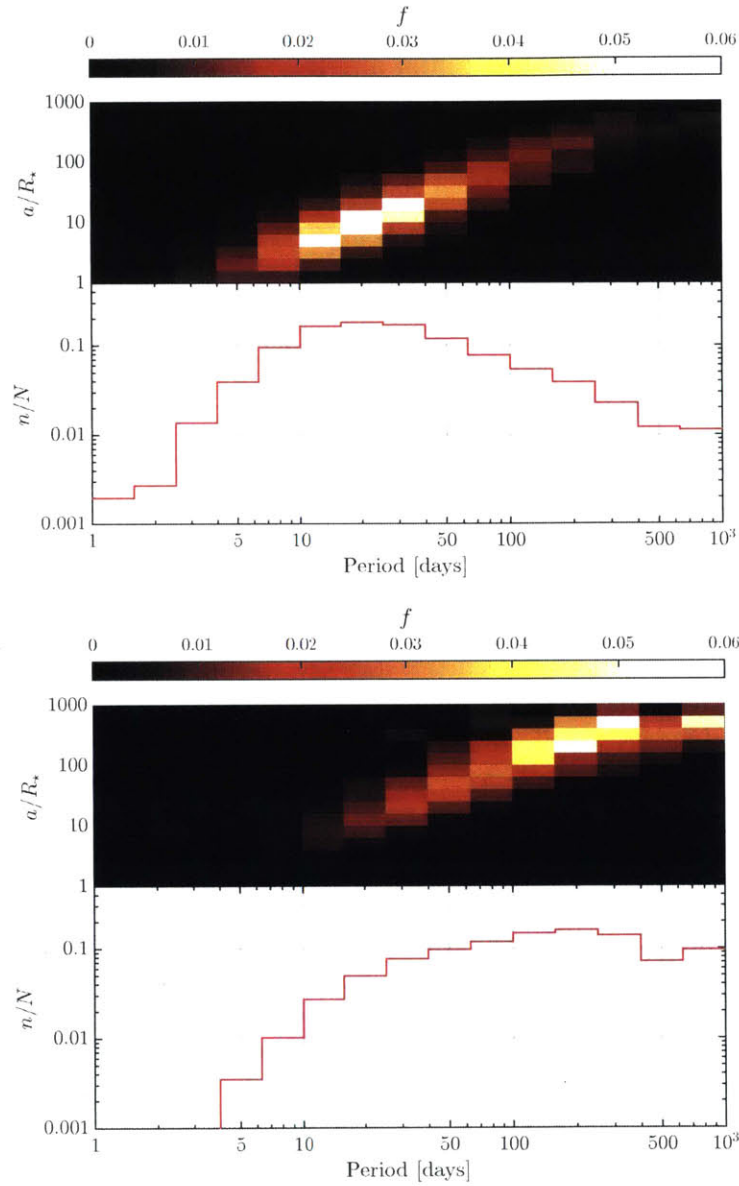
Figure 1-2: Exoplanet occurrence rate from transit measurements. In the top panel, we show the occurrence rate f as a function of the inverse transit probability $a/R_\star$ and period; immediately below, we show that occurrence rate summed over $a/R_\star$ and normalized. In the lower panel, we show the same, but with the occurrence rates multiplied by the $a/R_\star$ to account for transit probability. Data from the NASA Exoplanet Archive (Akeson, Chen, Ciardi et al. 2013).

21
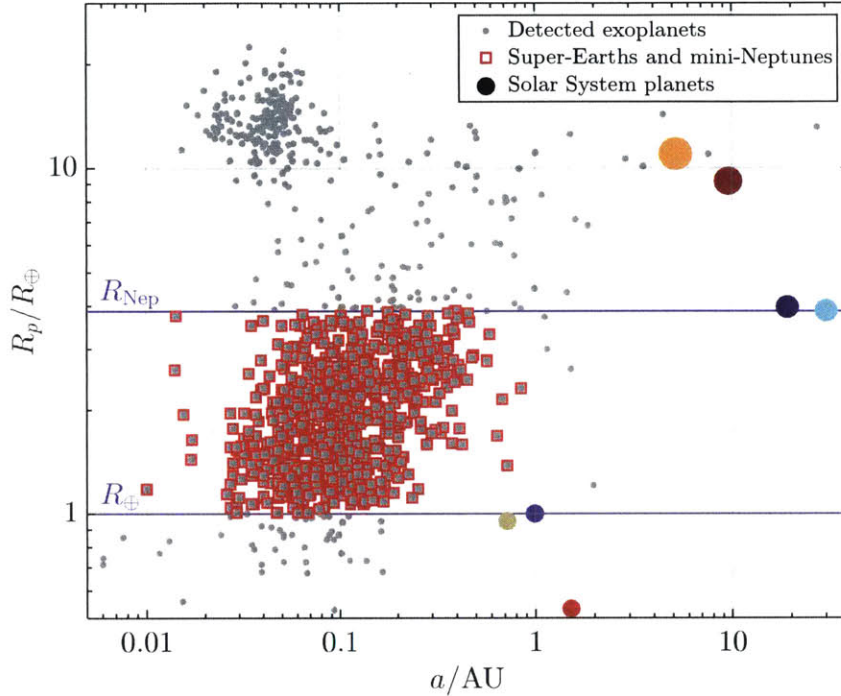
Figure 1-3: Radii R_p and semimajor axes a of discovered exoplanets and Solar Systems planets. Semimajor axis is given in units of AU and planet mass in units of an Earth radius $R_\oplus$. All exoplanets are shown as grey circles, while super-Earths and mini-Neptunes are highlighted as red squares. Solar System planets are shown with filled circles. The majority of super-Earths and mini-Neptunes reside between 0.03 AU and 0.3 AU of their host star. Exoplanet data from the NASA Exoplanet Archive (Akeson, Chen, Ciardi et al. 2013).

Super-Earths and mini-Neptunes for which both mass and radius measurements have been secured appear to have compositions consistent with most of the planet's mass in a rocky or heavy element core.[†]

[†]Throughout this thesis, we use the term "core" to refer to the underlying, non-gaseous mass in a planet. This astrophysical usage is in contrast to the geological use of the term, where "core" typically refers to the innermost region of a planet, regardless of its composition.

In some cases, the planet appears to be similar in composition to a terrestrial planet, only larger but still with a negligible gaseous envelope (Dressing, Charbonneau, Dumusque et al. 2015); given our fundamental desire to find planets similar to Earth, these planets probably offer our best hope of detecting life around other stars. In other cases, the planet appears to have a volatile-rich composition comprising up to several percent of the planet's total mass. This can be inferred by constructing a mass-radius diagram in which curves of constant composition are compared to observed mass-radius data (Weiss & Marcy 2014). In most cases, it is not sufficient to explain observed planet radii with just a heavy element composition (Fig. 1-4), requiring some light element envelope overlying a heavy element core.

The seemingly unexpected presence of these planets close to their host stars, however, has forced us to reevaluate aspects of planet formation theory. For those interested in finding Earth analogues and detecting life, a greater understanding of the formation and evolutionary history of super-Earths and mini-Neptunes will give us greater insight into the amenability for life to exist on these planets and—given their apparent ubiquity—in the Universe as a whole. This understanding will shed light on how rare or common we expect certain physical processes to be in planet formation.
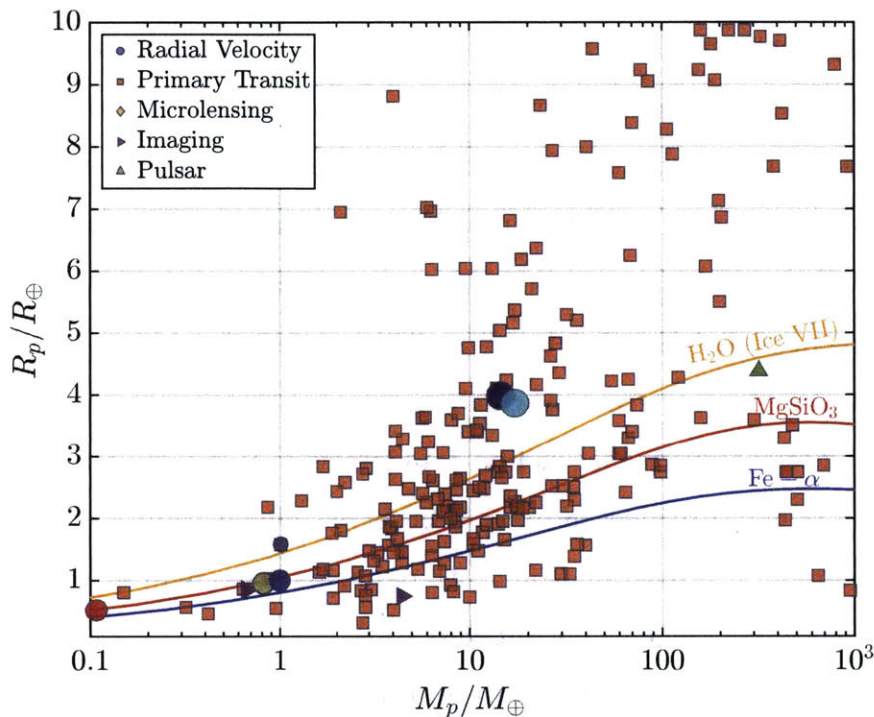
Figure 1-4: Mass-radius diagram of observed exoplanets with lines of constant composition and Solar System planets shown for reference. Curves of constant composition are shown for water (ice VII), perovskite (MgSiO$_3$), and ferrite (Fe $-\alpha$). Planets above the water line (and perhaps even those above the perovskite line) have significant hydrogen/helium envelopes. Isocomposition curves generated using equation of state data from Seager, Kuchner, Hier-Majumder et al. (2007) with planet masses and radii taken from the `exoplanet.eu` database (Schneider, Dedieu, Le Sidaner et al. 2011). The fact that most planets have radii greater than those for a rock or ice composition indicate that they have a light element envelope overlying a heavy element core.[‡]

[‡]For clarity, we have shown planet masses here without error bars. However, many of the reported planet masses have 1σ uncertainties approaching—or in some cases exceeding—100% (e.g. Weiss & Marcy 2014). As a result, care must be taken in interpreting this diagram, especially for planets whose masses are such that they lie below the Fe $-\alpha$ line, which is not physically realizable.

1.2 The Planet Formation Narrative

Piecing together a self-consistent formation history of planets is complicated by numerous factors. To first order, planet formation depends on the availability of material from which to build planets in the protostellar nebula and, after collapse of the initial gas cloud, in the protoplanetary disc. The gas disc surface density Σ_g and the metallicity Z in a protoplanetary system dictate the amount of gaseous and solid material, respectively, available to build planets. In addition, the amount of solid material that can ultimately agglomerate into a protoplanetary core is also dependent upon its location in the disc (parameterized by the semimajor axis a), and the mass of the host star $M_\star$. The growth and evolution of a planet are likewise mitigated by its interactions with its surroundings: a planet does not grow in isolation, but in the presence of a nebular gas disc and in the presence of other planetary and protoplanetary bodies. Interactions with both can have important consequences for the final architecture of a planetary system and the properties of observed exoplanets, leading to, for example, core-nucleated gas accretion (e.g., Stevenson 1982), and orbital migration and resonant interactions (Cossou, Raymond, Hersant et al. 2014). Both during and after the assembly phase, a planet continues to evolve with respect to its surroundings, ultimately influencing the observations we make of it and our interpretation thereof. To fully understand the observed population of exoplanets, then, it does not suffice to focus only upon (say) the dynamical interactions between planetary bodies, or gas accretion, but

we must consider a whole host of dynamical processes. This is especially complicated by the fact that many of the relevant physical timescales (e.g., the Kelvin-Helmholtz timescale for the cooling of a gaseous atmosphere, the dynamical interaction timescale, and the gas disc lifetime) are similar.

The purpose of this thesis is to better understand the variety of physical processes important to the formation and evolution of planetary systems, so that we can determine their effects and outline viable pathways that lead to the observed characteristics of planetary systems. Given the prevalence of super-Earths and mini-Neptunes, the focus of the first part of this thesis will be on exploring how this population of exoplanets formed and evolved. In the second part of this thesis, we consider planet formation within our own Solar System, in particular by outlining how observations of asteroids can help shed light on aspects of planet formation in our Solar System.

Based on the above discussion, it is worthwhile to summarize the general narrative of planet formation in slightly more detail. In Sec. 1.3, we consider planet formation within the context of the Solar System and in Sec. 1.4 discuss what asteroids and meteorites tell us about our Solar System, before closing the chapter.

1.2.1   The Gas Disc

Planet formation and evolution passes through a number of phases, beginning with a protostellar nebula and small micrometer-sized dust

grains, and ultimately resulting in fully assembled planetary bodies. It is comprised of the evolution of solid material in the gas disc, the gas itself, and the interaction between the gas and the solid material (see Fig. 1-5). The gas disc results from the collapse of the protostellar nebula, which itself is the remnant of an unstable galactic fragment.

The nonzero angular momentum of the initial nebula is conserved during collapse, so that the nebula becomes a disc (Fig. 1-6). The collapse of the disc happens on a relatively short timescale. A typical protostellar nebula might have a total mass of $\sim 1 M_\odot$ and a radius on the order of 10^4 AU (= 0.1 pc; Hartmann 2009), yielding a density $\rho_{\mathrm{neb}} \sim 10^{-17}$ kg/m^3. For a nebula in which gas pressure can be neglected, the collapse timescale is $\sim 1/\sqrt{G\rho_{\mathrm{neb}}} \sim 10^5$ yr. In reality, the pressure support of the nebula cannot be ignored, and the picture is rather more complicated. Nevertheless, more sophisticated models including pressure support, magnetic fields, and rotation (e.g., Shu 1977, Galli & Shu 1993) yield similar results for the collapse timescale of the protostellar nebula.
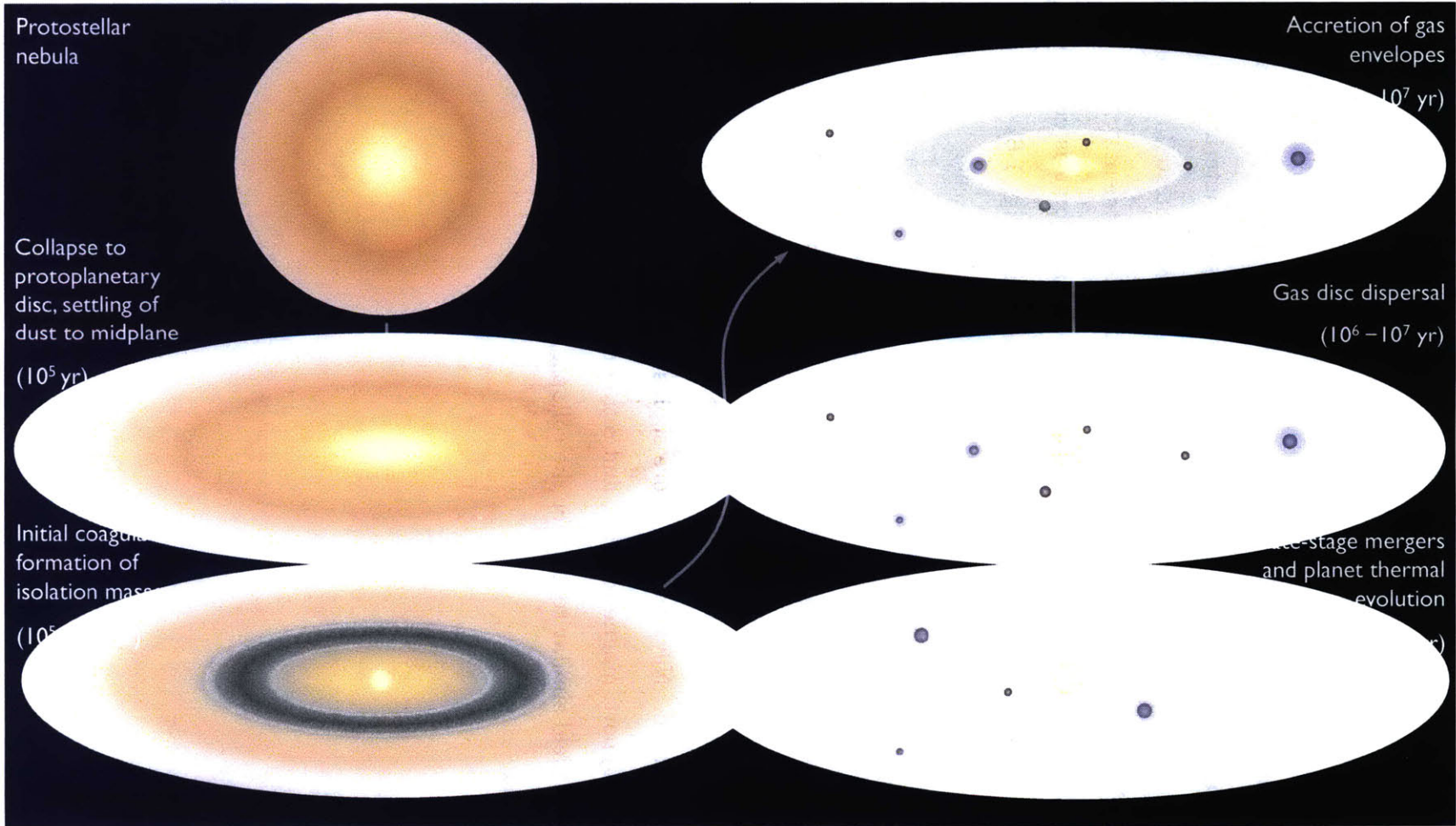
Figure 1-5: Planet formation narrative with relevant timescales indicated. For dynamical processes, the timescales shown are for ~ 1 AU assuming a minimum mass Solar nebula (see Sec. 1.3).
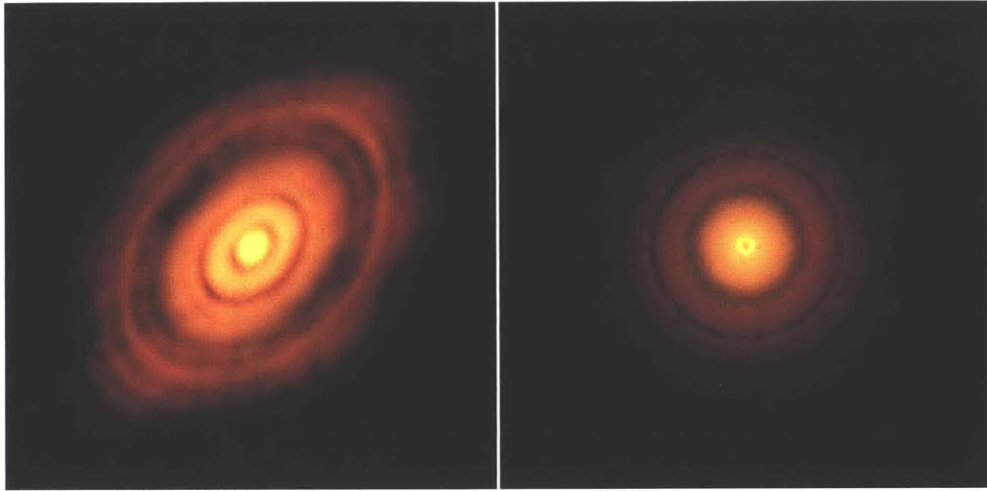
Figure 1-6: Observations of circumstellar discs taken by ALMA. On the left is HL Tauri, while on the right is TW Hydra. Both images from the ALMA Science Portal (Accessed: 2016-05-01).

As the nebula coalesces into a disc, the densest, hottest material in the center forms a star. While the timescale for the collapse of the nebula into a disc is itself relatively short, the lifetime of the disc itself is somewhat longer. During the disc lifetime, the gas accretes onto the host star. Gas accretion onto the host star, in addition to photoevaporative erosion of the disc by extreme ultraviolet (XUV) and X-ray radiation from the energetic host star, are ultimately responsible for the dispersal of the gas disc (e.g., Armitage 2013). Typical gas disc lifetimes are usually deduced by observing individual star-forming regions and determining the fraction of stars in that region that still possess discs. The age of that cluster may be deduced from stellar evolution models, or from dynamical considerations, by measuring the velocity dispersion of the stars in the cluster and mapping that back to an interaction timescale and hence

29

cluster lifetime. In any case, the typical age of protoplanetary discs measured in this way is on the order to $1 - 10$ Myr (e.g., Haisch, Lada, & Lada 2001, Mamajek 2009; see Fig. 1-7). More recently, measurements of remanent magnetism in meteorites have constrained the age of our own protostellar nebula, since the presence of a nebular magnetic field is incumbent upon the presence of the nebular gas. Such measurements have constrained the age of our own disc to $\lesssim 4$ Myr (Wang, Weiss, Downey et al. 2015), such that the Solar System disc age falls within the typical range.
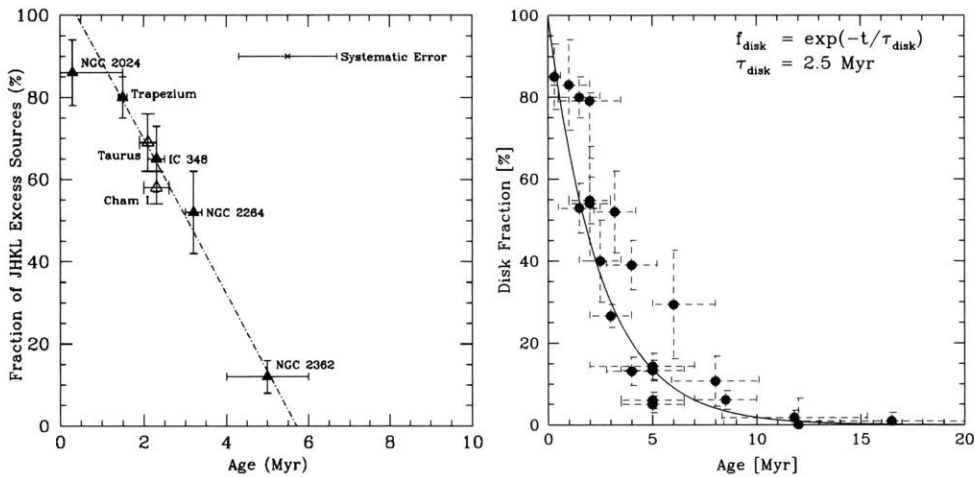
Figure 1-7: Protoplanetary disc lifetimes as determined by observation. The left panel is data from Haisch, Lada, & Lada (2001), while that on the right is from Mamajek (2009). In both cases, the typical disc lifetime is found to be on the order of $1 - 10$ Myr and is inferred from the fraction of stars in clusters showing evidence of circumstellar discs.

The timescales associated with many planet formation processes are constrained by the $1 - 10$ Myr gas disc lifetime. In particular, any process that involves interaction of the planet with a significant amount

of gaseous material must take place within the disc's lifetime. This includes the coagulation of enough solid material to accrete planets with a significant amount of mass in gas, including gas giants and Uranus or Neptune analogues, as well as the drift and migration of planetary-mass bodies through interaction with the gas disc. Both of these will have an impact in our work below. The $\sim 1 - 10$ Myr timescale also has a significant impact on aspects of planet formation and disc evolution which we briefly discuss here. In particular, disc lifetime measurements suggest an average mass accretion rate onto the host star of $\sim 10^{-7}$ $M_{\odot}/\mathrm{yr}$ if roughly 1% of the total system mass is in the gas disc itself (with the remaining 99% in the star); such an estimate based on core collapse models is supported much more directly from spectral measurements, with mass accretion rates onto young T Tauri stars typically $10^{-6} - 10^{-8}$ $M_{\odot}/\mathrm{yr}$ (Hartigan, Edwards, & Ghandour 1995). Since such accretion rates cannot be explained simply due to the molecular viscosity of the gas in the disc, means of angular momentum transfer in the disc other than viscous diffusion have been investigated extensively, including magnetic turbulence (Balbus & Hawley 1991) and gravitational instability (Gammie 2001).

Qualitatively, the standard picture of the gas disc about its host star is one in hydrostatic equilibrium and, at a given semimajor axis a, vertically isothermal at a temperature $T = T(a)$. Assuming an axially sym-

metric geometry, force balance then gives (Armitage 2013)

$$-\frac{v_g^2}{a} = -\frac{GM_\star}{a^2} - \frac{1}{\rho_g}\frac{\mathrm{d}P}{\mathrm{d}a}, \tag{1.1}$$

where v_g is the ciruclar speed of the gas, $M_\star$ is the mass of the host star, ρ_g is the local gas density, P is the gas pressure, and G is the gravitational constant. Simplifying then gives

$$v_g = v_K\sqrt{1-\eta} \approx v_K\left(1-\eta/2\right), \tag{1.2}$$

where v_K is the Keplerian speed $\sqrt{GM_\star/a}$ and $\eta \equiv -(a/\rho)\,\mathrm{d}P/\,\mathrm{d}a$. Taylor expansion on the right-hand side is permitted since η is typically on the order of 0.01. Since the pressure gradient is negative, the circular speed of the gas is sub-Keplerian. Solid bodies that are decoupled from the gas will then experience a head wind that will tend to remove angular momentum from the orbit (in the form of a drag-induced torque), hence shrinking the orbit.

Since the disc is assumed to be locally isothermal, there is an associated scale height $H_g \equiv c_s/\Omega$, where c_s is the sound speed in the gas and $\Omega = \sqrt{GM_\star/a^3}$ is the Keplerian orbital frequency. For a gas, the sound speed is given by

$$c_s = \sqrt{\gamma_a \frac{k_\mathrm{B}}{\mu m_\mathrm{p}} T}, \tag{1.3}$$

where γ_a is the adiabatic index of the gas, k_B is the Boltzmann con-

stant, μ is the molecular mass in atomic mass units, m_{p} is the proton mass, and T is the temperature of the gas. In this case, hydrostatic equilibrium yields for the vertical density structure of the disc $\rho_g = \rho_{g,0} \exp[-z^2/(2H_g^2)]$, where z is the distance (up or down) measured from the disc midplane and $\rho_{g,0}$ is the gas density at the midplane. Integrating this expression over all z gives for the surface density of gas $\Sigma_g = \rho_{g,0}\sqrt{2\pi}H_g$, so that the volumetric density at the midplane is roughly given by the surface density divided by the disc thickness $2H_g$. In our work below, we will likewise assume that $\rho_{g,0} \approx \Sigma_g/(2H_g)$.

How is the density structure of the early Solar System inferred? The standard procedure is to take all the mass in the present planets, and to spread the mass amongst annuli that extend to the midpoints between the planets' orbits (Weidenschilling 1977, Hayashi 1981). For the terrestrial planets, the planet masses are enhanced by a factor inverse to the Solar metallicity, in order to bring the composition to that of the Sun (and by assumption, the primordial disc). The resulting surface density Σ is then typically fit to a power law, with the power law proportional to $a^{-3/2}$ (see Fig. 1-8). This model is referred to as the "minimum mass Solar nebula" (MMSN); the density profiles given for the MMSN are typically

$$\Sigma_g = \Sigma_{g,0}(a/\text{AU})^{-3/2}, \tag{1.4}$$

$$\Sigma_{g,0} = 1400 \text{ g/cm}^2;$$

$$\Sigma_s = \Sigma_{s,0}(a/\text{AU})^{-3/2}, \tag{1.5}$$

$$\Sigma_{s,0} = 7 \text{ g/cm}^2,$$

where here, a metallicity of $1/200$ has been assumed (Hayashi 1981). The prefactor $\Sigma_{s,0}$ is enhanced by a factor of a few beyond the ice line. A survey of exoplanetary systems by Raymond & Cossou (2014) has shown that the power law exponent for surface density profiles generated in this way for exoplanetary systems is far from universal (e.g., Chiang & Laughlin 2013), although a "median" power law exponent is -1.45, similar to the MMSN's $-3/2$. These calculations assume that the current distribution of planets as a function of semimajor axis has remained unchanged from the time of planet assembly, an assertion which we ultimately find inconsistent with observed characteristics of exoplanets (Chapter 4).
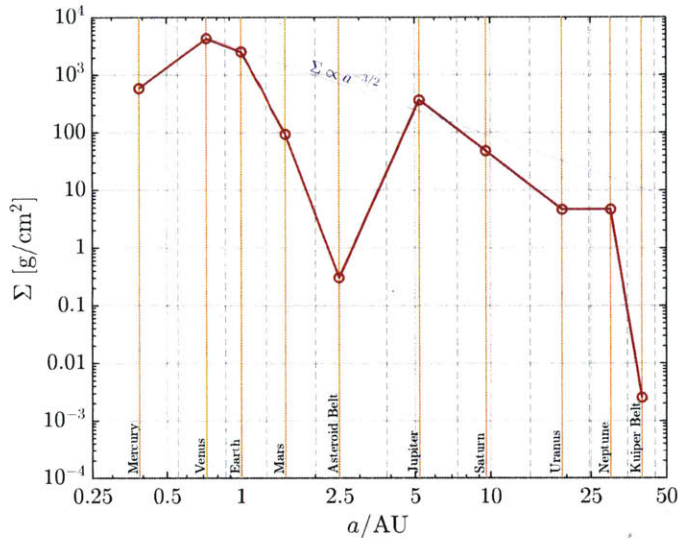
Figure 1-8: The minimum mass Solar nebula (MMSN). The surface density dependence on orbital radius for the MMSN from power law fits is typically found to be $\Sigma \propto a^{-3/2}$ (blue line).

1.2.2 The Evolution of Solids

Concurrent with the accretion of the gas in the disc onto the host star is the evolution of the solid mass in the disc. To first order, the fraction of solid mass available in the disc is simply the metallicity Z of the system. For most of our work, we will assume this, although in reality the picture is more nuanced and depends on the thermochemistry of the disc (see also the discussion of the snow line in Sec. 1.3). The solid material in the disc is initially in the form of small grains, on the order of a micron in size (Armitage 2013). These grains settle towards the disc midplane where eventually they accumulate and form into planetesimals, on the

order of meters or larger in size. How the dust accumulates into objects larger than a meter or so is a matter of ongoing work, with the physical mechanisms preventing growth beyond this size known as the "meter barrier" (e.g., Youdin & Kenyon 2013).

The meter barrier refers to two distinct phenomena. One relates to collisions between dust fragments on the order of one meter in size. In this case, collisions between dust tend to be destructive rather than constructive (Blum & Wurm 2008): meter-sized masses are well-decoupled from the gas, so that they attain higher relative velocities. Therefore in principle, dust particles that have coagulated to this size cannot grow larger due to collisions. The other meter barrier refers to the fact that meter-sized particles experience a significant headwind due to the sub-Keplerian speed of the gas [Eq. (1.2)]. Here, particles are not small enough that they couple quickly and efficiently to the gas (in which case they travel at v_g) and they are not large enough to have a significant amount of momentum to resist the oncoming headwind. These particles experience a significant amount of drag, which subsequently leads to the removal of angular momentum from the particle's orbit, so that the timescale for radial drift inwards can be less than the dynamical interaction timescale. In particular, in the inner disc at ~ 1 AU, infall occurs on timescales of $10^2 - 10^3$ yr (Armitage 2013).

Most favored lines of research for overcoming the meter barrier appeal to some form of gravitational instability that collects solid material in such a way that the wavelength of the lowest frequency mode is larger

than a meter (e.g., Goldreich & Ward 1973, Sekiya 1998, Youdin & Shu 2002). In any case, solid bodies large enough to interact with both the gas disc and with other solid bodies must form within the lifetime of the gas disc, after which the bodies undergo dynamical evolution. After this point, planet formation passes through several distinct dynamical stages, which we briefly review below in order to motivate some of our arguments later on. Although we focus on analytical arguments here, these dynamical stages are also found in numerical simulations of planet formation. The discussion below focuses mostly on gravitational interactions between planets, protoplanets, or planetesimals. The dynamical evolution, however, is also mediated by interaction with the gas disc, which is important for both the dynamical evolution of planetary orbits and the interaction of protoplanets; we highlight below situations in which this might be important. Further details may be found in Goldreich, Lithwick, & Sari (2004).

1.2.3 Dynamical Interactions of Protoplanets

We consider a simplified picture with large bodies (or protoplanets) of mass M and small bodies (or planetesimals) of mass m. The bodies are on orbits with semimajor axis a, and hence mean motion or orbital frequency $\Omega = \sqrt{GM_\star/a}$, where $M_\star$ is the mass of the star. The orbits will have deviations from circularity in the form of eccentricity and inclination, information for which is contained in the velocity dispersions v and u for large and small bodies, respectively.

The disc surface density of large and small bodies, respectively, is given by Σ_s and σ_s. To first order, these are related to the gas surface density Σ_g by the disc metallicity Z: $\Sigma_g Z \sim \Sigma_s + \sigma_s$.

Consider the accretion of small bodies by large bodies, and assume for simplicity that all solids of interest are of similar material density ρ_m so that $M \sim \rho_m R^3$ and $m \sim \rho_m s^3$, where R and s are the radii of large and small bodies, respectively. For a body of mass M and radius R encountering a swarm of smaller bodies of mass m with velocity dispersion u and number density n, the rate of interaction $\mathcal{R}_{\text{int}}$ in the absence of gravity is $\mathcal{R}_{\text{int}} = n\pi R^2 V = (\rho_s/m)\pi R^2 u$ where ρ_s is the volumetric mass density of small bodies. The volumetric mass density of small bodies can be related to its surface density σ_s by considering its scale height $H_s \sim \Omega/u$ so that the number of interactions per unit time are given by (dropping factors of order unity) $\mathcal{R}_{\text{int}} \sim \sigma_s \Omega R^2/m$. If we are interested in the rate of increase in mass of a larger body $M^{-1} \, \mathrm{d}M/\,\mathrm{d}t$ due to collisions with smaller bodies, we have

$$\frac{1}{M}\frac{\mathrm{d}M}{\mathrm{d}t} \sim \frac{\sigma_s \Omega}{\rho_m R},$$

where we have used $M \sim \rho_m R^3$. With this assumption, $M^{-1} \, \mathrm{d}M/\,\mathrm{d}t \sim R^{-1} \, \mathrm{d}R/\,\mathrm{d}t$ to within a factor of 3. This interaction rate assumes the interaction cross-section is the geometrical cross-section. However, R can be enhanced due to gravitational focusing. The impact parameter b when gravitational focusing is included is $b = R\sqrt{1 + (v_{\text{esc}}/u)^2}$, where

$v_{\text{esc}} = \sqrt{2GM/R}$ is the escape speed of the larger body. Hence, for velocity dispersion $u \gtrsim v_{\text{esc}}$, $b \sim R$ while for $u \lesssim v_{\text{esc}}$, the effective radius is enhanced to $b \sim R(v_{\text{esc}}/u)^2$. Then,

$$
\frac{1}{R} \frac{dR}{dt} \sim
\begin{cases}
\dfrac{\sigma_s \Omega}{\rho_m R}, & u \gtrsim v_{\text{esc}} \\[2ex]
\dfrac{\sigma_s \Omega}{\rho_m R} \left(\dfrac{v_{\text{esc}}}{u} \right)^2. & u \lesssim v_{\text{esc}}
\end{cases}
\tag{1.6}
$$

There are additional velocity regimes (for instance, when $u \lesssim v_H$, the Hill velocity of the larger body) which we ignore here. Since $v_{\text{esc}} \propto \sqrt{M/R} \sim R$, when $u \lesssim v_{\text{esc}}$, $R^{-1} \, dR/dt \propto R$: larger bodies grow faster than bodies smaller than them. This type of growth is termed "runaway growth".

The growth of planets is ultimately mediated by the velocity dispersions of large and small bodies. The velocity dispersion includes contributions from both deviations from circularity (eccentricity and inclination) and from Keplerian shear. Larger bodies have velocity dispersions v that are both excited by similar-sized bodies (through the process of viscous stirring) and damped by smaller bodies (dynamical friction). Smaller bodies have velocity dispersions u that are only appreciably excited by larger bodies; this is the case so long as the collision timescale for smaller bodies is longer than the timescale for larger bodies to excite u, since mutual collisions between smaller bodies (which can damp u) are not yet important.

The rates for velocity evolution are

$$\frac{1}{u}\frac{du}{dt} \sim \frac{\Sigma_s \Omega}{\rho_m R}\left(\frac{v_{esc}}{u}\right)^4 \tag{1.7}$$

$$\frac{1}{v}\frac{dv}{dt} \sim \frac{\Sigma_s \Omega}{\rho_m R}\left(\frac{v_{esc}}{v}\right)^4 - \frac{\sigma_s \Omega}{\rho_m R}\left(\frac{v_{esc}}{u}\right)^4. \tag{1.8}$$

In the first equation, the velocity dispersion of small bodies is increased due to viscous stirring by larger bodies. In the second equation, on the other hand, the velocity dispersion of larger bodies is increased by viscous stirring by other large bodies, but decreased due to dynamical friction from smaller bodies.

By examining the above equations qualitatively, we make the following general statements. The relevant timescale τ_{sys} is set by the mass growth rate of the largest body or bodies, when runaway growth is attained: $\tau_{sys}^{-1} \equiv R^{-1}\,dR/\,dt$. The corresponding velocity dispersion of small bodies is found by setting this rate equal to the small body velocity dispersion rate: $\tau_{sys}^{-1} \sim u^{-1}\,du/\,dt$. This yields $v \sim u(\Sigma_s/\sigma_s)^{1/2}$.

The system is excited until $v_{esc} \sim u$ and gravitational focusing ceases to be effective in accreting small bodies onto larger ones; this also implies then that the age of the system is set by the condition that $\Sigma_s \sim \sigma_s$. At any given time, the velocity dispersion of large bodies proceeds quasi-statically due to dynamical friction, so that $v^{-1}\,dv/\,dt \sim 0$ and $v \sim u(\Sigma_s/\sigma_s)^{1/4}$. Evolution under this picture ends when the largest bodies dominate their local feeding zones to accrete smaller bodies. The resulting masses are "isolation masses" M_{iso} whose masses are given by

adding up all the solid mass within a feeding zone (parameterized as a multiple of Hill radii). Doing so yields the expression (Lissauer 1993)

$$M_{\text{iso}} = \left[\frac{4\pi a^2 C \Sigma_s}{(3M_\star)^{1/3}} \right]^{3/2}, \qquad (1.9)$$

where C is a constant that describes the feeding zone's extent in units of Hill radii. In Fig. 1-9, we show isolation masses as a function of semimajor axis a for three cases in which the power-law dependence of the disc solid surface density $\Sigma_s \propto a^{-3/2}$ as in the minimum mass Solar Nebula model.

When isolation is attained, there may be perhaps a few dozen massive bodies separated by several mutual Hill radii. At 1 AU and with a minimum mass Solar nebula density profile, M_{iso} is roughly a Mars mass. These protoplanets can then excite one another into orbit crossing and hence giant impacts. For a collision, the impact speed v_{imp} is set by energy conservation:

$$v_{\text{imp}}^2 = v_{\text{esc}}^2 + v_\infty^2, \qquad (1.10)$$

where v_∞ is the speed of the impactor at infinity (its velocity dispersion). Collisions start to dominate over scattering when the collision rate equals the viscous stirring rate:
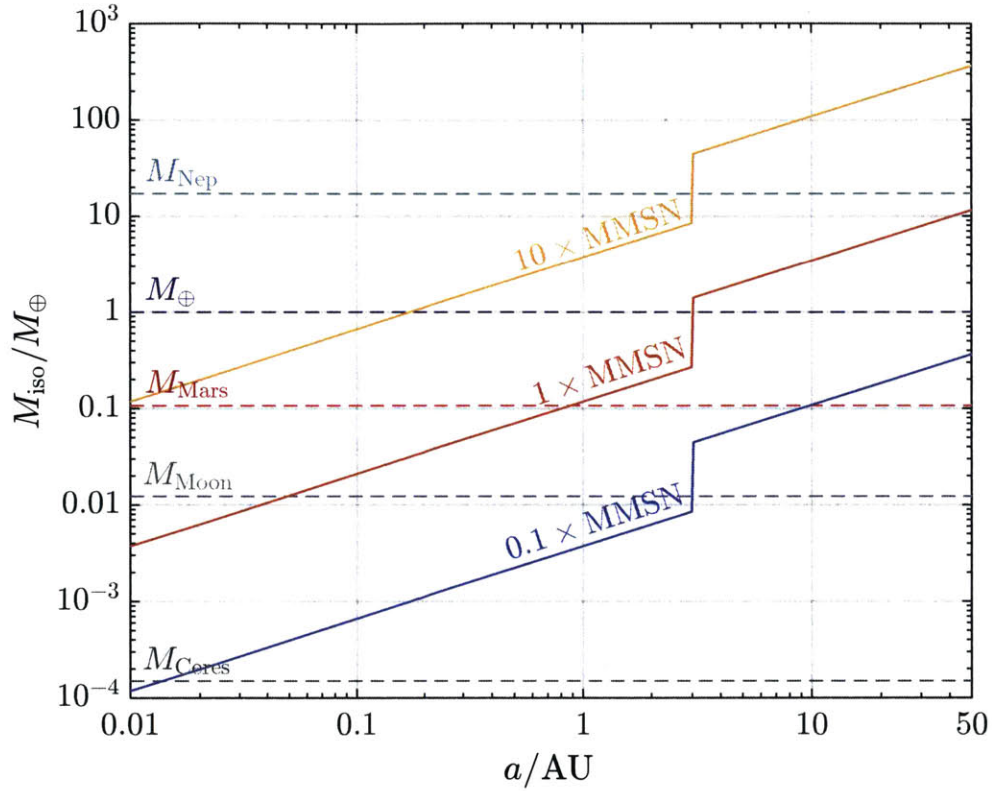
Figure 1-9: Isolation masses M_{iso} as a function of semimajor axis a. For each line, we assume $\Sigma_s \propto a^{-3/2}$ surface density distribution, while taking different $\Sigma_{s,0}$ values compared to that of the minimum mass Solar nebula. The masses of the asteroid 1 Ceres, the Moon, Mars, Earth, and Neptune are shown for comparison. At 1 AU, for a MMSN-type disc, an isolation mass is roughly a Mars mass.

$$\mathcal{R}_{\text{int,vs}} = \frac{\Sigma_s \Omega}{\rho_m R} \left(\frac{v_{\text{esc}}}{u} \right)^4$$

$$= \mathcal{R}_{\text{int,coll}}$$

$$= \frac{\Sigma_s \Omega}{\rho_m R} \left(\frac{v_{\text{esc}}}{u} \right)^2 , \tag{1.11}$$

where all bodies after isolation are now assumed "large". Hence, the velocity dispersion $v_\infty = u$, and at the onset of impacts, $u = v_\infty \sim v_{\text{esc}}$, so that at most, the impact speed $v_{\text{imp}} \sim \sqrt{2} v_{\text{esc}}$. We stress, however, that these collisions can only occur once the gas disc has dissipated to such an extent that the surface density of gas is similar to the surface density of solids: $\Sigma_s \sim \Sigma_g$. Intuitively, this is a result of the fact that a solid body no longer encounters a sufficient amount of mass in gas to oppose its own inertia. The rate of interactions for isolation masses may be estimated from the above expression for small bodies, but with the velocity dispersion for large bodies substituted for that of small bodies:

$$\mathcal{R}_{\text{int}} \sim \frac{\Sigma_s \Omega}{\rho_m R}. \tag{1.12}$$

Since in the minimum mass Solar nebula, $\Sigma_s \propto a^{-3/2}$ and $\Omega \propto a^{-3/2}$, $\mathcal{R}_{\text{int}}$ scales strongly with a, $\mathcal{R}_{\text{int}} \propto a^{-3}$.

1.3 The Solar System

As we have remarked, the architecture of observed exoplanetary systems appears radically different from our own. Nevertheless, features of the Solar System give important clues to processes that may be important to planet formation, so that our Solar System is a worthwhile starting point. An obvious fact about our Solar System is that it seems relatively ordered in the sense that rocky, terrestrial planets are in the inner Solar System (interior to a few AU), with the two gas giants Jupiter and Saturn, and the ice giants Uranus and Neptune following. In the picture of planet formation outlined above, Earth and Venus formed from a series of giant impacts, with final planet masses being attained in perhaps a few million to tens of millions of years. In the case of the Earth, the final giant impact gave rise to the moon (Ćuk & Stewart 2012, Canup 2012), although there is still a controversy surrounding the specifics of this impact (see, e.g., Pahlevan & Stevenson 2007, Mastrobuono-Battisti, Perets, & Raymond 2015). The gas giants, on the other hand, formed beyond the water ice line, beyond a few AU (Lewis 1972). Here, isolation masses are massive enough to instantiate core-nucleated instability, wherein the mass of gas that can be gravitationally bound by the core is on the order of the core mass itself, hence contributing significantly to the gravitational mass and leading to runaway accretion (Stevenson 1982). Gas giants therefore possess significant *primordial* H/He envelopes derived from the gas disc, while the terrestrial planets do not; their *secondary* atmospheres are derived instead from internal

processes such as outgassing, or relatively late delivery of volatiles from elsewhere. Terrestrial mass-planets or their embryos are often too small to capture any significant primordial envelope from the disc (see Section 4.2.3, Fig. 4-3). Terrestrial planet envelopes that were captured from the nebula may have been eroded due to interaction with the Solar wind (Chassefière & Leblanc 2004), the Moon-forming giant impact (Halliday 2008), a barrage of late impacts by smaller bodies (Schlichting, Sari, & Yalinewich 2015), Jeans escape (Hunten 1973), or some combination thereof.

Beyond the gas giants lie Uranus and Neptune, whose masses are thought to be mostly solid, with about $10 - 20\%$ in low molecular mass envelopes, although the interior structure of both is still highly uncertain (Helled, Anderson, Podolak et al. 2011). The masses of Uranus and Neptune are consistent with the formation of their underlying high molecular mass cores as isolation masses at their current semimajor axes. Indeed, substituting the MMSN values from Eq. (1.5) into Eq. (1.9) yields

$$M_{\text{iso}} = \begin{cases} 0.02 M_\oplus \times (a/\text{AU})^{3/4} & \text{(Interior to ice line)}, \\ 0.2 M_\oplus \times (a/\text{AU})^{3/4} & \text{(Beyond ice line)}. \end{cases} \qquad (1.13)$$

Based on the framework we have outlined above, it appears that the Solar System formed as we might expect in a very general sense. However, the question remains to what extent the framework is self-

fulfilling. Certainly most of the physical arguments above must be correct, based both on analytical and numerical arguments. Nevertheless, as we have remarked, observed exoplanets and exoplanetary systems are markedly different from our own, and so there must be some differences in the overall picture. For instance, we have totally neglected disc-driven migration, which has been shown theoretically to be an efficient mechanism for bringing planets from the outer disc into the inner disc (e.g., Ward 1997). In the case of our own Solar System, perhaps the assembly of rocky cores in the outer disc occurred late enough that disc-driven migration did not play a significant role in sculpting the Solar System's architecture [although *planetesimal*-driven migration has been shown to be a possible source for early rearrangement of the Solar System (Walsh & Morbidelli 2011)]. The key question we consider in Part I of this thesis is then how does the framework explaining the Solar System's architecture fit into the larger context of planet formation now that we have thousands of other observed planetary systems?

1.4   Asteroids and Meteorites

While in our Solar System, the large bodies tell us much about planet formation and evolution, the small bodies (such as comets and asteroids) can reveal, too, clues about how planetary systems form and evolve. Bodies which are large enough to assume hydrostatic equilibrium and to undergo sufficient heating to differentiate and be geologically active have

many of the details of their formation history either lost or incapable of being observed. Small bodies, on the other hand, can reveal details about the thermochemical and dynamical environments in which they formed. In some cases, these smaller bodies may have originated as broken pieces from a larger body, thus giving insight into the impact history of Solar System bodies and their subsequent evolution thereof.

In this section, we briefly review aspects of asteroids and meteorites (most of which are thought to have originated from the Asteroid Belt) that are important to planet formation in our own Solar System. In anticipation of Part II of this thesis, we discuss how greater insight into the nature of these bodies can give us insight into how our Solar System and other planetary systems formed and evolved.

1.4.1 Meteorites

Meteorites come broadly in two flavors, "chondrites" and "achondrites" (Weisberg, McCoy, & Krot 2006). Chondrites are those whose elemental compositions match essentially those of the Sun (or more properly, the Solar photosphere), with the exception of the most highly volatile elements. Achondrites on the other hand are meteorites that have been significantly processed, and are either melts, partial melts, or melt residue of igneous rock, or are brecciated fragments from differentiated bodies. Some meteorites have structures characteristic of achondrites, but chemistries more similar to chondrites, and are called primitive achondrites. Based on these broad similarities, different meteorite specimens

are grouped into *meteorite classes* and subclasses (see Fig. 1-10). Given their compositional similarity to the Solar photosphere, chondrites are believed to represent primitive unprocessessed or unmelted material from early in the Solar System's formation history.

The similarities amongst meteorites in a given class or subclass suggest a common origin. In some cases, a *parent body* from which meteorites in a given subclass originated can be identified. For instance, the achondritic eucrites, diogenties, and howardites have chemical signatures suggesting that they originated in the main best asteroid Vesta, while the shergottites, nakhlites, chassignites, and orthopyroxenites are believed to have originated on Mars. Otherwise, although there are some outstanding discrepancies in tightening the picture (Chapman 1996), it is generally accepted that most meteorites originate from asteroids. This fact has important consequences, since meteorites are directly accessible to us for in-depth study on Earth. If we can identify asteroids that are similar to the meteoritic specimens on Earth, we can then begin to make inferences about the nature of the asteroids from which the meteorites originated.
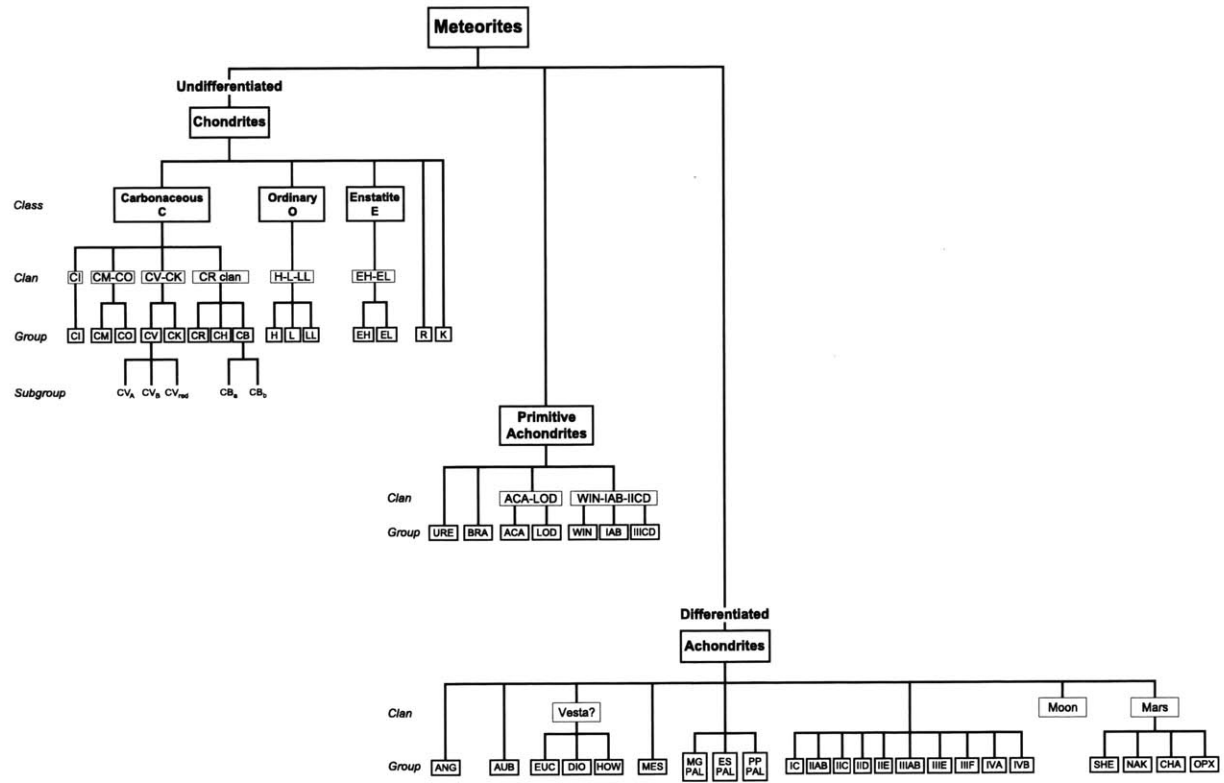
Fig. 1. Diagram expressing the systematics of meteorite classification and showing the major meteorite divisions, classes, clans, and groups and relationships among meteorite groups. URE — ureilite, ACA — acapulcoite, LOD — lodranite, ANG — angrite, AUB — aubrite, BRA — brachinite, WIN — winonaite, HED — howardite-eucrite-diogenite, MES — mesosiderite, MG PAL — main-group pallasite, ES PAL — Eagle Station pallasite, PP PAL — pyroxene pallasite, SHE — shergottite, NAK — nakhlite, CHA — chassignite, OPX — orthopyroxenite.

Figure 1-10: Meteorite classification scheme, indicating chondrites, achondrites, and primitive achondrites. Figure taken from Weisberg, McCoy, & Krot (2006).

1.4.2 Asteroids

Determining the links between asteroids and meteorites is one of the primary goals of both asteroid and meteorite science. Asteroids, like meteorites, are classified based on spectral characteristics (Bus, Vilas, & Barucci 2002), and comparison between laboratory spectra taken from meteorites and both ground- and space-based observations of asteroids confirms this, with several caveats.[§]

Most asteroids reside within the Main Asteroid Belt between Mars and Jupiter. Nevertheless, an increasing population of asteroids with orbits interior to this, on near-Earth or Earth-crossing orbits—so-called "near-Earth objects" (NEOs)—is being discovered (Binzel, Lupishko, di Martino et al. 2002; see Fig. 1-11). The latter population in particular is of interest for future exploration due to its accessibility and the potential for collisions with Earth in the future. Asteroids display both rich variety and common features, both of which shed light on their history and which reveal the influence of the evolving Solar System around them. We discuss some of these below.

[§]Many of the differences between laboratory meteorite spectra and asteroid spectra can be linked with confidence to space weathering—the chemical alteration of the asteroid surface—by radiation and high-energy particles in space (Bus, Vilas, & Barucci 2002). Other differences appear to be related to the fact that asteroids often possess a deep regolith on their surface. While some connections between specific asteroid types and meteorite classes have been made, others are still debated.
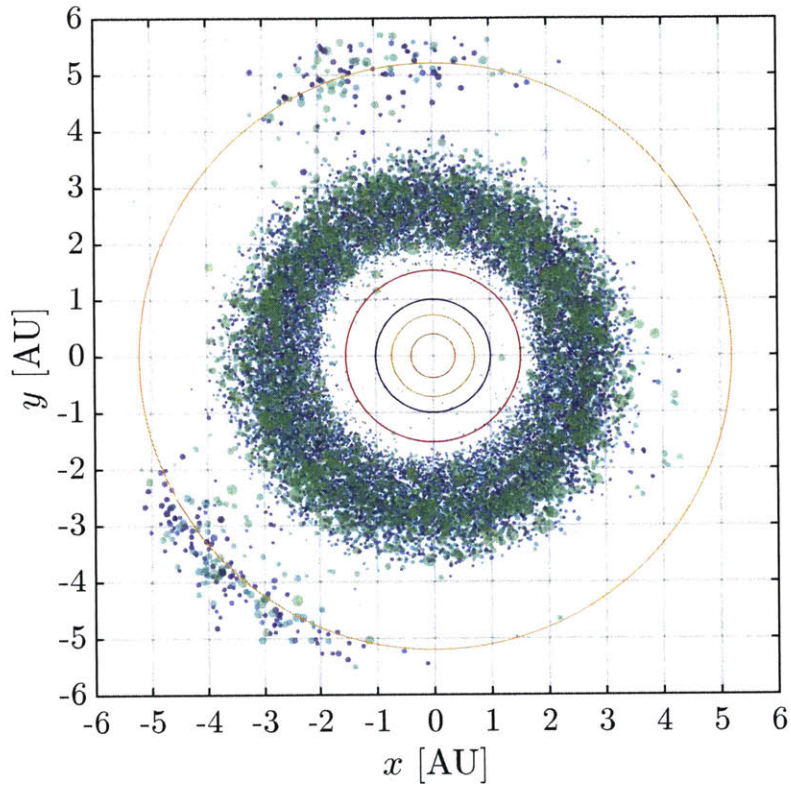
Figure 1-11: The observed population of asteroids in the main belt at an arbitrarily chosen epoch. The colored points indicate the 20,000 largest asteroids in the IAU Minor Planet Center database (International Astronomical Union Accessed: 2016-05-23). The colors indicate absolute magnitude (and hence size), with the dimmest objects shown in blue and the brightest in green. The orbits of Mercury, Venus, Earth, Mars, and Jupiter are indicated by circles. The Trojan asteroids are those clustered on Jupiter's orbit.

1.4.3 Dynamics

The minimum mass Solar nebula model shown in Fig. 1-8 has the striking property that, between the orbits of Mars and Jupiter, there is a significant depletion of mass compared to what a naïve power-law fit of the mass distribution would suggest. Calculations for the accretion timescale of the largest asteroids and their similarity to meteorite solidification ages suggest the Asteroid Belt was more populated in the past, with $\sim 0.1 - 1 M_\oplus$ total mass, compared to about $5 \times 10^{-4} M_\oplus$ at the present (Wetherill 1989; see also Fig. 1-9).

The significant depletion of mass in the Asteroid Belt is now believed to be a result of the dynamical evolution of Jupiter early in the Solar System's history (Petit, Chambers, Franklin et al. 2002), during the late stages of planet formation. The influence of resonances sweeping through the Main Belt (and in fact through the Solar System) had the effect of dynamically exciting the Belt, pumping up eccentricities and inclinations, and also ejecting asteroids from the Solar System. The increase in the random velocities of asteroids also had the effect of creating destructive collisions, generating more and smaller bodies. Many asteroids are known to reside in dynamical families: asteroids with similar orbital elements, suggesting that any one family potentially originated from the same body or from a collision between two bodies.

1.4.4 Structure

The importance of collisions to asteroid formation and evolution, and the subsequently formed meteorite record has further led to the recognition of several important facts and some new conjectures for both asteroids and meteorites. First, in concert with observations of asteroids, it is recognized that a significant fraction of asteroids have either very porous structures or are gravitationally bound agglomerates of rocks, or "rubble piles" (Sánchez & Scheeres 2014). This in turn has led to investigation of the importance of effects such as the Yarkovsky and modified Yarkovksy (or YORP) effects (Bottke, Vokrouhlický, Rubincam et al. 2006, Vokrouhlický, Bottke, Chesley et al. 2015), in which radiation pressure forces alter the orbits and spins of asteroids (Polishook 2014), the latter sometimes leading to breakup. In the case of porous asteroids, pores can have important consequences both for the composition of the asteroid (pores being capable of holding water) and the collisional response of the asteroid (porous rocks behaving differently due to collisions). The fact that impact cratering occurs differently on porous bodies compared to solid bodies has been recognized (Housen & Holsapple 2003). In particular, large craters on asteroids such as 433 Eros that are comparable to the size of the asteroid itself are believed to be evidence of a porous structure, as large impacts on porous bodies tend to compact rather than eject material (Britt, Yeomans, Housen et al. 2002).

The asteroids 1 Ceres (McCord & Sotin 2005) and 4 Vesta (Russell, Raymond, Coradini et al. 2012) are likely differentiated. In the case of

the Vesta, it is believed that the diversity in the HED/Vestoid meteorites are due to different parts of Vesta being ejected due to impacts. More broadly, some of the variety in the meteorite record is now believed to be due to a combination of both glancing impacts and the differentiated structure of a previous generation of asteroids, with iron- or metal-rich meteorites probing interior or metal-rich regions, and rocky or chondritic meteorites the remnants of stripped-off upper layers (Weiss, Berdahl, Elkins-Tanton et al. 2008).

1.4.5 Chemistry

The chemistry of asteroids is largely inferred by visible and near infrared reflectance spectra (Bus, Vilas, & Barucci 2002), and the connection to meteorites is usually made based on the similarities between the reflectance spectra of the two. The analysis of asteroid reflectance spectra is subject to some interpretation, however (DeMeo, Binzel, Slivan et al. 2009), which has led to the recognition, for instance, that space weathering has a significant impact on the surfaces of asteroids. Another alternative remote sensing technique, which is discussed more in depth in Chapter 6, is X-ray spectroscopy, which can provide more unambiguous information regarding the chemical composition of an asteroid via sharp spectral features (Lim & Nittler 2009, Inamdar, Binzel, Hong et al. 2014). There exist in some cases asteroids and meteorites that possess reflectance spectra that are intermediate between different asteroids, suggesting the possibility that some objects originated from

either a transitional object in the protoplanetary disc (Clark, Binzel, Howell et al. 2011) or are the surviving remnants of a collision that led to the disruption of a previous asteroid into the currently observed bodies. These possibilities will be revisited in Chapter 6 in connection to the near-Earth asteroid 101955 Bennu.

1.4.6 Magnetism

Lastly, we briefly touch upon magnetic fields in asteroids. The fact that at least some asteroids possessed ancient core dynamos has been recognized by comparing laboratory experiments investigating the remanent magnetism in meteorites from known parent bodies (e.g., Fu, Weiss, Shuster et al. 2012 in the case of Vesta). In other cases, dating meteorites from a given class and investigating their magnetic properties can reveal more complex behavior. In the case of the angrite parent body, for instance, a low remanent magnetism consistent with the nebular field has been inferred for a relatively young specimen, while a slightly older angrite specimen has a magnetic field consistent with a dynamo, suggesting a delayed or late-onset dynamo (Wang, Weiss, Downey et al. 2015).

1.5 Organization of This Thesis

In the previous sections, we have outlined observations that hold clues to the formation of both our own Solar System and exoplanetary systems.

In this thesis we consider both in two distinct parts, with the overarching goal of constructing a framework that allows us to identify viable formation pathways for planets and planetary systems.

This thesis is comprised of two parts. In Part I, we consider the formation and evolution of exoplanetary systems, with a particular focus on the population of super-Earths and mini-Neptunes. We use what we know about inferred exoplanet bulk composition (namely volatile-rich envelopes and their masses) and bulk structure to motivate potential planet formation pathways. Chapters 2 and 3 provide the requisite background for Chapters 4 and 5. In Chapter 2, we present an overview of the thermal evolution of gaseous envelopes. The discussion is thorough and includes a number of classical results, as well as several new results. The purpose of this chapter is to develop the background necessary for describing how planetary-mass bodies accrete their gaseous envelopes in the presence of the gas disc, and how they evolve after the dissipation of the gas disc. Since we know that the availability of material from which we can form planets depends on the availability of gaseous material, this has important implications for the observation and interpretation of exoplanet properties.

In Chapter 3, we briefly review the theory of hydrodynamic shock waves. As we have mentioned above, late-stage impacts are a regular feature of planet formation theory and figure prominently in planet formation in the inner disc, where interaction rates are relatively high, and where impacts are a natural consequence of viscous stirring. These

impacts have the potential to alter the bulk composition of the planet, especially its volatile mass, which can be partially or completely ejected due to such an impact. We therefore develop a framework by which we can parameterize volatile loss due to a giant impact in a computationally efficient way.

In Chapter 4, we apply the results of the previous two chapters to the formation of close-in super-Earths and mini-Neptunes in the presence of the gas disc. In order to determine whether these close-in exoplanets could have formed *in situ* at their observed semimajor axes, we construct a self-consistent planet formation history including core accretion by isolation masses, envelope mass loss due to giant impacts, and core accretion during gas disc dispersal by a fully assembled planet core.

Our results in Chapter 4 show that core accretion theory predicts a relatively narrow mass-radius (or mass-bulk density) relationship for planets with H/He envelopes. On the other hand, observational data show a substantial amount of diversity in the observed exoplanet population, especially those in tight orbital configurations. While photoevaporation is typically invoked to explain such diversity (Lopez & Fortney 2013a), in reality, the XUV and X-ray activity of young stars is highly uncertain. Furthermore, even after accounting for photoevaporation, in such tightly packed systems, there is far more diversity than expected. In Chapter 5, we explore another method by which to bring about bulk density change. As it has been shown that giant impacts can bring orbital stability to observed, closely packed exoplanetary systems, we pro-

pose and show that such impacts can lead to the observed diversity in exoplanet bulk structural properties. Furthermore, we derive envelope mass fractions for sub-Neptune exoplanets with reported mass and radius measurements and use these to identify multiplanet systems whose architectures have been shaped by late-stage impacts.

In Part II, we switch our focus to the formation of our own Solar System. We do so in the context of NASA's OSIRIS-REx asteroid sample return mission. OSIRIS-REx, which launches in September 2016 and arrives at the near-Earth asteroid 101955 Bennu in 2019, will orbit and characterize Bennu before ultimately retrieving a sample of the asteroid regolith and returning it to Earth. On the basis of Earth-based visible and infrared spectroscopy, Bennu is believed to be a chondrite, with affinities in different spectral bands to the asteroids 3 Pallas and 24 Themis, perhaps pointing to it being a transitional body between the two.

An X-ray imaging spectrometer called REXIS was designed and built at MIT in order to constrain the surface composition of Bennu while OSIRIS-REx orbits the asteroid via X-ray fluorescence spectroscopy. In Chapter 6, we investigate the ability of REXIS to carry out its science mission of characterizing Bennu and ultimately placing Bennu within an analog meteorite class. We do so by constructing models of Solar X-ray and asteroid fluorescence spectrum to anticipate the data product received by REXIS. We then simulate the data processing and reconstruction to show that REXIS is capable of identifying Bennu as a

carbonaceous chondrite. Identifying the composition of Bennu and any similarities it has to both the meteorite record and observed asteroids can shed a great deal of light on the origin of Bennu and compositional variation amongst asteroids.

In Chapter 7, we summarize the work presented in this thesis. We highlight outstanding questions in planet formation as well as promising routes for further study. We close with conclusions.

Part I

The Formation and Evolution

of Exoplanetary Systems

Chapter 2

The Thermal Evolution of

Gaseous Envelopes

The two most direct observables for exoplanets are their radii R_p and masses M_p. Radii are measured by the transit technique, while masses are measured using either radial velocities or transit timing variations. If we have both the mass and radius of a planet, we immediately possess another important piece of information, the bulk, or mean, density $\bar{\rho}$:

$$\bar{\rho} = \frac{3M_p}{4\pi R_p^3}.$$ (2.1)

A planet's bulk density allows us—sometimes with a few general assumptions—to make several powerful inferences about the planet. In our Solar System, we see a variety of bulk densities, bracketed by those with significant gaseous components ($\bar{\rho} \lesssim 2\,\mathrm{g/cm^3}$) and those with negligible gaseous masses ($\bar{\rho} \gtrsim 3\,\mathrm{g/cm^3}$); see Table 2.1. If we assume a primarily H/He com-

position for the gaseous envelope, then with a model in place for the evolution of the envelope as a function of time, the mean density $\bar{\rho}$ allows us to infer the fraction of the planet's mass that is contained in the envelope. As we will show in Chapters 4 and 5, atmospheric mass can serve as an important proxy for where and how an exoplanet formed.

Table 2.1: Mass and radius data for Solar System planets.

Planet	M_p [$M_\oplus$]	R_p [$R_\oplus$]	$\bar{\rho}$ [g/cm^3]
Mercury	0.055	0.38	5.43
Venus	0.82	0.95	5.24
Earth	1	1	5.51
Mars	0.11	0.53	3.91
Jupiter	317.83	10.97	1.33
Saturn	95.16	9.14	0.69
Uranus	14.54	3.98	1.27
Neptune	17.14	3.87	1.64

In this chapter, we will present the framework that allows us to theoretically track the thermal history of a bound, gaseous envelope. This framework has two important applications in understanding the formation and evolution of a planet. First, it can be applied to the accretion and subsequent thermal evolution of a gravitationally bound envelope in the presence of the gas disc. The envelope mass that can be accreted depends critically on when and where the planet formed in the primordial disc.

Secondly, after the dissipation of the gas disc, an envelope cools and contracts under the action of the core's gravity. The radius, which is a function of time, also depends on both the total mass and envelope mass

of the planet, so that mass and radius measurements can be used to infer the bulk composition of the planet. Since envelope mass is a strong function of formation history, this in turn can be used to infer or place constraints on the formation history of the planet given its current state.

We begin this chapter by reviewing the general equations necessary to track the evolution of a gaseous envelope, including the equations of hydrostatic equilibrium, energy transfer, and thermal evolution. Along the way, we detail specifics regarding the equation of state and the behavior of the gas within the envelope. We close by exploring analytically the equations of evolution by deriving some analytical expressions for the thermal evolution of the gaseous envelope, applicable early in the planet's history. We close by presenting some numerical results from our models, although we hold off the presentation of our context-specific results for Chapters 4 and 5.

We assume throughout that our planets are made up of hydrogen and helium envelopes. We justify this on cosmochemical grounds due to the fact that molecular hydrogen H_2 and atomic helium He likely dominate the vast majority of gaseous mass in any protoplanetary nebula (Lewis 1972). Envelopes that make up a non-negligible fraction of the total planet mass are therefore most likely primarily H/He. Nevertheless, the contribution of heavy molecules and solids to the thermal evolution of gas envelopes is not totally ignorable, and we account for the presence of both in our opacity models, which we discuss in Section 2.4.1.

2.1 General Concepts

A gravitationally bound gaseous envelope is in quasihydrostatic equilibrium with the mass below it. The structure of the envelope depends non-trivially on a host of variables, including the mass of the core, the incident stellar flux (which deposits energy into the envelope), the modes of energy transport within the envelope, and how energy is exchanged between the envelope and the underlying core (Bodenheimer & Pollack 1986). Initially, heated by the energy of formation, the envelope will be extended. With time, the envelope will contract under the influence of the core's gravity. The energy lost due to contraction to a lower entropy state is released to the environment in the form of radiation. In the presence of the gas disc, the cooling of the envelope allows an increasing mass of gas to be accreted. In its initial hot state, the envelope will likely be unstable to convection, so that the envelope structure lies along an isentrope, or adiabat. The outermost portion of the envelope must be in radiative equilibrium with the environment around it (whether that be the protoplanetary disc or the vacuum), so that its temperature is set at either the disc temperature or the equilibrium temperature based on the incident stellar flux. As the envelope cools from the outside in, the location where the outermost radiative portion meets the adiabat—the radiative-convective boundary (RCB)—moves inwards, and the envelope isothermalizes as it releases energy (Henyey, Vardya, & Bodenheimer 1965).

The radiative portion of the envelope regulates the cooling of the en-

tire envelope, since energy transport in the adiabatic region occurs very efficiently via convection. Energy transport in the radiative region, on the other hand, occurs via radiative diffusion, so that the opacity (which dictates how energy is attenuated by radiation-matter interactions) can serve as a bottleneck to cooling.* This cooling process is summarized in Fig. 2-1. In the presence of the gas disc, the cooling and contraction of the RCB lead to more atmospheric mass accretion by the planet, which we discuss in Chapter 4.

2.2 Hydrostatic Equilibrium

While the cooling and thermal evolution of a gaseous envelope are time-dependent processes, the hydrodynamic timescale is much shorter than the timescale over which order unity energy changes in the gas envelope take place: the sound crossing time is R_p/c_s (where R_p is the planet radius and c_s the speed of sound of the gas) and is perhaps on the order of tens of minutes to hours, while the contraction (Kelvin-Helmholtz) time is on the order of millions to billions of years. As a result, the thermal evolution of a gaseous envelope can be considered quasi-statically: for a fixed energy budget in the planet, its envelope structure can be computed assuming hydrostatic equilibrium. By considering a string of

---

*Strictly speaking, depending on the lapse rate within the envelope, there may be more than one radiative-convective boundary. The cooling of the envelope as a whole, however, will depend on the characteristics of the outermost radiative region, and so for simplicity, when speaking of the RCB, we will always refer to the RCB that lies furthest from the core.
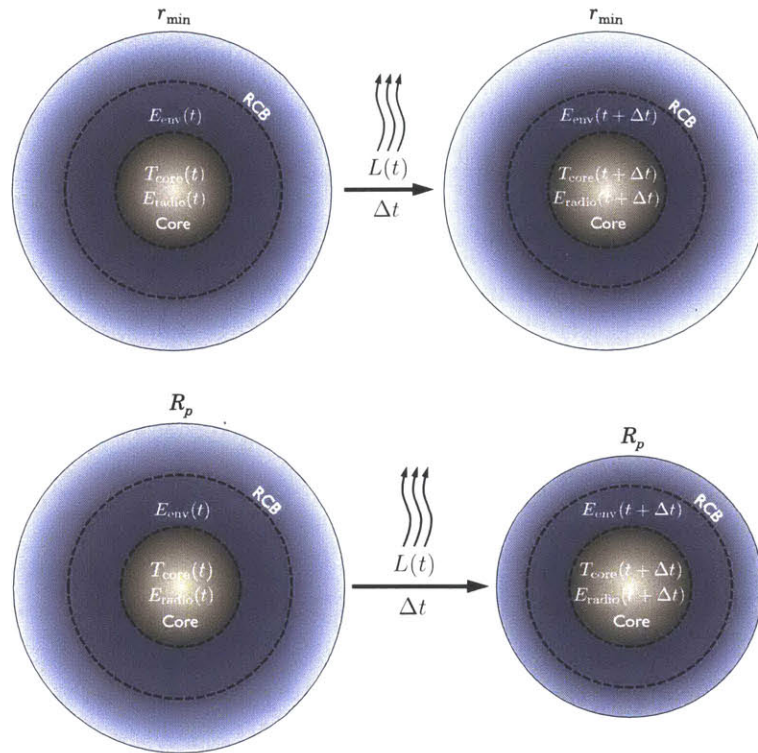
Figure 2-1: Schematic diagram of the cooling and contraction of a gaseous envelope. In the upper panel, thermal evolution in the presence of the gas disc is illustrated, while in the lower panel, we show thermal evolution after the gas disc has dissipated. The passage from the left hydrostatic state at time t to that at time $t + \Delta t$ is associated with a release of luminous energy L. The outermost edge is given either by $r_{\min}$, the radius within which the core can gravitationally bind an envelope in the presence of the gas disc (see Chapter 4), or by R_p, the observable planet radius in the absence of the gas disc (see Chapter 5). The radiative-convective boundary is shown as RCB, as are the sources of energy to the system: the envelope energy E_{env}, the core energy (parameterized by the core temperature T_{core}), and radiogenic heating (E_{radio}). Energy exchange between the core and envelope and general thermal evolution are discussed in depth in Section 2.5.

energy states, we can construct a sequential thermal evolution history for the planet, which we do in Section 2.5.

The equation of hydrostatic equilibrium is (Chandrasekhar 1960)

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -g(r)\rho, \tag{2.2}$$

where P is the gas pressure and the gravity $g(r) = GM(<r)/r^2$. Here, $M(<r)$ is the total mass enclosed within a radius r, including the core mass, ρ is the local gas density, and

$$\frac{\mathrm{d}M}{\mathrm{d}r} = 4\pi r^2 \rho. \tag{2.3}$$

If the atmospheric mass is sufficiently smaller than the core mass M_{core}, we may replace Eqs. (2.2) and (2.3) with

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{GM_{\mathrm{core}}}{r^2}\rho. \tag{2.4}$$

The gas envelopes of super-Earths and mini-Neptunes—which are the focus of our work—comprise at most $\sim 20\%$ of their total mass in their gas envelope, with a typical envelope mass fraction roughly $0.1-10\%$ (see Chapter 5 and especially Fig. 5-13). As a result, the assumption made in Eq. (2.4) that $M_{\mathrm{core}} \gg M_{\mathrm{atm}}$ is a good one. This assumption has additional practical and physical underpinnings. Solving for the structure of a self-gravitating atmosphere using Eq. (2.3) requires an iterative solution. Given other iterative procedures we employ, this requires con-

siderable computational time, which Eq. (2.4) alleviates. Furthermore, in the presence of the full gas disc, when the atmosphere to core mass ratio $M_\text{atm}/M_\text{core} \gtrsim 0.2 - 0.3$, the envelope begins to undergo runaway gas accretion. This phenomenon—also referred to as "core nucleated instability" (Mizuno 1980, Stevenson 1982, Rafikov 2006, Piso & Youdin 2014)—typically results in the formation of gas giants, which make up a small fraction of the exoplanet population and are outside the scope of this work.

Eq. (2.4) above is supplemented with two additional equations, one of state and one of energy conservation.

2.3 Equation of State

The equation of state for a given material relates its state or thermodynamic variables to one another. The simplest nontrivial equation of state is that of an ideal gas:

$$P = \rho \frac{k_\text{B}}{\mu m_\text{p}} T, \qquad (2.5)$$

where T is the temperature of the gas, k_B is the Boltzmann constant, μ is the molecular mass of the gas species in atomic mass units, and m_p is the proton mass. For all of our applications, we assume a metallicity $Z = 0.02$ and hydrogen and helium mass fractions of $X = 0.8$ and $Y = 0.18$, respectively. Ignoring dissociation and ionization, the mean molecular mass is $\mu = 2.3$, where we have assumed that all hydrogen is in the form

of H_2.

For many applications, the ideal gas law is extremely accurate, from stellar interiors to terrestrial planet atmospheres. For the gas accretion phase of close-in planets, the ideal gas law remains a good approximation because disc temperatures are high enough so that temperatures and pressures in the envelope avoid regimes in which the ideal gas law breaks down and metallic hydrogen is obtained. Even if the gas dissociates (see Section 2.4.2), for a fixed pressure and temperature, the density changes (due to change in mean molecular weight) at most by a factor order unity. The gas still remains ideal and the ramifications on the thermal evolution of the envelope of ignoring dissociation are small.

One extreme in which the gas law is not applicable, however, is for relatively low temperatures but high pressures (Stevenson & Salpeter 1977); this extreme will be achieved at the base of an envelope thermally evolving in the absence of the gas disc. In this case, hydrogen reaches a condensed, degenerate state.[†] These conditions are obtained for thermal evolution in the absence of the gas disc, and so are important for us to consider in Chapter 5.

The temperature at the base of the envelope will range from roughly that which joins the rocky core adiabatically with the gas disc to one which is completely isothermal with the equilibrium temperature. For close-in planets, the latter is on the order of $\sim 500 - 1000$ K, while the former can be as high as $\sim 10^4$ K. Calculations of the phase struc-

---

[†]Degenerate in the sense that the equation of state becomes weakly dependent or independent of temperature.

ture of hydrogen indicate that for pressures $P \gtrsim 10^{-2} - 10^{-1}$ Mbar ($=$ $10^{10} - 10^{11}$ dyne/cm^2 $= 10^9 - 10^{10}$ Pa) and densities $\rho \gtrsim 10^{-2} - 10^{-1}$ g/cm^3, hydrogen is a metal, either as a solid or liquid (Stevenson & Salpeter 1977). At the lower end of this temperature range, and for basal pressures typical of sub-Neptune gas envelopes, metallic hydrogen is obtained. Indeed, a simple order of magnitude estimate from the hydrostatic equation for the pressure in the gas envelope at its base gives

$$P_{\mathrm{base}} \sim \frac{GM_{\mathrm{core}}M_{\mathrm{atm}}}{4\pi R_{\mathrm{core}}^4}$$
$$= 1.15 \times 10^{11} \left(\frac{M_{\mathrm{core}}}{M_{\oplus}}\right) \left(\frac{M_{\mathrm{atm}}}{M_{\mathrm{core}}}\right) \mathrm{Pa}. \tag{2.6}$$

As such, this region of phase space is unpopulated in popular, tabulated equations of state used for planetary or stellar evolution models [e.g., OPAL (Rogers, Swenson, & Iglesias 1996) and Saumon, Chabrier, & van Horn (1995)]. The equation of state data of Saumon, Chabrier, & van Horn (1995) that we show in the upper panel of Fig. 2-2 only extend to the non-metallic portion of phase space. The low-temperature, high-pressure portion in which the Saumon data are not compiled is indicated with a dashed line in the upper panel of Fig. 2-2.

Nevertheless, thermochemical calculations indicate that metallic hydrogen is in fact compressible, although with a pressure dependence that is weaker than is found by application of the ideal gas law or by simple power-law extrapolation of the table data (Nettelmann, Holst, Kietzmann et al. 2008). Our approach therefore is to interpolate between

the tabulated equation of state data of Saumon, Chabrier, & van Horn (1995) within the tabulated regions, and where the tabulated data do not exist (i.e., where pressures and temperatures are in a metallic regime) to assume the equation of state is independent of the temperature and to interpolate between the endpoints of the data in pressure-density space. These endpoints are shown in the lower panel of Fig. 2-2; as is evident from the tabulated data, the isotherms converge, indicating the weak temperature dependence of the equation of state.

2.3.1 Radiation Pressure

Strictly speaking, the pressure P in Eq. (2.4) must include not only the mechanical pressure of a gas p_{gas} [reflected in Eq. (2.5)], but also the effect of radiation pressure p_{rad}, so that $P = p_{\mathrm{gas}} + p_{\mathrm{rad}}$. Radiation pressure, which may be non-negligible in stellar and giant planet interiors, is negligible here: the ratio of the two is given by (Chandrasekhar 1960)

$$\frac{p_{\mathrm{rad}}}{p_{\mathrm{gas}}} = \frac{\dfrac{4\sigma_{\mathrm{B}}}{3c}T^4}{\rho\dfrac{k_{\mathrm{B}}}{\mu m_{\mathrm{p}}}T} = \left(\frac{4\mu m_{\mathrm{p}}\sigma_{\mathrm{B}}}{3k_{\mathrm{B}}c}\right)\frac{T^3}{\rho}, \qquad (2.7)$$

where c is the speed of light. In Fig. 2-3 we show a plot of Eq. (2.7) as a function of temperature T, assuming the gas state lies along an adiabat characteristic of an envelope embedded in the gas disc (and therefore representing the highest temperature and lower density case for the interior of the envelope). The radiation pressure is typically three or more
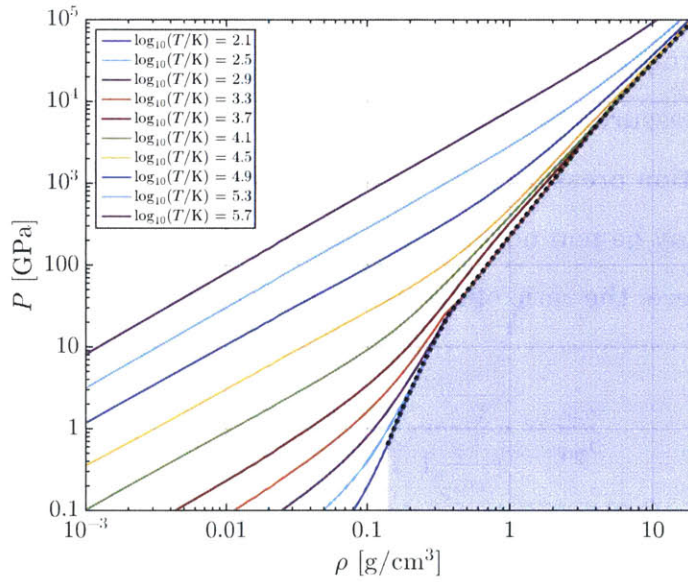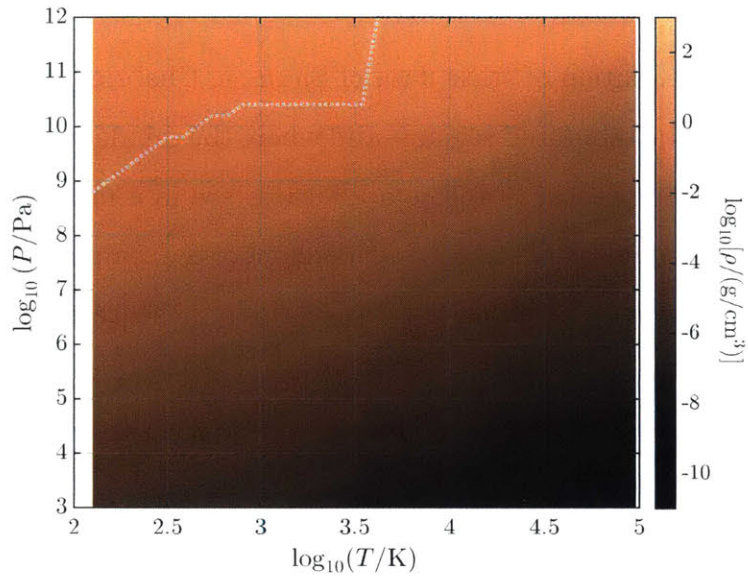
Figure 2-2: Tabulated equation of state data from Saumon, Chabrier, & van Horn (1995). In the top panel, we show a contour map. In the bottom panel, we show the contours projected as isotherms. The isotherms converge at the metallic hydrogen phase transition.

orders of magnitude smaller than the gas pressure, and hence negligible, so that we take $P = p_{\text{gas}}$.
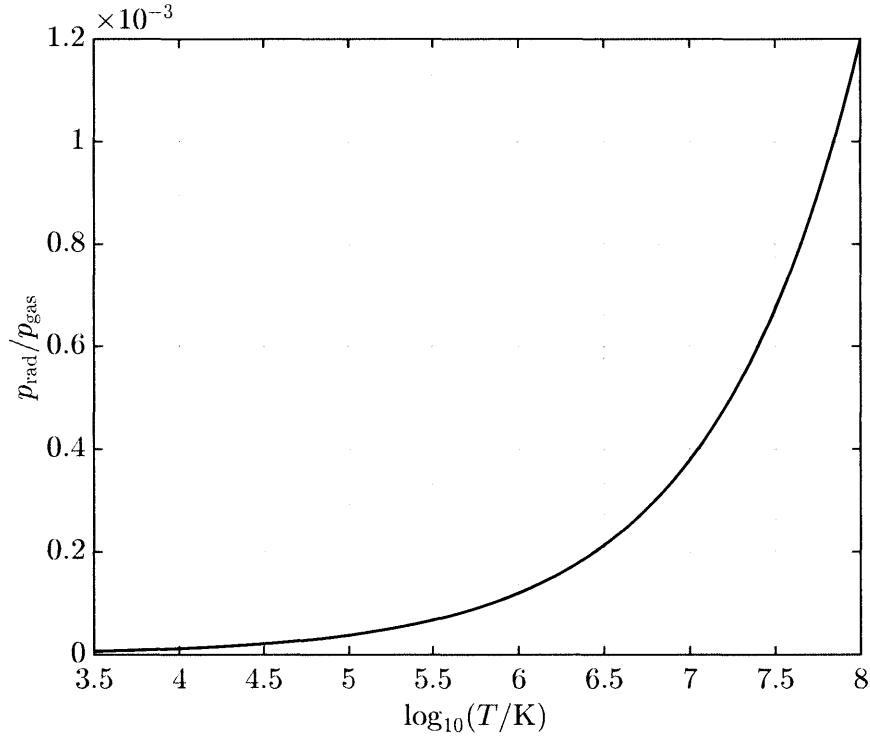
Figure 2-3: Ratio of radiation pressure p_{rad} to gas mechanical pressure p_{gas} using Eq. (2.7) and assuming the envelope state is isentropic, joining the temperature and the pressure at the envelope base with the gas disc adiabatically.

2.4 Energy Transport in Gaseous Envelopes

In addition to the equations of hydrostatic equilibrium [Eq. (2.4)] and state [Eq. (2.5)], we require an additional equation to describe energy transport in the envelope, which can occur in two ways. In the first,

73

energy is transported via radiative diffusion. Here (Eddington 1926),

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{4\sigma_\mathrm{B}}{3c}T^4\right) = -\frac{\kappa_\mathrm{R}L\rho}{4\pi cr^2},$$ (2.8)

where κ_R is the Rosseland mean opacity of the gas at a given radius (see Section 2.4.1 below) and L is the luminous energy released by a mass shell of the envelope at a given r. Differentiating the left-hand side, we get (Chandrasekhar 1960)

$$\frac{\mathrm{d}T}{\mathrm{d}r} = -\frac{3\kappa_\mathrm{R}\rho}{16\sigma_\mathrm{B}T^3}\frac{L}{4\pi r^2},$$ (2.9)

which allows us, with Eq. (2.4) and (2.5) to specify the state of the gas at any r. In certain circumstances, however, the temperature gradient $\mathrm{d}T/\mathrm{d}r$ may be unstable locally to convection. Let γ_a be the ratio of specific heats of the gas

$$\gamma_a \equiv c_p/c_v,$$ (2.10)

where c_p the heat capacity at constant pressure and c_v is the heat capacity at constant volume. Defining the adiabatic gradient $\nabla_\mathrm{ad} \equiv (\gamma_a - 1)/\gamma_a$, the Schwarzschild criterion states that convective instability sets in when (Kippenhahn, Weigert, & Weiss 2012)

$$\frac{\mathrm{d}\log T}{\mathrm{d}\log P} \geq \nabla_\mathrm{ad}.$$ (2.11)

74

If at any r, $\mathrm{d}\log T/\,\mathrm{d}\log P$ is greater than ∇_{ad}, i.e.,

$$\frac{\mathrm{d}\log T}{\mathrm{d}\log P} = \frac{-3\kappa_{\mathrm{R}}PL}{64\pi G\sigma_{\mathrm{B}}M_{\mathrm{core}}T^4} > \nabla_{\mathrm{ad}}, \qquad (2.12)$$

convection takes over as the dominant method of energy transport. Typically, in a gaseous configuration unstable to convection, the convective motion efficiently reduces a superadiabatic gradient $\mathrm{d}\log T/\,\mathrm{d}\log P > \nabla_{\mathrm{ad}}$ to an adiabatic one in which $\mathrm{d}\log T/\,\mathrm{d}\log P \equiv \nabla_{\mathrm{ad}}$.[‡] Thus, when Eq. (2.11) is satisfied, we set $\mathrm{d}\log T/\,\mathrm{d}\log P = \nabla_{\mathrm{ad}}$ locally. The radius at which the transition between the two energy transport regimes (radiative and convective) occurs is termed the radiative-convective boundary (RCB).

2.4.1 Opacities

The opacity that enters into our equation of radiative energy transport is the Rosseland mean opacity κ_{R}; it is an average of the frequency-dependent opacity κ_ν over all frequencies (Rybicki & Lightman 1986):

$$\frac{1}{\kappa_{\mathrm{R}}} \equiv \frac{\displaystyle\int_0^\infty \frac{1}{\kappa_\nu}\frac{\nu^3 e^{2h_{\mathrm{P}}\nu/k_{\mathrm{B}}T}}{\left(e^{h_{\mathrm{P}}\nu/k_{\mathrm{B}}T}-1\right)^3}\frac{h_{\mathrm{P}}\nu}{k_{\mathrm{B}}T^2}\,\mathrm{d}\nu}{\displaystyle\int_0^\infty \frac{\nu^3 e^{2h_{\mathrm{P}}\nu/k_{\mathrm{B}}T}}{\left(e^{h_{\mathrm{P}}\nu/k_{\mathrm{B}}T}-1\right)^3}\frac{h_{\mathrm{P}}\nu}{k_{\mathrm{B}}T^2}\,\mathrm{d}\nu}, \qquad (2.13)$$

[‡]See, e.g., Cowling (1934) for a stability analysis derivation of this fact. An alternative treatment accounting for superadiabatic heat transport employs mixing-length theory, which examines the heat transport by eddies in a convective medium. Again, the deviations from adiabaticity are negligible (Chandrasekhar 1960, Kippenhahn, Weigert, & Weiss 2012).

where h_P is Planck's constant and ν the frequency of radiation. The mean opacity depends on the temperature, pressure, and composition of the envelope gas. For instance, temperature and pressure can affect spectral absorption features and hence enhance or diminish opacity. For a fixed X and Y and elemental abundances, using photoattenuation data, it is possible to evaluate Eq. (2.13). This has been done and produced in tabular form in the OPAL dataset (Iglesias & Rogers 1996), which we use. The Rosseland mean opacity for a H/He mixture of cosmic abundances as a function of density and temperature is shown in Fig. 2-4.

While the opacities shown in Fig. 2-4 are for a primarily H/He gas mixture, the affect of molecular absorption is accounted for. Molecular abundances for species other than H_2 and He in the mixture are determined assuming chemical equilibrium and cosmic elemental abundances so that opacities resulting from hydrogen, helium, and molecular transitions are all reflected in the data. Nevertheless, the temperatures and pressures typically reached in the outmost regions of the envelope are not tabulated in the OPAL data. In that case, we take a simplified approach, and use power-law fits for κ_R derived from calculations similar to those used to derive the OPAL data. The power-law expressions are of the form $\kappa_R \propto T^\alpha P^\beta$ for power-law indices α and β (Zhu, Hartmann, & Gammie 2009). We account for contributions from both molecular scattering and dust grains. In particular, at temperatures $\gtrsim 1800$ K, when dust grains sublimate, we expect molecular scattering to be the domi-
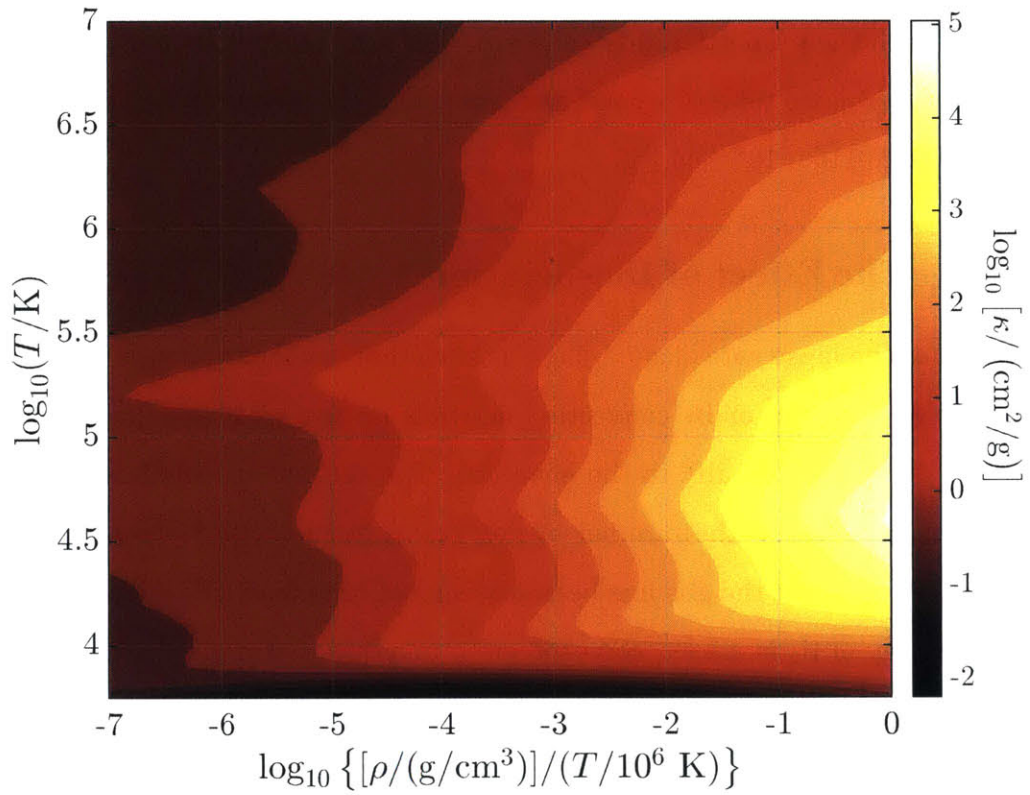
Figure 2-4: Rosseland mean opacity κ_R for a H/He mixture of $X = 0.8$ and $Y = 0.18$. The data are from the OPAL dataset (Iglesias & Rogers 1996).

nant source of opacity (Movshovitz, Bodenheimer, Podolak et al. 2010). At temperatures lower than this, both molecular scattering and grain opacity are accounted for. We assume in the latter case that opacities can be summed by mass, so that $\kappa_R \rho = \left(\kappa_{R,\text{dust}} Z + \kappa_{R,\text{mol}}\right) \rho$, where Z is the assumed metallicity. For the dust opacity $\kappa_{R,\text{dust}}$, we take $\alpha = 0.738$ and $\beta = 0$, with the proportionality constant $0.0528 \text{ cm}^2/\text{g}$. For the molecular opacity $\kappa_{R,\text{mol}}$, we take $\alpha = 2.905$ and $\beta = 0.498$, with the proportionality constant $1.0116 \times 10^{-14} \text{ cm}^2/\text{g}$.

2.4.2 The Effect of Dissociation

Whether the lapse rate $\mathrm{d}\log T / \mathrm{d}\log P$ is such that energy transport is via radiative diffusion or by convection depends on the adiabatic gradient $\nabla_{\text{ad}} = (\gamma_a - 1)/\gamma_a$ locally in the envelope. The ratio of specific heats γ_a in turn depends on the number of degrees of freedom in the gas, and therefore whether the gas has been dissociated or ionized. The dissociation energy of H_2 is 4.5 eV, corresponding to a dissociation temperature $T_{\text{diss}} \sim 4.5 \text{ eV}/k_{\text{B}} = 5.2 \times 10^4$ K, which is obtainable in super-Earth or mini-Neptune atmospheres. For an ideal gas possessing N degrees of freedom, $\gamma_a = (N + 2)/N$. Hence, dissociation of a diatomic gas (for which $N = 5$) into a monatomic gas ($N = 3$) changes γ_a from 7/5 to 5/3. In Fig. 2-5, we show numerically tabulated data for ∇_{ad} from Saumon, Chabrier, & van Horn (1995). The "troughs" at roughly $T \sim 10^{3.5}$ K and $T \sim 10^4$ K are explained analytically below.
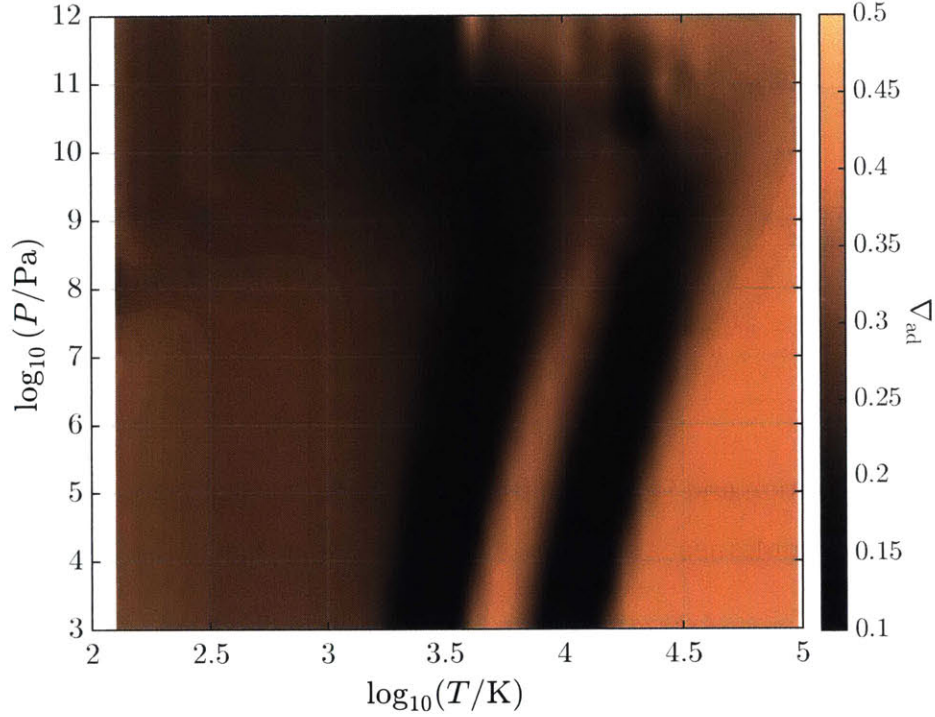
Figure 2-5: Adiabatic gradient from tabulated equation of state data (Saumon, Chabrier, & van Horn 1995). The troughs characterizing the dissociation and ionization fronts are explained from the calculations based on the Saha equations (Fig. 2-7).

For the following discussion, we assume for simplicity that we have a gas that is initially H_2. For hydrogen dissociation, $H_2 \xleftrightarrow{\chi_D} 2H$, where $\chi_D = 4.476$ eV is the dissociation energy of H_2. For hydrogen ionization, $H \xleftrightarrow{\chi_I} p^+ + e^-$, where the ionization energy $\chi_I = 13.6$ eV. If x represents the ionization fraction and y the dissociation fraction, χ_D the dissociation energy, and χ_I the ionization energy, then the adiabatic gradient $\nabla_{ad} \equiv$

$(\gamma_s - 1)/\gamma_s = Z_1/Z_2$, where Z_1 and Z_2 are given by (Vardya 1960)

$$Z_1 \equiv \left(\frac{1+y}{2} + x\right)\left\{1 + \left(\frac{1-y}{2}\right)\left(y + \frac{x}{2}\right)\frac{\chi_D}{k_B T} + x\left(\frac{3}{2} - y - \frac{x}{2}\right)\frac{\chi_I}{k_B T}\right\}$$

(2.14a)

$$Z_2 \equiv \frac{7}{2}\left(\frac{1-y}{2}\right) + \frac{5}{2}(x+y) + \frac{1-y}{8}\left[2y(1+y) - x(1-y)\right]\left(\frac{\chi_D}{k_B T}\right)^2 + ...$$
$$+ \frac{x}{2}\left[(1 - x^2) + 2x(1-y)\right]\left(\frac{\chi_I}{k_B T}\right)^2 + \frac{x}{2}(1-y)(1+x+2y)\left(\frac{\chi_I}{k_B T}\right)\left(\frac{\chi_D}{k_B T}\right).$$

(2.14b)

The dissociation and ionization fractions y and x, respectively, are determined by simultaneous solution of two Saha equations (Saha 1921):

$$\frac{[2(x-y)]^2}{1-x} = \frac{1}{n_0}\left[\frac{2\pi(m_p/2)}{h_P^2 k_B T}\right]^{3/2}\exp\left(-\frac{\chi_D}{k_B T}\right) \qquad (2.15)$$

$$\frac{(2y)^2}{2(x-y)} = \frac{1}{n_0}\left(\frac{2\pi m_e}{h_P^2 k_B T}\right)^{3/2}\exp\left(-\frac{\chi_I}{k_B T}\right), \qquad (2.16)$$

where n_0 is the number density of H_2, m_p is the mass of a proton, k_B is the Boltzmann constant, and h_P is Planck's constant. The number densities of the potential constituents are $[n_{H_2}] = (1 - x)n_0$, $[n_H] = 2(x - u)n_0$, $[n_{p^+}] = 2yn_0$, and $[n_{e^-}] = 2yn_0$.

In the extremes, solution of Eq. (2.15) would give $x, y \approx 0$ for $k_B T \lesssim \chi_D$, $y \approx 1$ and $x \approx 0$ for $\chi_D \lesssim k_B T \lesssim \chi_I$, and $y \approx 0$ and $x \approx 1$ for $\chi_I \lesssim k_B T$. The temperature at which ionization occurs $\sim \chi_I/k_B \approx 1.6 \times 10^5$ K. If temperatures are sufficiently low that we may ignore ionization (which

80

is usually the case), then the expression for ∇_{ad} reduces to

$$\nabla_{\mathrm{ad}} = \frac{(1+y)\left[1 + \frac{y}{2}(1-y)\frac{\chi_D}{k_{\mathrm{B}}T}\right]}{\frac{7}{2}(1-y) + 5y + \frac{y}{2}(1-y^2)\left(\frac{\chi_D}{k_{\mathrm{B}}T}\right)^2}, \tag{2.17}$$

so that $\gamma_s = (1 - \nabla_{\mathrm{ad}})^{-1}$ is $5/3$ when $y = 1$ (complete dissociation and hence a monatomic gas) and $7/5$ when $y = 0$ (full molecular hydrogen and hence diatomic). In Fig. 2-6, we show in the upper panel the number fraction of an $H_2 - H - p^+$ mixture as a function of temperature for several pressures. In the lower panel, we show the mean molecular weight of the same mixture as a function of temperature for a variety of temperatures. In Fig. 2-7, we show the adiabatic index γ_a (upper panel) and the adiabatic gradient ∇_{ad} (lower panel) as a function of temperature for a variety of pressures. The troughs at the dissociation and ionization fronts evident in Fig. 2-5 are clear in Fig. 2-7. In the partially dissociated regime, the increase in degrees of freedom of the system leads to a decrease in $\gamma_a \approx 1.1$.

2.4.3 The Adiabatic Envelope

In an adiabatic envelope, as a consequence of the relation $\mathrm{d}\log T / \mathrm{d}\log P = \nabla_{\mathrm{ad}}$ and the ideal gas law (Chandrasekhar 1960),

$$P \propto \rho^{\gamma_a} \propto T^{\gamma_a/(\gamma_a-1)}. \tag{2.18}$$

81
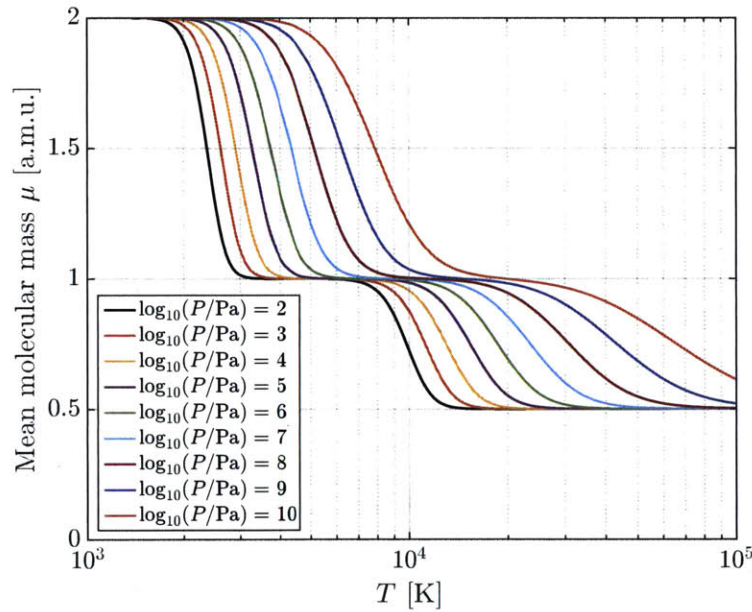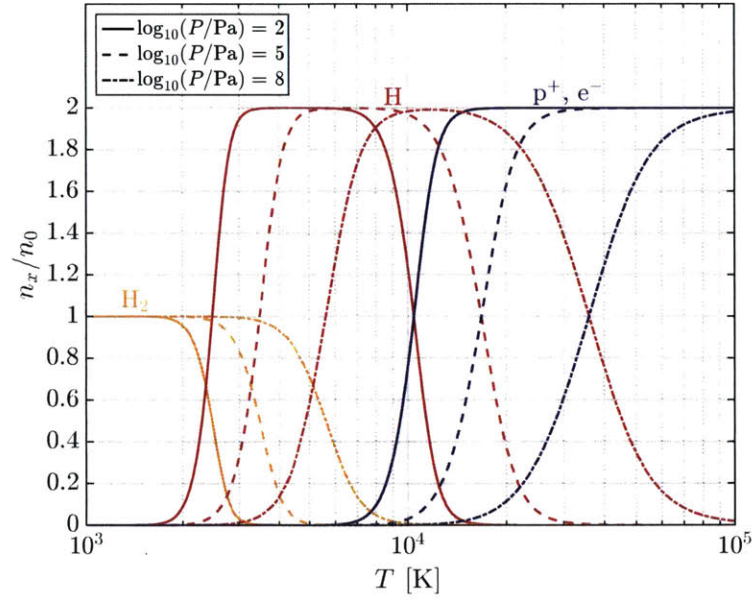
Figure 2-6: **Effect of dissociation and ionization on mean molecular mass.** In the upper panel, we show the number fraction n_x/n_0 of a $H_2 - H - p^+$ mixture as a function of temperature for a variety of pressures. The number fraction of the xth species is normalized to that of H_2 (n_0) at low temperatures. In the lower panel, we show the effect of dissociation and ionization on the mean molecular mass of the same mixture.
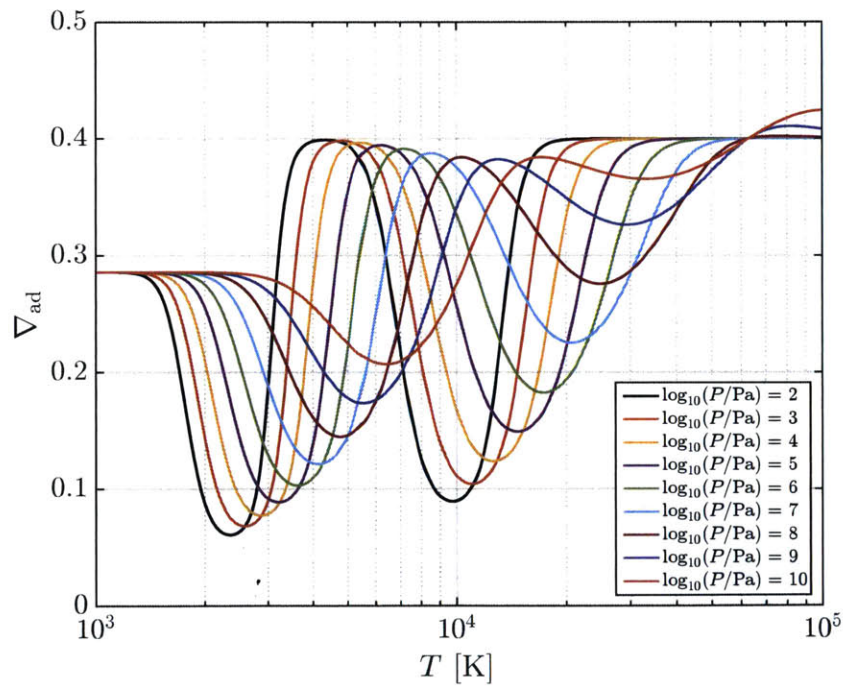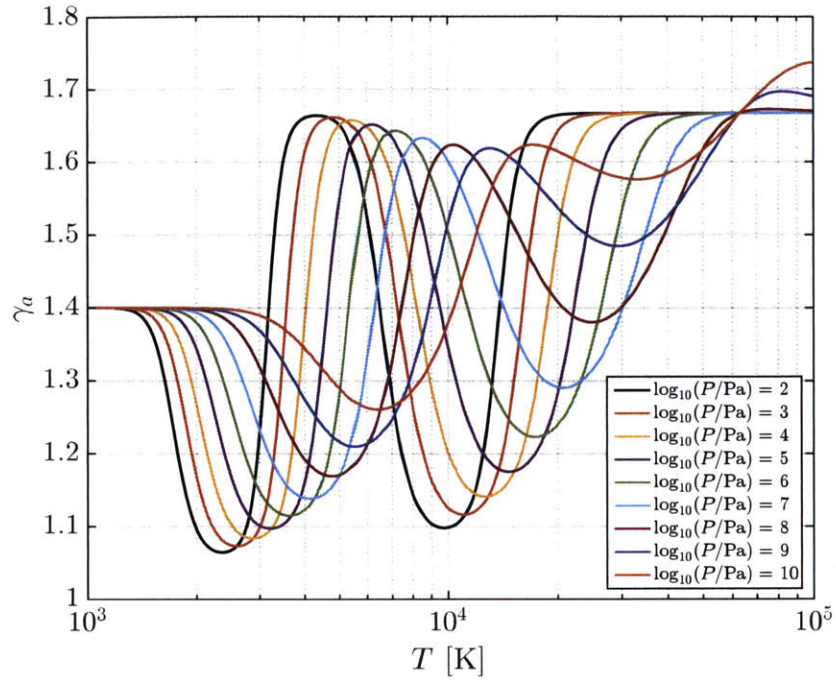
Figure 2-7: Analytically calculated adiabatic index and adiabatic gradient. The troughs evident in Fig. 2-5 are clear in the lower panel.

When this is the case, combined with hydrostatic equilibrium, we can solve analytically for the density, pressure, and temperature of the envelope if γ_a is constant. Suppose we have a boundary $r = R^*$ with corresponding density, pressure, and temperature ρ^*, P^*, and T^*, respectively. Then if r is measured inwards from R^*,

$$\frac{\rho}{\rho^*} = \left[\nabla_{\mathrm{ad}} \lambda \left(\frac{R^*}{r} - 1 \right) + 1 \right]^{1/(\gamma_a - 1)} \tag{2.19a}$$

$$\frac{P}{P^*} = \left[\nabla_{\mathrm{ad}} \lambda \left(\frac{R^*}{r} - 1 \right) + 1 \right]^{\gamma_a/(\gamma_a - 1)} \tag{2.19b}$$

$$\frac{T}{T^*} = \nabla_{\mathrm{ad}} \lambda \left(\frac{R^*}{r} - 1 \right) + 1. \tag{2.19c}$$

Here λ is the escape parameter at $r = R^*$, given by the ratio of the gas gravitational potential energy to its thermal energy:

$$\lambda = \frac{G M_{\mathrm{core}} \mu m_{\mathrm{p}}}{R^* k_{\mathrm{B}} T^*}. \tag{2.20}$$

Eqs. (2.19) will be useful since we will find that most of the mass of gaseous envelopes around super-Earths and mini-Neptunes are contained interior to the RCB. Furthermore, we will find that accretion in the presence of the full gas disc leads to envelopes that are almost entirely adiabatic, so that the simple expressions above lead to results indistinguishable from those achieved by numerical integration (see Chapter 4).

2.5 Planetary Thermal Evolution

The quantity $L(r)$ is the luminous intensity transported through the envelope at a given radial position r. The origin of this energy is several-fold. First, entropy changes in the envelope will lead to energy release from it. The most obvious case of this is the energy released due to contraction of the envelope (leading to an overall decrease in E_{tot}); this is referred to as the *intrinsic luminosity*. Sources of energy within the envelope itself can also contribute to L. In this case, for instance, planetesimal accretion can release energy into the envelope. Energy release due to thermochemical changes—chemical reactions—can also contribute to L. Lastly, the core underlying the envelope is a source of luminous energy which can be exchanged with the envelope. Sources of luminous energy other than intrinsic luminosity generally delay the cooling and contracting of the envelope, as we discuss below.

Consider a position in the envelope r. Denote the rate of energy generation per unit mass at that radial position by $\dot{\epsilon}$ and the envelope specific entropy by S. Then energy conservation gives (Kippenhahn, Weigert, & Weiss 2012)

$$\frac{\partial L}{\partial r} = 4\pi r^2 \rho \left(\dot{\epsilon} - T \frac{\mathrm{d}S}{\mathrm{d}t} \right). \tag{2.21}$$

As mentioned, the envelope structure typically (though not always) contains an outer region in radiative equilibrium with the environment and an inner, convective region; the transition between the two regions is

termed the radiative-convective boundary (RCB). Occasionally, there will be several radiative-convective boundaries in an envelope; for the discussion that follows, RCB refers to the outermost (measured from the planet interior) radiative-convective boundary. Integrating Eq. (2.21) from the core radius R_{core} to RCB, we get

$$\int_{R_{\text{core}}}^{\text{RCB}} \frac{\partial L}{\partial r}\, dr = \int_{R_{\text{core}}}^{\text{RCB}} 4\pi r^2 \rho \left(\dot{\epsilon} - T\frac{dS}{dt} \right) dr,$$
$$L\,(\text{RCB}) - L\,(R_{\text{core}}) = L^* - \int_{R_{\text{core}}}^{\text{RCB}} 4\pi r^2 \rho T \frac{dS}{dt}\, dr. \qquad (2.22)$$

In the second line, we have rewritten $L^* \equiv \int 4\pi r^2 \dot{\epsilon}\, dr$. We note that the envelope energy and entropy change are connected by

$$\frac{dE_{\text{env}}}{dt} = \int_{R_{\text{core}}}^{\text{RCB}} 4\pi r^2 \rho T \frac{dS}{dt}\, dr. \qquad (2.23)$$

The term $L\,(\text{RCB})$ in Eq. (2.22) is the luminosity which subsequently passes from the convective region into the radiative region and which is emitted to space. When $L\,(\text{RCB})$ is nonzero, then the apparent or effective blackbody temperature T_{eff} of the planet is different from its equilibrium temperature T_{eq}, the latter of which is the temperature the planet would be if it were in complete thermal equilibrium with its surroundings. The effective temperature is (de Pater & Lissauer 2015)

$$T_{\text{eff}} = \left[\frac{L\,(\text{RCB})}{4\sigma_{\text{SB}}\pi R_p^2} + T_{\text{eq}}^4 \right]^{1/4}. \qquad (2.24)$$

This additional source of energy due to the thermal evolution of the en-

velope beyond that which might be expected naively from thermal equilibrium accounts for the fact that the gas giants and Neptune in our Solar System have $T_{\text{eff}} > T_{\text{eq}}$ (Pollack, Rages, Baines et al. 1986).

The term $L\left(R_{\text{core}}\right)$ determines the energy exchange of the envelope with the core: if F_{core} is the energy flux at the core-envelope boundary, then $L\left(R_{\text{core}}\right) = 4\pi R_{\text{core}}^2 F_{\text{core}}$. When this term is nonzero, it slows the cooling of the envelope. The thermal state of the rocky core can be understood to first order by considering its temperature at formation. The energy of formation for an Earth mass or larger core can lead to extremely large formation temperatures T_{form} in the core. Equating the energy of formation $\sim GM_{\text{core}}^2/R_{\text{core}}$ with $M_{\text{core}}c_{p,\text{core}}T_{\text{form}}$ (where $c_{p,\text{core}}$ is the heat capacity of the core) yields

$$T_{\text{form}} \simeq 6.25 \times 10^4 \text{ K} \left(\frac{M_{\text{core}}}{M_{\oplus}}\right)^{3/4} \left(\frac{c_{p,\text{core}}}{10^3 \text{ J/kg/K}}\right)^{-1}, \qquad (2.25)$$

well above the melting temperature of rock. In this expression, use has been made of the fact that for rocky planets, $R_{\text{core}} \propto M_{\text{core}}^{1/4}$ (Seager, Kuchner, Hier-Majumder et al. 2007). Even accounting for the latent heat of fusion for unmelted rock would result in elevated core temperatures above the melting point, leading to very efficient energy transport throughout the core.

Throughout this thesis, we assume that the heat capacity $c_{p,\text{core}} = 10^7$ erg/g/K $= 10^3$ J/kg/K (Alfè, Gillan, & Price 2002). This is consistent with values used in the literature (e.g., Lopez & Fortney 2014,

Howe & Burrows 2015) and can be recovered to within a factor order unity by dividing the universal gas constant by the molar mass of a typical core constituent molecule (Fermi 1956). In this case, $L\left(R_{\mathrm{core}}\right) = c_{p,\mathrm{core}}M_{\mathrm{core}}\, dT_{\mathrm{core}}/\, dt$ so that Eq. (2.22) becomes

$$L\left(\mathrm{RCB}\right) = c_{p,\mathrm{core}}M_{\mathrm{core}}\frac{dT_{\mathrm{core}}}{dt} + L^* - \int_{R_{\mathrm{core}}}^{\mathrm{RCB}} 4\pi r^2 \rho T\frac{dS}{dt}\, dr. \qquad (2.26)$$

A luminosity due to planetesimal accretion can be accounted for by setting a term $L_{\mathrm{acc}} = L^*$ on the right-hand side of Eq. (2.26), where (Mizuno 1980)

$$L_{\mathrm{acc}} = \frac{GM_{\mathrm{core}}\dot{M}}{R_{\mathrm{core}}} \qquad (2.27)$$

and $\dot{M}$ the rate of accretion. In Chapter 4 we take $L^* = L_{\mathrm{acc}}$, although most of the results we present in the following chapters assume $L^* = 0$. Lastly, in certain cases, radiogenic heating of the envelope due to radioactive isotopes in the core may be relevant. In this case, we add a term $\dot{E}_{\mathrm{radio}}$ to the right-hand side of Eq. (2.26). The decay properties of radioisotopes relevant to the long-term thermal evolution of an Earth-mass or larger planet are summarized in Table 2.2 (Anders & Grevesse 1989, Lopez, Fortney, & Miller 2012).

Table 2.2: Radioisotope decay data.

Species	Abundance n_i [# per 10^6 Si atom]	Half-life λ_i [Gyr]	Energy released ε_i per decay [10^{-13} J]
^{235}U	0.704	0.0057	7.473
^{40}K	1.27	5.48	2.126
^{238}U	4.47	0.0181	6.825
^{232}Th	14.1	0.042	6.526

The total luminosity released by radiogenic heating is then

$$\dot{E}_{\mathrm{radio}}(t) = \frac{\log 2}{10^6} \frac{f_{\mathrm{Si}} M_{\mathrm{core}}}{\mu_{\mathrm{Si}} m_{\mathrm{p}}} \sum_i \frac{\varepsilon_i}{\lambda_i} \exp\left(-\frac{t \log 2}{\lambda_i}\right), \tag{2.28}$$

where λ_i is the half-life of the ith species, ε_i the energy released by the ith species per each decay, f_{Si} the fraction of the core mass in silicon, and μ_{Si} the molecular mass of the silicon atom in atomic mass units. The sum is extended over the long-lived species tabulated in Table 2.2. Ultimately, however, we find that the impact of radiogenic heating is typically negligible (see also Lopez & Fortney 2014).

2.6 Entropy of the Gaseous Envelope

Eq. (2.26) shows that the gas envelope evolves to move to a lower entropy state. More precisely, this entropy change is due to irreversible energy changes in the system due to cooling and contracting. In typical usage, it is meaningful to discuss only the difference in entropy between two states of gas. The specific entropy of a gas is typically known only up to

a constant (Fermi 1956):

$$S = c_v \log T + R \log V + \text{const.}, \qquad (2.29)$$

where R is the universal gas constant and V the volume of the gas configuration. By virtue of Nernst's postulate, which stipulates that the entropy of a system goes to zero as $T \to 0$ (Fermi 1956), we might expect to recover this constant. However, the entropy term diverges as $T \to 0$ if c_v is constant. From statistical mechanics, however, the temperature dependence of c_v can be calculated, and the entropy constant can therefore be recovered. We quote only the results here (which are generally known as Sackur-Tetrode formulae) for a monatomic and diatomic ideal gas (Fermi 1956):

$$S = N k_{\mathrm{B}} \times \begin{cases} \dfrac{5}{2} + \log \dfrac{V \left(2\pi\mu m_{\mathrm{p}} k_{\mathrm{B}} T\right)^{3/2}}{N h_{\mathrm{P}}^3} & \text{(monatomic gas)} \\[2ex] \dfrac{7}{2} + \log \dfrac{V \left(2\pi\mu m_{\mathrm{p}} k_{\mathrm{B}} T\right)^{5/2} 4\pi I_r}{N h_{\mathrm{P}}^5 \mu m_{\mathrm{p}}} & \text{(diatomic gas).} \end{cases}$$

$$(2.30)$$

Here, N is the number of atoms or molecules in the collection, V is the volume of the collection, I_r the rotational moment of inertia for a diatomic molecule, and h_{P} is Planck's constant. An adiabat coincides with an isentrope, which can be shown as follows: substituting $\rho = N/V$ and using the ideal gas law $\rho \propto P/T$ gives $\exp(S) \propto T^{5/2}/P$ in the monatomic case and $\propto T^{7/2}/P$ in the diatomic case. For an ideal gas, the adiabat

is defined as $\mathrm{d}\log T/\mathrm{d}\log P = \nabla_{\mathrm{ad}}$ where $\nabla_{\mathrm{ad}} = 2/5$ for a monatomic gas and $2/7$ a diatomic gas. Hence, in a monatomic gas, $P \propto T^{5/2}$, while in a diatomic gas, $P \propto T^{7/2}$; $\exp(S)$ is therefore constant along an adiabat. The entropy as a function of state has also been tabulated by Saumon, Chabrier, & van Horn (1995); the results are shown in Fig. 2-8. Lines of constant entropy are shown; it is clear that outside of the regions defined by the ionization and dissociation fronts, the lines of constant S follow the power law behavior given by Eqs. (2.30).

The fact that an adiabatic envelope follows an isentrope has interesting implications for the evolution of a gaseous envelope in the presence of a gas disc. If an envelope accreted adiabatically from the nebula and remained in an adiabatic state throughout the lifetime of the gas disc, as the gas disc dissipated, the accreted envelope could expand adiabatically and be lost to the gas disc. If the entropy in the envelope lowers as it cools, however, work must be done in order to expand it out. The specifics underlying transition phase are discussed in detail in Ginzburg, Schlichting, & Sari (2015).
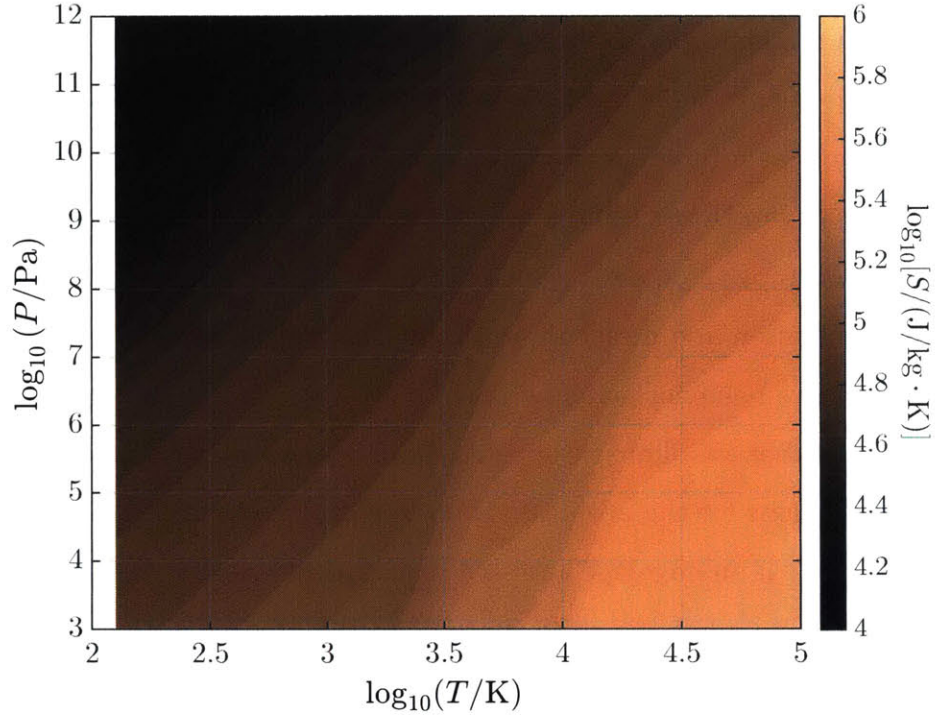
Figure 2-8: Tabulated entropy data for a H/He gas (Saumon, Chabrier, & van Horn 1995).

2.7 Example Thermal Evolution Results

In this section, we briefly present some results from our thermal evolution models. Again, we leave the context-specific results to Chapters 4 and 5. In all cases, we solve for the thermal evolution [Eq. (2.26)] using a 4$^{\text{th}}$ Runge-Kutta order method assuming $L^* = 0$. We generate a grid of intrinsic luminosities L, and for each L, solve iteratively for the radius assuming a fixed envelope mass M_{atm}. The radius space is discretized

into 500 points. In these calculations, we assume an initial hot start, with the initial intrinsic luminosity $L = 10^{25}$ W. As our baseline case, we show in Fig. 2-9 density and pressure profiles and a temperature-pressure diagram for a core mass of $4M_\oplus$ and an envelope mass fraction of $M_{\mathrm{atm}}/M_{\mathrm{core}} = 5\%$. We assume the planet is around a Sun-like star at 0.1 AU, so that the incident stellar flux $F_p = 100F_\oplus$, or equivalently, that the equilibrium temperature $T_{\mathrm{eq}} = 880$ K.

We see that the initial radius is much larger than the core radius, after which the envelope rapidly cools and contracts; this fact underlies the hot start model which otherwise hides our ignorance of the initial state of the thermal evolution. As cooling occurs, the radiative-convective boundary (black dot) moves inwards, while the radiative region (thick lines) likewise move in as the envelope isothermalizes with its environment. The temperature-pressure figure shows clearly how the outermost region of the envelope is isothermal with the equilibrium temperature, and how, interior to the radiative-convective boundary, the envelope lies along an adiabat, which is linear on the logarithmic-logarithmic axes shown in the figure.
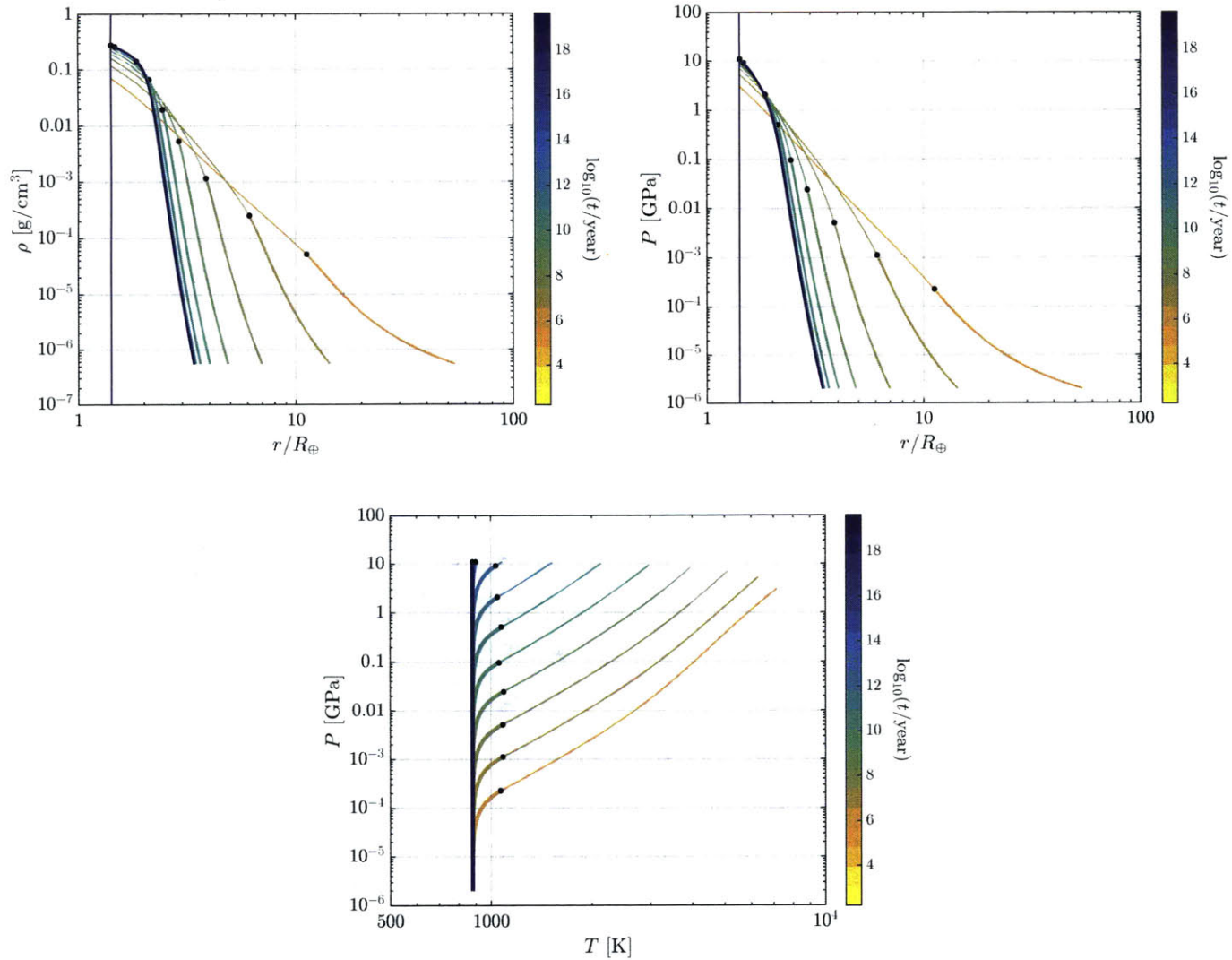
Figure 2-9: **Thermal evolution of a planet with** $M_{core} = 4M_\oplus$ **and** $M_{atm}/M_{core} = 5\%$. **Planet is at** 0.1 **AU after the dissipation of the gas disc. In the upper left panel, we show the density profile as a function of radial distance from the core center. In the upper right panel, we show the pressure profile as a function of radial distance. In the lower panel, we show the temperature-pressure profile. In each case, the outermost radiative region is indicated in bold, with the outermost radiative-convective boundary indicated with a black dot. In the** ρ **vs.** r **plot, note the flattening of the density profile near the core as the envelope cools and compresses, indicating that the transition to metallic hydrogen has occurred.**

Chapter 3

The Ejection of Gaseous Envelopes via Hydrodynamic Shock

The formation history of exoplanets requires giant impacts between protoplanets: collisions of similarly sized bodies at speeds equal to or greater than their mutual escape velocity v_{esc} (see Section 1.2.3). These impacts—which are a natural consequence of the dynamical evolution of planets—can lead to substantial changes in the composition and structure of the resulting planetary body, including catastrophic disruption and ejection of the gaseous envelope, as well as disruption and ejection of the rocky mass. If we wish to consider planet formation in the most general way, we must therefore account for the possibility of giant impacts between planet-sized bodies and the effect it has on the volatile content of the

95

planet. As we will see in Chapters 4 and 5, the consequences are significant and thus important to consider when constructing viable planet formation pathways.
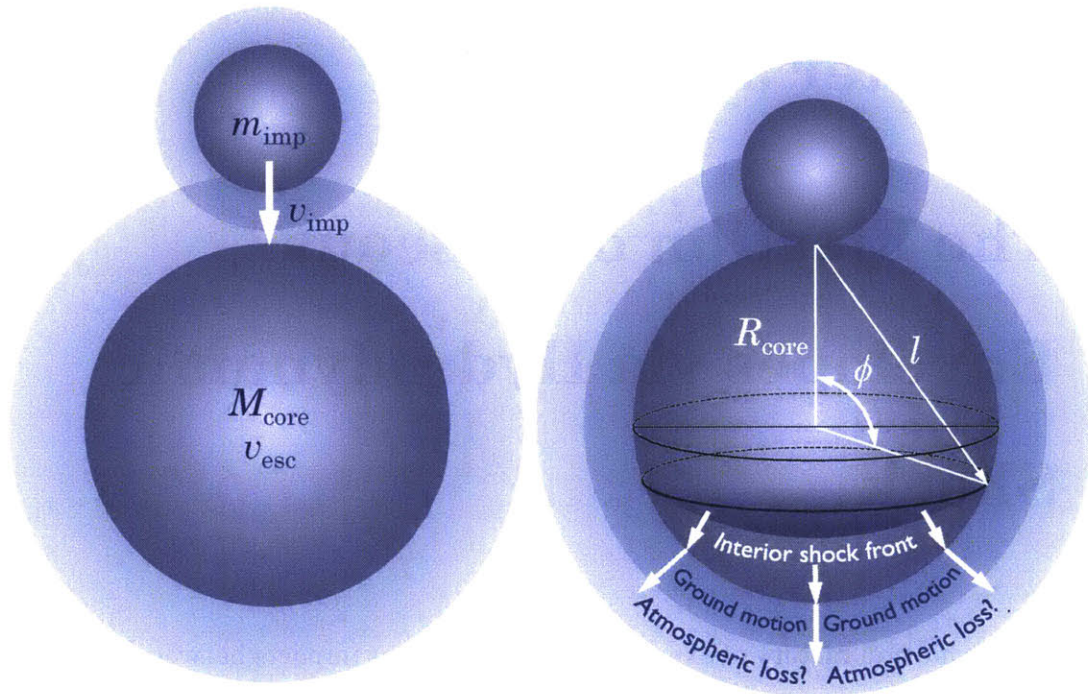
Figure 3-1: Schematic of giant impact-induced atmospheric loss. An impactor of mass m_{imp} and velocity v_{imp} approaches a planetary embryo of core mass M_{core} and escape velocity v_{esc} (left-hand panel). The collision generates a shock wave that propagates through the interior of the embryo core (right-hand panel). As this interior shock wave travels through the core, it subsequently initiates a global ground motion of the planet. This ground motion then launches a shock into the atmosphere above it, potentially leading to hydrodynamic escape.

The geometry of a giant impact is summarized in Fig. 3-1. Consider two planet-sized bodies or planetary embryos encountering one another. The impactor, which we assume for clarity is the less massive of the

two, has a mass m_{imp} and an impact velocity likely between the mutual escape velocity of the two bodies, v_{esc}, and $\sqrt{2}v_{\mathrm{esc}}$ (Section 1.2.3). The impact between the two bodies drives a shock front through the core of the target body. This interior shock in turn leads to global ground motion of the target body. This ground motion in turn launches a shock front into the atmosphere, potentially leading to atmospheric escape.

Our calculations for hydrodynamic shock propagation are one-dimensional, applicable to a given position on the core "surface". We connect these one-dimensional results with global envelope loss by using results from analytical scalings and smooth-particle hydrodynamics (SPH) simulations (Leinhardt & Stewart 2012) to determine how the surface velocity varies globally due to an impact. This parameterization, which produces results in good agreement with much more computationally intensive SPH results (Stewart, Lock, & Mukhopadhyay 2014a;b), also has the advantage of allowing us to probe a broad range of parameter space with much less computational time than SPH.

In this chapter, we present our method for calculating the effect of impacts on envelope mass loss. We begin by introducing the equations of fluid motion, and the standard expressions for mass, momentum, and energy conservation across a shock front. Except for some cases (for instance, very thin atmospheres), we cannot make use of the standard analytical similarity solutions and instead make recourse to numerical methods. We therefore discuss our one-dimensional numerical shock propagation model in some detail and present example results from our

calculations. We describe how we connect our one-dimensional calculation to global envelope mass loss and present example numerical integrations of the kinds we employ in Chapters 4 and 5. We close by calculating how *rocky* mass can be lost due to giant impact. The loss of silicate material and the possible subsequent formation of a debris disc around a planet is of interest because it can be observed and thus provide direct evidence of an important planet formation process (Wyatt & Jackson 2016).

3.1 Fluid Dynamical Equations

The continuum equations describing the evolution of inviscid fluid are given by the usual Euler equations. Consider a fluid particle whose velocity vector $\mathbf{u}$ is given by (u_x, u_y, u_z). It has a local density of ρ (and hence a specific volume $V = 1/\rho$), is subject to a fluid pressure P, and possesses an internal energy ε. If we denote the time derivative of a quantity moving in space with a fluid particle (the Lagrangian derivative) as d/dt, the equations of mass, momentum, and energy conservation, respectively, are (e.g., Landau & Lifshitz 1959, Zel'dovich & Raizer

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \nabla \cdot \mathbf{u} = 0 \tag{3.1a}$$

$$\rho \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} + \nabla P = \mathbf{F} \tag{3.1b}$$

$$\frac{\mathrm{D}\varepsilon}{\mathrm{D}t} + P \frac{\mathrm{D}V}{\mathrm{D}t} = Q, \tag{3.1c}$$

where $\mathbf{F}$ is a body force term, and where Q is an energy source or sink term. If we wish to include gravity, we have $\mathbf{F} = -\nabla\Phi$, where Φ is the gravitational potential. The second term on the left-hand side of Eq. (3.1c) is a feature of a compressible fluid and denotes the work done on the fluid element in compressing it. For an adiabatic process $Q \equiv 0$, and the specific entropy S of the particle is conserved: $\mathrm{d}S/\mathrm{d}t = 0$. The specific entropy of an ideal gas is, up to an arbitrary constant [cf. Eq. (2.29)],

$$S = c_p \log PV^{\gamma_a} + \text{const.}, \tag{3.2}$$

so that the entropy conservation equation gives

---

* The equations can be written in terms of Eulerian derivatives using $\mathrm{D}/\mathrm{D}t = \partial/\partial t + \mathbf{u} \cdot \nabla$:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho}\nabla P + \mathbf{F}$$

$$\frac{\partial}{\partial t}\left[\rho\varepsilon + \frac{\rho(\mathbf{u} \cdot \mathbf{u})}{2}\right] = -\nabla \cdot \left[\rho \mathbf{u}\left(\varepsilon + \frac{\mathbf{u} \cdot \mathbf{u}}{2}\right) + P\mathbf{u}\right].$$

$$\frac{\mathrm{D}p}{\mathrm{D}t} = -\gamma_a \frac{P}{\rho} \frac{\mathrm{D}\rho}{\mathrm{D}t}.$$ (3.3)

3.1.1 Jump Conditions[†]

The differential equations of mass, momentum, and energy conservation of the preceding section apply to fluids in which the quantities vary continuously. In some cases, however, the quantities vary discontinuously. The point, line, or surface (for one-, two-, or three-dimensional motion, respectively) over which the state variables for the fluid vary discontinuously is called the "shock wave" or "shock front". We will primarily be concerned with shock waves traveling through atmospheres. In this case, a shock front is launched into a planetary atmosphere due to ground motion caused by an impact. The shock front then propagates into the atmosphere, sweeping up compressed material behind it (see Fig. 3-2).

Across the shock front, which for the moment we regard as infinitesimally thin, mass, momentum, and energy flux must be conserved. Since the volume bounding the wave is infinitely thin, there are no mass, momentum, or energy sources or sinks. If we denote the state variables of the undisturbed fluid before the shock front with the subscript "+" and

---

[†]The presentation here essentially follows that of Zel'dovich & Raizer (1967).

100
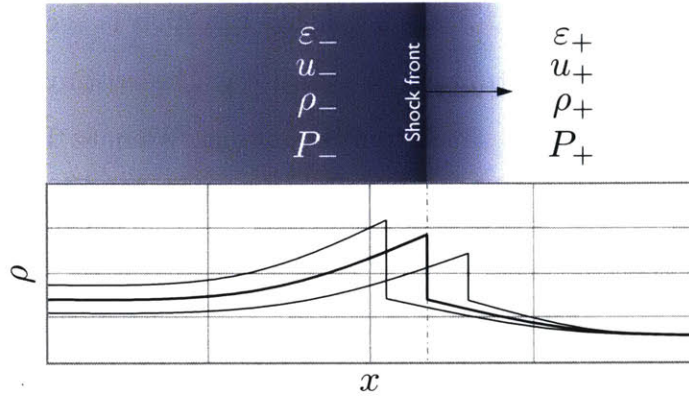
Figure 3-2: The geometry of shock front and the state variables in front of (+) and behind (-) the shock front. The propagation is from left to right, and in the lower panel, density profiles at an earlier and later time have been shown for reference.

those of the compressed material behind it "−", the *shock equations* are

$$\rho_- u_- = \rho_+ u_+ \tag{3.4a}$$

$$P_- + \rho_- u_-^2 = P_+ + \rho_+ u_+^2 \tag{3.4b}$$

$$\varepsilon_- + \frac{P_-}{\rho_-} + \frac{u_-^2}{2} = \varepsilon_+ + \frac{P_+}{\rho_+} + \frac{u_+^2}{2}. \tag{3.4c}$$

Here, the u quantities are the projection of $\mathbf{u}$ normal to the shock surface. These equations can be formally derived by fixing the shock front in space, taking the dot product of the equations with a unit vector normal to the shock front, and integrating the Eulerian forms of the equations of mass, momentum, and energy conservation over an infinitesimally thin control volume bounding the shock surface. In the case of the momentum equation, our derivation of Eqs. (3.4) has assumed that $\mathbf{F}$ is continuous in space. If $\mathbf{F}$ is only due to gravity, then we may use Laplace's equation

101

to note that $\nabla^2 \Phi = 4\pi G \rho$. Since Laplace's equation is second order in space, even if ρ is discontinuous, Φ will always be continuous. Hence, gravity does not affect the shock front equations. We note that using the definition of specific enthalpy $h \equiv \varepsilon + P/\rho$, we may rewrite Eq. (3.4c) as $h_- + u_-^2/2 = h_+ + u_+^2/2$.

3.1.2 Hugoniot Relations and Shock Relations for an Ideal Gas

The algebraic equations Eqs. (3.4) can be combined into several other useful algebraic relationships. In particular, using Eq. (3.4a), we have $V_+/V_- = \rho_-/\rho_+ = u_+/u_-$ and, using Eq. (3.4b), we have

$$u_+ = V_+ \sqrt{\frac{P_- - P_+}{V_+ - V_-}} \tag{3.5a}$$

$$u_- = V_- \sqrt{\frac{P_- - P_+}{V_+ - V_-}} \tag{3.5b}$$

The flow speed of the shock front into the undisturbed fluid is given by Eq. (3.5b). The relative speed of the gas behind the shock front and that before it, $|u_+ - u_-| \equiv u$ is given by

$$|u| = \sqrt{(P_- - P_+)(V_+ - V_-)}; \tag{3.6}$$

for a shock launched into an atmosphere, immediately after the shock is launched, $|u| = v_g$, where v_g is the ground speed. Lastly, Eq. (3.4c) can

be rewritten as

$$h_- - h_+ = \frac{1}{2}(P_- - P_+)(V_- + V_+); \qquad (3.7)$$

this expression is termed the "shock adiabatic" or Hugoniot relation.

For an ideal gas, many of the preceding equations take a simplified form. Since we assume for the sake of our calculations here that our gas is ideal,[‡] we reproduce here some of the key results. The internal energy and enthalpy of an ideal gas have the simple forms (Fermi 1956)

$$\varepsilon = c_v T = \frac{PV}{1 - \gamma_a} \qquad (3.8a)$$

$$h = c_p T = \frac{\gamma_a PV}{1 - \gamma_a}. \qquad (3.8b)$$

If we define the Mach number on each side of the shock front $\mathcal{M}_{+/-} \equiv (u_{+/-})/(c_{s,+/-})$, then we have the relations

$$\frac{\rho_+}{\rho_-} = \frac{(\gamma_a + 1)\mathcal{M}_-^2}{(\gamma_a - 1)\mathcal{M}_-^2 + 2} \qquad (3.9a)$$

$$\frac{P_+}{P_-} = \frac{2\gamma_a \mathcal{M}_-^2 - \gamma_a + 1}{\gamma_a + 1} \qquad (3.9b)$$

$$\mathcal{M}_+^2 = \frac{2 + (\gamma_a - 1)\mathcal{M}_-^2}{2\gamma_a \mathcal{M}_-^2 - \gamma_a + 1}. \qquad (3.9c)$$

There are two limiting regimes in which a shock can propagate: the "weak" and "strong" limits. In the weak limit, pressures across the shock front are essentially equal, so that $\mathcal{M}_- \approx 1 \approx \mathcal{M}_+$ and the densities

---

[‡] Deviations from ideality are discussed at the end of this chapter.

are continuous across the front. Thus, the disturbance propagates as a sound wave at the sound speed, and any discontinuities are small, of third order in pressure difference (Landau & Lifshitz 1959). On the other hand, in the strong limit, the pressure in front of the shock front is formally taken to 0 compared to that behind the shock front, so that $\mathcal{M}_- \to \infty$. This is often the assumption made in the analysis of shock waves due to bombs detonated in Earth's atmosphere (e.g., Taylor 1950, Sedov 1959), and is the case applicable to giant impact-induced shock propagation. We note that interestingly, the density contrast across the shock front in this case takes a finite value, $(\gamma_a + 1)/(\gamma_a - 1)$.

We lastly make some remarks on the *thickness* of a shock front. We have thus far assumed the front is infinitely thin (Fig. 3-2). In reality, however, the shock front has a nonzero thickness. In order to correctly account for this, we must allow for a nonzero viscosity and finite heat conductivity across the front. The momentum and energy conservation expressions, Eqs. (3.4b) and (3.4c), are modified to include momentum flux due to viscous stresses and to include viscous energy dissipation and heat transfer by conduction. The problem is solved in full in Landau & Lifshitz (1959), with the primary qualitative result that the thickness of the shock front is smaller for stronger shocks. In order to handle the discontinuities intrinsic to shock propagation through a fluid, our numerical scheme introduces an *artificial viscosity*, which serves the same purpose as true viscosity (giving the shock front a nonzero thickness), but which is otherwise not physically motivated (see Section 3.3.1 be-

low).

3.2 Fluid Dynamical Equations in a Planetary Atmosphere

In order to solve the hydrodynamical equations applied to the case of a planetary atmosphere, we again adopt a Lagrangian framework in which we track a particle of fixed mass. Assuming spherical symmetry, we therefore track a shell of mass $dm = 4\pi r^2\, dr$, where r is the distance from the center of the coordinate axes (taken here to be the center of the core). Eqs. (3.1) applied to a planetary atmosphere become

$$\frac{Dr}{Dm} = \frac{1}{4\pi r^2 \rho}, \tag{3.10a}$$

$$\frac{Du}{Dt} = -4\pi r^2 \frac{\partial P}{\partial m} - \frac{GM_{\text{core}}}{r^2} - 4\pi G \int_{R_{\text{core}}}^{r} \rho M(< r) dr, \tag{3.10b}$$

$$\frac{Dr}{Dt} = u, \tag{3.10c}$$

where m is the mass enclosed in a mass shell at r, and t is time. Here, $M\,(< r)$ is the atmospheric mass interior to r. Since radiative losses are negligible,[§] we assume that the shock propagates adiabatically, in which case the energy equation Eq. (3.3) is rewritten

$$\frac{DP}{Dt} = -4\pi \rho \gamma P \frac{\partial \left(r^2 u\right)}{\partial m}. \tag{3.11}$$

[§]See Sec. 3.5.2 for an order of magnitude justification of this.

3.2.1 Initial and Boundary Conditions

In order to solve Eqs. (3.10) and (3.11), it is necessary to prescribe initial and boundary conditions. For initial conditions, we always assume that the envelope density structure is that of an adiabatic envelope [Eqs. (2.19)]. We do this for several reasons. For one, we will always be interested in impacts that happen relatively early in the formation history of the planet. In this case, practically all of the envelope mass is in the convective region of the envelope, which follows an adiabat (see also Chapter 2). This also has the benefit of computational ease, since the profile is easily prescribed by an analytical formula. We discuss in further detail how we instantiate the density profile in Sec. 3.3.1.

The dynamical evolution of the shock front and the fluid it encounters depend on how we initiate the motion at the base of the envelope. Here, we follow the method of Genda & Abe (2003) by fixing the ground velocity v_g ($= \mathrm{d}r_g/\mathrm{d}t$, with r_g the radial position of the ground) at $t = 0^+$ to some multiple C of the escape velocity v_{esc} at the core surface, and then force the ground to return back to its initial radial position R_{core} according to a ballistic trajectory. The velocity of the mass parcel immediately adjacent to the ground (which we denote m_0) is set to v_g:

$$u(m_0) = v_g \tag{3.12}$$

$$v_g(t = 0^+) = C \times v_{\mathrm{esc}} \tag{3.13}$$

$$\frac{\mathrm{d}v_g(t > 0)}{\mathrm{d}t} = \begin{cases} -\dfrac{GM_{\mathrm{core}}}{r_g^2} & \text{if } r_g > R_{\mathrm{core}} \\ 0 & \text{else.} \end{cases} \tag{3.14}$$

This method suffers from the shortcoming that, in cases in which the impulsive ground velocity is greater than the escape velocity, r_g still returns to R_{core}. In this regard, our results are conservative, since the return of r_g to R_{core} has the effect of "sucking" the base of the envelope back. A purely piston-like impulse (which would correspond to $v_g(t \geq 0) = \mathrm{const.}$), on the other hand, would likely overestimate local envelope mass loss.

3.3 Solution to the Hydrodynamic Equations

Eqs. (3.10) are not amenable to analytical treatment except in some limiting cases. When the problem is in fact analytically tractable, it is usually approached using similarity methods (Taylor 1950, Sedov 1959, Sari 2006, Schlichting, Sari, & Yalinewich 2015). In these cases, the equations of motion are amenable to solutions in which the state variables are functions of similarity variables, dimensionless scaling parameters that vary as a function of time. One case in which analytic solutions

exist is the plane-parallel case in which the curvature of the envelope is negligible; these solutions have been explored in the context of terrestrial planet atmosphere mass loss in Schlichting, Sari, & Yalinewich (2015).

The plane-parallel approximation becomes physically relevant when the adiabatic index γ_a (which, practically speaking, defines the initial mass distribution in the envelope) is sufficiently small, or when the atmospheric mass is sufficiently small so that the extent of the envelope beyond the core radius R_{core} is small ($R_p/R_{\text{core}} \approx 1$, where R_p is the total planet radius). In the former case, the density ρ profile in an adiabatic envelope scales roughly as $r^{1/(\gamma_a-1)}$ [Eq. (2.19)], so that the mass in the envelope goes as $r^3\rho \sim r^{(4-3\gamma_a)/(\gamma_a-1)}$. Thus if $\gamma_a \lesssim 4/3$ the atmospheric mass is concentrated at smaller radii, while for $\gamma_a \gtrsim 4/3$, the mass is concentrated further away from the core. Mass concentration near the core leads to a situation physically similar to the small radius case. There are additional consequences to this, which we discuss further below. For the cases that we are interested in—in which the gas envelopes constitute a substantial fraction of the total planet mass—curvature and radius-dependent gravity effects become important. Therefore we appeal to numerical methods.

3.3.1 Numerical Methods

The numerical method we use here is a modification of that used by Genda & Abe (2003) and that used by Bodenheimer, Laughlin, Różyczka

et al. (2007). We employ a Lagrangian, finite-difference scheme implementing artificial viscosity. In this section, we review the implementation of our numerical method.

In our hydrodynamic simulations, we assume the initial atmospheric profile is adiabatic, although in principle the initial condition can be arbitrarily set (say to an isothermal profile, or any other hydrostatic configuration). We choose an adiabatic profile because in gas envelopes relevant to sub-Neptune type planets, the majority of the mass lies in a convective region that constitutes (especially during early times) the majority of the envelope mass and radius (see Fig. 2-9).

In order to facilitate numerical solution of the equations of fluid motion, all variables are non-dimensionalized. To do this, we first identify physical variables important to the problem at hand. The planet core radius R_{core} provides a natural scaling for the length. The adiabatic equations Eqs. (2.19) are different from, say, an isothermal profile in that there is a definite radius R_{max} at which the density and pressure vanish. This occurs when $\nabla_{\mathrm{ad}}\lambda\left(R_{\mathrm{core}}/R_{\mathrm{max}} - 1\right) + 1 = 0$, or

$$\lambda = \left[\nabla_{\mathrm{ad}}\left(R_{\mathrm{core}}/R_{\mathrm{max}} - 1\right)\right]^{-1}, \qquad (3.15)$$

so that fixing the planet radius fixes the escape parameter. (Note that, compared to Eqs. (2.19), which were derived by integrating the equations of hydrostatic equilibrium inwards towards the core, the expression above is arrived at by integrating the equations from the core out-

wards, introducing a negative sign into the term after the escape parameter λ and making $R^* \to R_{\mathrm{core}}$.) Specifying the envelope mass then fixes ρ at the core (denoted ρ_0), so that the complete envelope state is specified. The free parameters of the problem are thus M_{core}, M_{atm}, R_{core}, and R_{max}. If the core radius can be parametrized in terms of the core mass, then the number of free parameters is further reduced by one, and by the Buckingham-Π theorem (Buckingham 1914), the physics of the problem can be completely specified by the dimensionless groupings $M_{\mathrm{atm}}/M_{\mathrm{core}}$ and $R_{\mathrm{max}}/R_{\mathrm{core}}$.

To non-dimensionalize the equations of fluid motion, we use the aforementioned quantities and several other derivable quantities. From the core mass, we get $g = GM_{\mathrm{core}}/R_{\mathrm{core}}^2$. We also have the reference pressure $P_0 = gR_{\mathrm{core}}\rho_0/\lambda$ and the reference sound speed $c_{s,0} = \sqrt{\gamma_a P_0/\rho_0}$. Thus, we scale velocities by $c_{s,0}$, lengths by R_{core}, mass by M_{core}, density by ρ_0, and time by $R_{\mathrm{core}}/c_{s,0}$. We finally introduce another dimensionless escape parameter $\lambda_s \equiv G\rho_0 R_{\mathrm{core}}^2/c_{s,0}^2 = \lambda_0/\gamma_a$, which is useful for calculating the acceleration due to gravity from the core and the self-gravity of the envelope.

At each time step, indexed by n, the state variables—position, velocity, pressure, temperature, and density—are updated for a given mass element to a new $(n+1)^{\mathrm{th}}$ value. Each of the state quantities is discretized. Since our scheme is Lagrangian, each discretization corresponds to an initial position in the envelope density profile, or, equivalently, a mass shell. For the ith discretized mass element, the non-

dimensionalized velocity is updated as

$$u_i^{(n)} = u_i^{(n-1)} - \delta t^{(n-1)} \left\{ \frac{8\pi}{\gamma_s} \left(r_i^{(n-1)} \right)^2 \left(\frac{P_i^{(n-1)} - P_{i-1}^{(n-1)} + q_i^{(n-1)} - q_{i-1}^{(n-1)}}{\delta m_{i-1}} \right) \right.$$

$$\left. + \lambda_s \left[\frac{1}{\left(r_i^{(n-1)} \right)^2} + \sum_{j=1}^{i-1} \frac{\delta m_j}{\left(r_j^{(n-1)} \right)^2} \right] \right\} .$$

In this equation, all the quantities are their dimensionless equivalents, while q is the artificial viscosity parameter. δm_i represents the ith mass shell and $\delta t^{(n)}$ the timestep at the nth iteration. All the other symbols have their usual meanings, with the superscripts indicating the timestep, and the subscripts indicating the discretization. The first term in the square brackets is the acceleration due to pressure, while the second term is that due to the gravitational acceleration from the core, and the final term, that due to the self-gravity of the envelope.

The pressure is updated as

$$P_i^{(n)} = P_i^{(n-1)} - 4\pi \rho_i^{(n-1)} \left[\gamma_s P_i^{(n-1)} + (\gamma_s - 1) q_i^{(n-1)} \right] \times$$

$$\left[\frac{\left(r_{i+1}^{(n-1)} \right)^2 u_{i+1}^{(n-1)} - \left(r_i^{(n-1)} \right)^2 u_i^{(n-1)}}{\delta m_i} \right] \delta t^{(n-1)}, \quad (3.16)$$

while the new position and density are

$$r_i^{(n)} = r_i^{(n-1)} + \delta t^{(n-1)} u_i^{(n)} \tag{3.17}$$

$$\rho_i^{(n)} = \frac{3}{4\pi} \frac{\delta m_i}{\left(r_i^{(n)}\right)^3 - \left(r_{i-1}^{(n)}\right)^3}. \tag{3.18}$$

In order to mitigate the numerical discontinuities that naturally arise in propagating the shock front, it is necessary to introduce an artificial viscosity term into the discretized equations of motion. The artificial viscosity has the effect of smoothing the discontinuity at the shock front very slightly, so that the shock front has a finite thickness. In principle, this mimics the effect of real viscosity, which as we have discussed above, gives the shock front a nonzero thickness, although we stress the viscosity we employ here is a numerical artifice and has no direct physical interpretation. In any case, if it is chosen judiciously, sharpness in the profiles can be maintained. The artificial viscosity at time step n is

$$q_i^{(n)} = \begin{cases} -c_q \gamma_s \rho_i^{(n)} \left(u_{i+1}^{(n)} - u_i^{(n)}\right) \left[\sqrt{\frac{P_i^{(n)}}{\rho_i^{(n)}}} - \frac{\gamma_s+1}{2}\left(u_{i+1}^{(n)} - u_i^{(n)}\right)\right], & \text{if } u_{i+1}^{(n)} - u_i^{(n)} < 0 \\ 0, & \text{otherwise.} \end{cases}$$

$$\tag{3.19}$$

Here, c_q is a viscosity coefficient. By virtue of the viscosity model, mass elements are prevented from overtaking a mass element adjacent to it.

In order to ensure numerical stability, time step $\delta t^{(n)}$ is updated according to a Courant-Friedrichs-Lewy-type criterion (Courant, Friedrichs,

& Lewy 1928), which in our non-dimensional system is

$$\delta t^{(n)} = K \times \min \left(\frac{\Delta \mathbf{r}^{(n)}}{\max \mathbf{u}^{(n)}} \right), \qquad (3.20)$$

where K is a numerical constant, $\Delta \mathbf{r}^{(n)}$ is the vector of position differentials $r_i^{(n)} - r_{i-1}^{(n)}$ at n, and $\mathbf{u}^{(n)}$ is the vector of velocities at n. In our simulations, we typically take $K = 1/14$, although other values as high as $1/4$ and as low $1/20$ have also been used successfully. Since the elapsed time after n integrations is $t^{(n)} = \sum_{m=1}^{n} \delta t^{(m)}$, it is advantageous to choose K as large as possible to reduce the computation time. Our simulation parameters are summarized in Table 3.1.

Table 3.1: Summary of baseline parameters for hydrocode simulations.

Parameter	Value	Description
N	200,000	Number of time steps
M	300	Number of radial discretizations
K	1/4	CFL stability coefficient
c_q	0.25	Viscosity parameter

3.4   Local and Global Atmospheric Mass Loss

3.4.1   Local Atmospheric Mass Loss

In our simulations, we consider the one-dimensional propagation of a shock through a gaseous envelope as a function of the ground motion

induced by an impact. Our strategy is to determine the atmospheric mass loss at one location on the planet using our numerical model, and, by connecting the impact characteristics to the global ground motion, to then determine the global envelope mass loss by integrating over the surface of the planet.

Locally, an atmospheric mass parcel is lost from the planet if, after a long time, the velocity it attains is larger than its initial, radius dependent escape velocity $v_{\mathrm{esc}}(r_0) = \sqrt{2GM_{\mathrm{core}}/r_0}$, where r_0 is the initial radius. This condition can be relaxed somewhat if the planet is very close in to its host star, so that its Roche or Rill radius is accessible. In this case, the mass parcel needs only enough kinetic energy at large times such that it can overcome the potential at the Hill radius. Thus the most general criterion for envelope mass loss is

$$\frac{u^2(t \gg 0)}{2} - \frac{GM_{\mathrm{core}}}{R_H} \geq -\frac{GM_{\mathrm{core}}}{r_0}, \tag{3.21}$$

or

$$u(t \gg 0) \geq v_{\mathrm{esc}}(r_0)\sqrt{1 - \frac{r_0}{R_H}}. \tag{3.22}$$

Since typically $r_0 \ll R_H$ (approaching it only if $a \lesssim 0.01$ AU), the second term under the radical can usually be ignored. All mass parcels with velocities greater than $v_{\mathrm{esc}}(r_0)$ are considered lost to the system, and the mass fraction lost locally χ_{local} is simply the sum of all mass parcels lost over the total mass of the envelope column.

3.4.2 Global Atmospheric Mass Loss

Connecting local atmospheric mass loss to the global atmopsheric mass loss fraction, which we denote χ_{global}, requires us to determine the distribution of ground velocities over the surface of the core. In order to do this, we must determine how ground velocities are a function of impact characteristics.

To proceed, we make use of numerical and analytical results that indicate for disruptive collisions in gravity-dominated bodies (as opposed to strength-dominated bodies), shock propagation through the body is consistent with conservation of the momentum of the impactor (Leinhardt & Stewart 2012). In a center of mass frame, the momentum of the impactor p_{imp} is given by

$$p_{\mathrm{imp}} = \frac{m_{\mathrm{imp}} v_{\mathrm{imp}}}{1 + m_{\mathrm{imp}}/M_{\mathrm{core}}}, \tag{3.23}$$

where the impact velocity v_{imp} is given by Eq. (1.10) and the direction of p_{imp} is along the direction of the relative impact velocity. As the shock sweeps up mass through the target body, then, the velocity of the mass that is swept out must have a velocity consistent with conserving p_{imp}.

Referring to Fig. 3-1, it is straightforward to show that the volume swept out by a shock traversing a distance l from the point of impact is given by

$$\frac{4\pi}{3} R_{\mathrm{core}}^3 \left(\frac{l}{2R_{\mathrm{core}}} \right)^3 \left(4 - \frac{3l}{2R_{\mathrm{core}}} \right). \tag{3.24}$$

115

If we assume that the density in the core is constant (which is reasonable if the core is well-mixed),[¶] then using momentum conservation, the velocity of the shocked core mass v_s is

$$v_s = v_{\text{imp}} \left(\frac{m_{\text{imp}}}{M_{\text{core}} + m_{\text{imp}}} \right) \left(\frac{l}{2R_{\text{core}}} \right)^3 \left(4 - \frac{3l}{2R_{\text{core}}} \right). \qquad (3.25)$$

This speed is again, along the direction of the relative impact velocity. Projecting this normal to the ground surface (which is the component that goes into launching the shock) and normalizing against the mutual escape velocity gives for the ground speed v_g at the moment the shock is launched into the envelope

$$\frac{v_g}{v_{\text{esc}}} = \left(\frac{v_{\text{imp}}}{v_{\text{esc}}} \right) \left(\frac{m_{\text{imp}}}{M_{\text{core}} + m_{\text{imp}}} \right) \left(\frac{l}{2R_{\text{core}}} \right)^2 \left(4 - \frac{3l}{2R_{\text{core}}} \right). \qquad (3.26)$$

If the angle subtended by a given point on the core surface is given by ϕ (with $l/(2R_{\text{core}}) = \sqrt{(1 - \cos\phi)/2}$; Fig. 3-1), then χ_{global} is related to χ_{local} by

$$\chi_{\text{global}} = \frac{1}{2} \int_0^\pi \chi_{\text{local}}(v_g/v_{\text{esc}}, \phi) \sin\phi \, d\phi. \qquad (3.27)$$

3.5 Example Numerical Integrations

In this section, we briefly present the results of some numerical shock integrations. The results shown here are for fixed parameters repre-

---

[¶]See also Section 2.5 for a discussion on how well-mixed the core might be in early times.

sentative of exoplanet atmospheres. We show the results only for local shock profiles, while the remainder of our results for χ_{local} and χ_{global} are presented in context in Chapters 4 and 5.

In Fig. 3-3 we show density, pressure, and velocity profiles for various times for a $4M_\oplus$ core and an envelope mass fraction of 5%. The core radius R_{core} is $1.41R_\oplus$, and the total planet radius is $10R_{\mathrm{core}} = 14.1R_\oplus$, chosen to represent an extended radius at early times. The ratio of specific heats $\gamma_a = 7/5$, representing molecular hydrogen, and the ground velocity $v_g = 0.5v_{\mathrm{esc}}$. For this case, we calculate the local atmospheric mass loss fraction (after the shock front has passed through the envelope) to be $\chi_{\mathrm{local}} = 0.22$.

In contrast, in Fig. 3-4, we show how the state variables evolve as a function of time with the same simulation parameters (including $v_g = 0.5v_{\mathrm{esc}}$) but with $\gamma_a = 1.1$. Here, the choice of $\gamma_a = 1.1$ is motivated by the case in which partial dissociation of H_2 occurs within the envelope (see Fig. 2-7). For $\gamma_a = 1.1$, the initial density profile has far more mass concentrated towards the envelope base. In particular, we note the vastly different scales for ρ in the upper left panels in Figs. 3-3 and 3-4. In this case, after the shock front has passed through the envelope, $\chi_{\mathrm{local}} = 0.38$.

In each case, we note a relatively sharp shock front passing through the envelope. In the $\gamma_a = 1.1$ case, the shock reaches the envelope edge more quickly than in the $\gamma_a = 7/5$ case. Here, the core can transfer more momentum to the envelope than in the $\gamma_a = 7/5$ case because of the

greater concentration of envelope mass near the core. As a result, the effect of gravity in slowing down the shock is less relative to the case in which mass is concentrated further.

3.5.1 Some Other Features of Shock Propagation in Atmospheres

We note here briefly some other interesting aspects of variations in the parameter space, which will also be evident in the results presented in Chapters 4 and 5.

As indicated above, γ_a plays an interesting role in how shocks develop in planetary atmospheres and how much envelope mass loss occurs due to a giant impact. More generally, either a low envelope mass atmosphere or a small γ_a atmosphere will behave as a plane-parallel atmosphere. This has the effect of essentially reducing the free variables in the problem to just the characteristics of the ground velocity, since the geometrical or curvature effects associated with an extended envelope are negligible.

Consider now the case in which $\gamma_a \gtrsim 4/3$, in which the envelope is more extended than the plane-parallel case. Suppose a gas envelope has fixed $M_{\mathrm{atm}}/M_{\mathrm{core}}$ and consider two planetary radii, $R_p = R_L$ and $R_p = R_S$, where $R_L > R_S$. We assume momentarily that the escape velocity v_{esc} from the planet is fixed at $\sqrt{2GM_{\mathrm{core}}/R_{\mathrm{core}}}$. In the case in which $R_p = R_L$, since the envelope is more extended than in the R_S case, there is less mass closer to the core than for the $R_p = R_L$ case. For a fixed ground

velocity then, the envelope near the core has less initial momentum and can therefore be slowed down more quickly by the gravity from the core (and to a lesser extent, the envelope below it), leading to less mass loss. If we now take $v_{\mathrm{esc}} = v_{\mathrm{esc}}(r)$, however, the picture is somewhat more complicated, and in fact there are some instances in which the atmospheric mass loss increases at very small ground velocities compared to the $v_{\mathrm{esc}} = \sqrt{2GM_{\mathrm{core}}/R_{\mathrm{core}}}$ case.

There is ultimately a competition between where most of the atmospheric mass is located and the range over which the launched envelope mass is slowed down based on its initial momentum. As we will see in Chapter 5, if we assume $M_{\mathrm{atm}}/M_{\mathrm{core}} = 0.01$ and $M_{\mathrm{core}} = 4M_{\oplus}$, there is more mass loss at low ground velocities for an extended envelope ($R_p = R_L$) since it is possible to take advantage of the height-dependent escape velocity at higher altitudes, whereas for the more condensed atmosphere ($R_p = R_S$), the case is more similar to the plane-parallel situation since v_{esc} changes relatively little over the extent of the envelope. At higher ground velocities, the envelope at the base has initially more momentum, which the more condensed atmosphere (having more mass towards the base) can transport more efficiently to outer parts of the envelope than the extended atmosphere, leading to more mass loss.
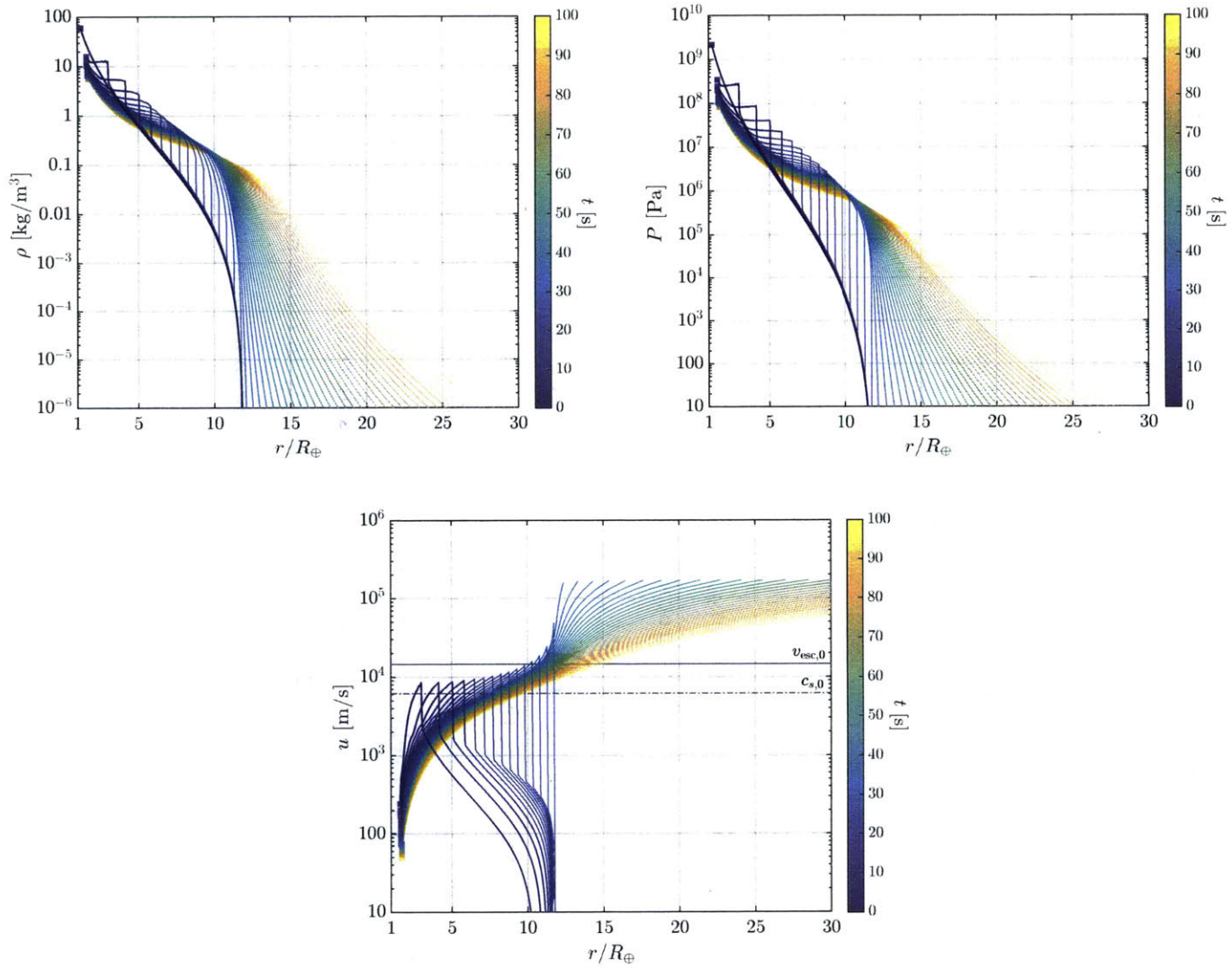
Figure 3-3: Density, pressure, and velocity profiles as a function of time with $\gamma_a = 7/5$.
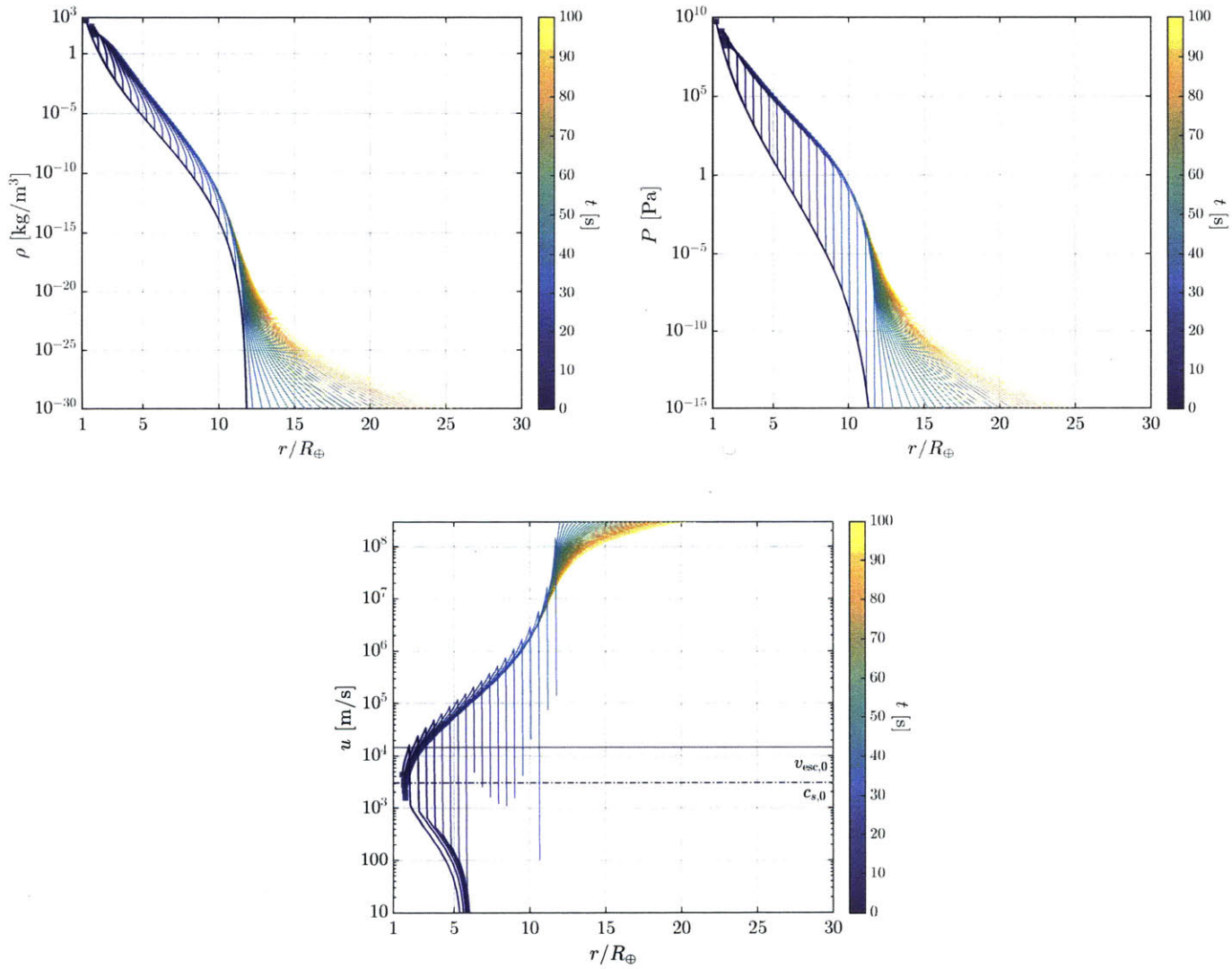
Figure 3-4: Density, pressure, and velocity profiles as a function of time with $\gamma_a = 1.1$.

3.5.2 Radiative Losses

In our calculations, we have assumed the shock propagates adiabatically, without any heat being lost to the environment via radiation. We justify this assumption with an order of magnitude calculation.

The relative effect of radiative losses can be estimated as

$$\frac{L_{\mathrm{rad}}}{E_{\mathrm{shock}}/\tau_{\mathrm{shock}}} \equiv \eta \tag{3.28}$$

where L_{rad} is the radiative power, E_{shock} is the energy propagated by the shock, and τ_{shock} is the shock propagation timescale calculated without accounting for radiative losses. τ_{shock} is typically on the order of minutes or hours: $\tau_{\mathrm{shock}} \sim 10^2 - 10^3$ s (see, e.g., Fig. 3-3). The energy of the shock can be calculated from the ground velocity:

$$E_{\mathrm{shock}} \sim \left(\frac{M_{\mathrm{atm}}}{M_{\mathrm{core}}}\right) M_{\mathrm{core}} v_{\mathrm{g}}^2, \tag{3.29}$$

while the radiative power $L_{\mathrm{rad}} \sim 4\pi R_{\mathrm{core}}^2 \sigma_{\mathrm{B}} T_{\mathrm{shock}}^4$, where the characteristic shock temperature T_{shock} can be found by considering the heat capacity $c_{p,\mathrm{env}}$ of the shocked gas:

$$T_{\mathrm{shock}} \sim \frac{E_{\mathrm{shock}}}{c_{p,\mathrm{env}} M_{\mathrm{atm}}}. \tag{3.30}$$

Solving for $L_{\rm rad}$ and η, we find

$$L_{\rm rad} \sim \frac{4\pi R_{\rm core}^2 \sigma_{\rm B} E_{\rm shock}^4}{c_{p,{\rm env}}^4 \left(M_{\rm atm}/M_{\rm core}\right)^4 M_{\rm core}^4} \tag{3.31}$$

$$\eta = 4\pi \frac{\tau_{\rm shock} \sigma_{\rm B}}{c_{p,{\rm env}}^4} \left(\frac{M_{\rm atm}}{M_{\rm core}}\right)^{-1} \frac{GM_{\rm core}^2}{R_{\rm core}} \left(\frac{v_g}{v_{\rm esc}}\right)^6. \tag{3.32}$$

In familiar units (and using $R_{\rm core} \propto M_{\rm core}^{1/4}$), we have

$$\eta = 10^{-8} \left(\frac{\tau_{\rm shock}}{100~{\rm s}}\right) \left(\frac{M_{\rm core}}{M_\oplus}\right)^{1.75} \left(\frac{v_g}{v_{\rm esc}}\right)^6 \left(\frac{c_{p,{\rm env}}}{10^4~{\rm J \cdot kg/K}}\right)^{-4} \left(\frac{M_{\rm atm}}{M_{\rm core}}\right)^{-1}. \tag{3.33}$$

Even with $c_{p,{\rm env}} \sim k_{\rm B}/(\mu m_{\rm p}) = 3.5 \times 10^3~{\rm J \cdot kg/K}$, η is typically very small. Cases in which radiative losses are non-negligible are those in which the envelope mass is small (say, $M_{\rm atm}/M_{\rm core} \sim 10^{-6}$), and locations on the core surface at which the ground velocity is very high (near the impact point, for instance).

3.6   Core Mass Loss

Thus far, we have considered envelope mass loss from planetary embryos due to giant impacts. However, extensive numerical work has shown that the heavy element mass of a planet can also be lost due to a collision. Quantifying this is of interest since the ejection of silicate material after a collision can lead to the formation of debris discs around young planets, which can then in turn be observed (Wyatt & Jackson 2016). In
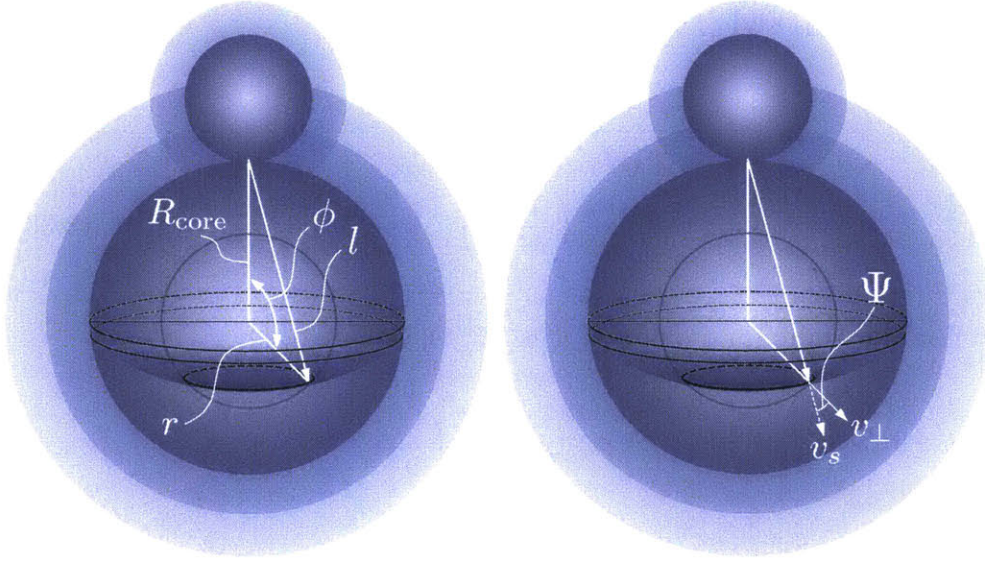
Figure 3-5: Core mass loss geometry.

this section, we construct an analytical model (similar to the model for envelope mass loss discussed above) to quantify core mass loss due to collisions.

We refer to Fig. 3-5 for the geometry used to determine core mass loss. A volume dV of a ring at a position defined by the coordinates (r, ϕ) is $2\pi r^2 \sin\phi \, d\phi \, dr = 2\pi R_{\mathrm{core}}^3 (r/R_{\mathrm{core}})^2 \sin\phi \, d\phi \, d(r/R_{\mathrm{core}})$ (Fig. 3-5, left panel). The shocked velocity v_s of a mass parcel a distance l from the point of impact is again

$$v_s = v_{\mathrm{imp}} \left(\frac{m_{\mathrm{imp}}}{m_{\mathrm{imp}} + M_{\mathrm{core}}} \right) \left(\frac{l}{2R_{\mathrm{core}}} \right)^{-3} \left[4 - 3 \left(\frac{l}{2R_{\mathrm{core}}} \right) \right]^{-1}, \qquad (3.34)$$

124

while the pair (r, ϕ) is related to l through

$$\left(\frac{l}{2R_{\text{core}}}\right)^2 = \frac{1}{4}\left[1 + \left(\frac{r}{R_{\text{core}}}\right)^2\right] - \frac{1}{2}\left(\frac{r}{R_{\text{core}}}\right)\cos\phi. \qquad (3.35)$$

The velocity $v_\perp$ perpendicular to the sphere enclosed by r is $v_s \cos\Psi$ (right panel of Fig. 3-5). The mass enclosed in dV acquires a velocity perpendicular to the sphere given by

$$\frac{v_\perp}{v_{\text{esc}}} = \left(\frac{v_{\text{imp}}}{v_{\text{esc}}}\right)\left(\frac{m_{\text{imp}}}{m_{\text{imp}} + M_{\text{core}}}\right)\left(\frac{l}{2R_{\text{core}}}\right)^{-3}\left[4 - 3\left(\frac{l}{2R_{\text{core}}}\right)\right]^{-1}\cos\Psi,$$

$$(3.36)$$

where

$$\cos\Psi = \frac{\left(\dfrac{l}{2R_{\text{core}}}\right)^2 - \dfrac{1}{4}\left[1 - \left(\dfrac{r}{R_{\text{core}}}\right)^2\right]}{\left(\dfrac{l}{2R_{\text{core}}}\right)\left(\dfrac{r}{R_{\text{core}}}\right)}. \qquad (3.37)$$

For each mass ring defined by (r, ϕ), if $v_\perp \geq v_{\text{esc}}$, it is considered lost. The results for the core mass loss fraction $\chi_{\text{loss,core}}$ for a range of normalized impact velocities are shown in Fig. 3-6.

In the calculations above, it is implicit that the escape velocity is constant. However, the gravitational potential Φ within a uniform sphere is not the same as a point source, but instead is given by

$$\Phi(r) = -\frac{GM_{\text{core}}}{2R_{\text{core}}}\left[3 - \left(\frac{r}{R_{\text{core}}}\right)^2\right]. \qquad (3.38)$$

125

In this case, the escape velocity for a mass parcel at r depends on r, so that the criterion for loss is instead

$$\frac{v_\perp}{v_{\rm esc}} \geq \sqrt{\frac{1}{2}\left[3 - \left(\frac{r}{R_{\rm core}}\right)^2\right]}, \qquad (3.39)$$

where here $v_{\rm esc}$ indicates the escape velocity at the surface. The difference in core mass loss between the case in which $v_{\rm esc} = \sqrt{2GM_{\rm core}/R_{\rm core}}$ and in which it varies as a function of r is shown in Fig. 3-6.

In Fig. 3-7, we show a contour map of a normalized $v_\perp$ on a cross-section of the core. Mass within the sphere in the positive y-axis is not lost to the system since its velocity is not positive, but rather inwards towards the core center. The line for which $v_\perp = v_{\rm esc}$ (for the case in which $v_{\rm imp} m_{\rm imp}/[v_{\rm esc}(m_{\rm imp} + M_{\rm core})] = 1$) is indicated with a black line. The asymptotic limit that the mass loss curve in Fig. 3-6 attains can therefore be understood: in our model, the mass outside of this sphere can never be lost because any velocity vector pointing there has no radially outward projection. The maximum mass that can be lost is the difference in masses enclosed in a volume of a sphere of radius $R_{\rm core}$ and that within $R_{\rm core}/2$, or $\max(\chi_{\rm loss,core}) = 0.875$. An impact between bodies of similar mass (such that $m_{\rm imp} \approx M_{\rm core}$) with a random velocity approximately $v_{\rm esc}$ (such that $v_{\rm imp} \approx \sqrt{2}v_{\rm esc}$) might result in $\sim 10\%$ of the core mass being lost, while an impact with greater mass contrast (say between Earth- and Mars-sized bodies) might result in about 0.1% of the core mass being lost. On the other hand, an impact occurring at about

126

twice the escape velocity might result in nearly all of the core mass being lost.

Our simple analytical model can be recast in a form comparable to SPH simulations. Asphaug (2010) defines an accretion efficiency parameter $\xi \equiv (M_1 + M_{\mathrm{core}})/m_{\mathrm{imp}}$, where M_1 is the largest accreted mass after an impact. He calculated ξ for a range of impact mass ratios $\gamma \equiv m_{\mathrm{imp}}/M_{\mathrm{core}}$ and impact velocities using SPH for a variety of impact angles. In our simplified model, we assume the collisions are head-on, and we set $M_1 = M_{\mathrm{core}}(1 - \chi_{\mathrm{loss,core}}) + m_{\mathrm{imp}}$, so that ξ simplifies to

$$\xi = 1 - \frac{\chi_{\mathrm{loss,core}}}{\gamma}. \tag{3.40}$$

A value $\xi = 1$ implies that no mass is lost, i.e., the collision is a perfect merger and that accretion is perfect. A value of $\xi = 0$ on the other hand implies that $\chi_{\mathrm{loss,core}} = \gamma$, or that the amount of mass that is ejected is equal to that of the impactor. $0 < \xi < 1$ implies net mass accretion for the largest body. Finally, $\xi < 0$ implies that the collision is erosive. This simple analytical model, while not capturing subtleties such as those associated with hit-and-run collisions, nevertheless generally agrees with Asphaug's results (Fig. 3-8).
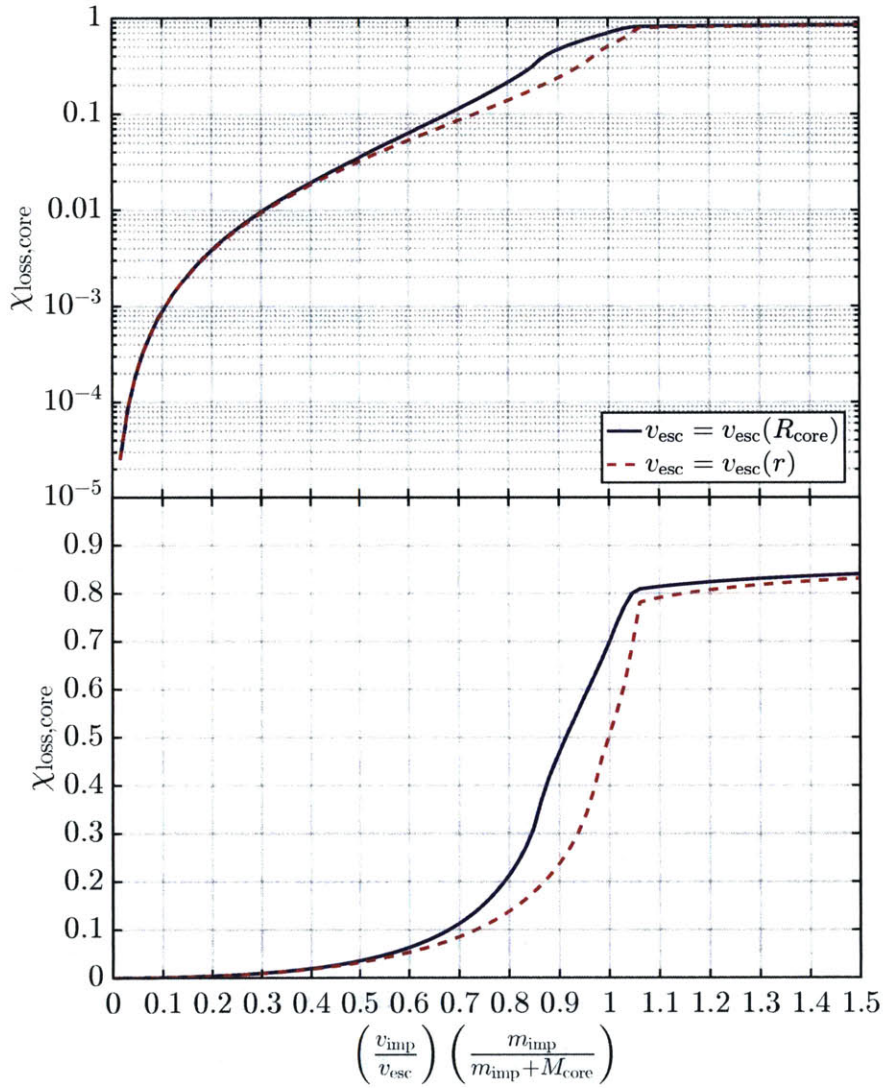
Figure 3-6: Core mass loss from an impact assuming momentum conservation. In both panels, we show the core mass loss fraction $\chi_{\text{loss,core}}$ as a function of normalized impactor momentum. In the upper panel, $\chi_{\text{loss,core}}$ is shown on a logarithmic scale, while in the lower panel, we show the same on a linear scale. In each case, we show mass loss results assuming a constant escape velocity (solid blue line) and a radius-dependent velocity in the core interior (dashed crimson line). The maximum mass loss fraction attained is 0.875.

128
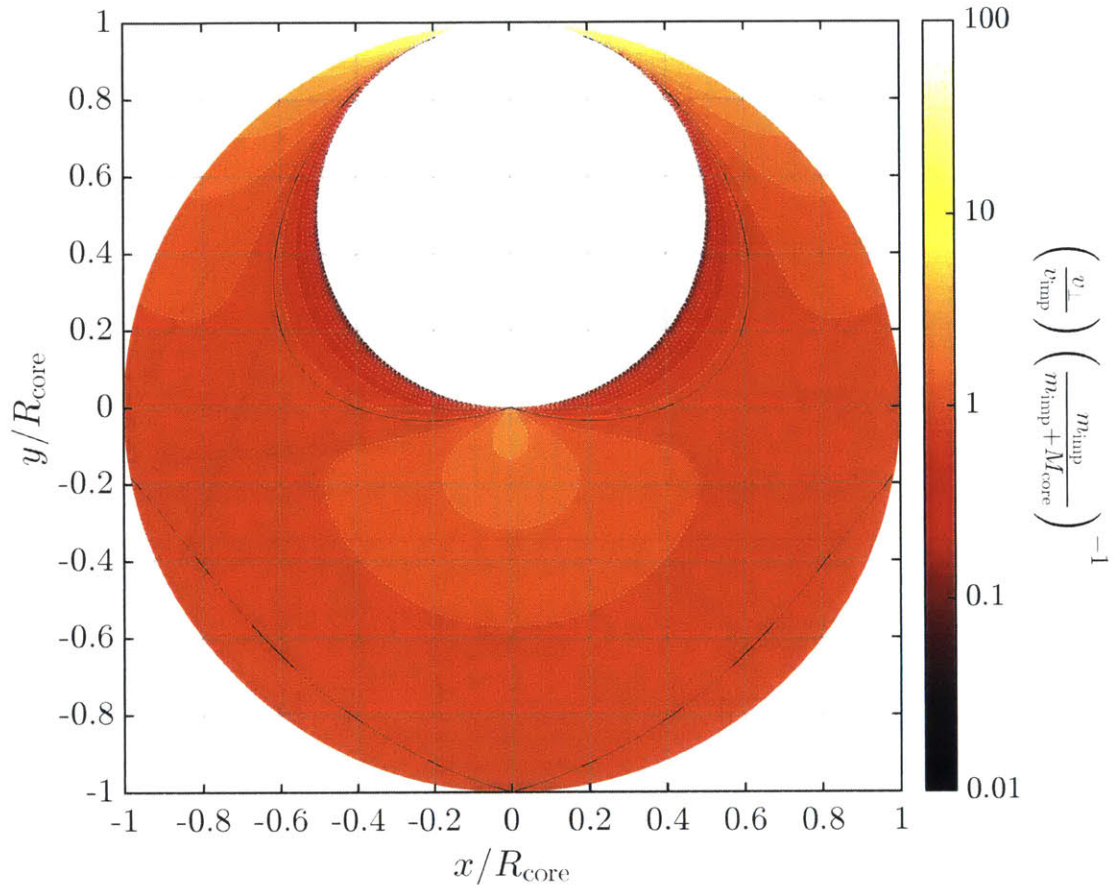
Figure 3-7: Interior velocity map for core mass loss. We show a cross-section of the velocity field $v_\perp$ in the core, normalized according to Eq. (3.36). The black line indicates where $(v_\perp/v_{\rm imp})[m_{\rm imp}/(m_{\rm imp} + M_{\rm core})]^{-1} = 1$.
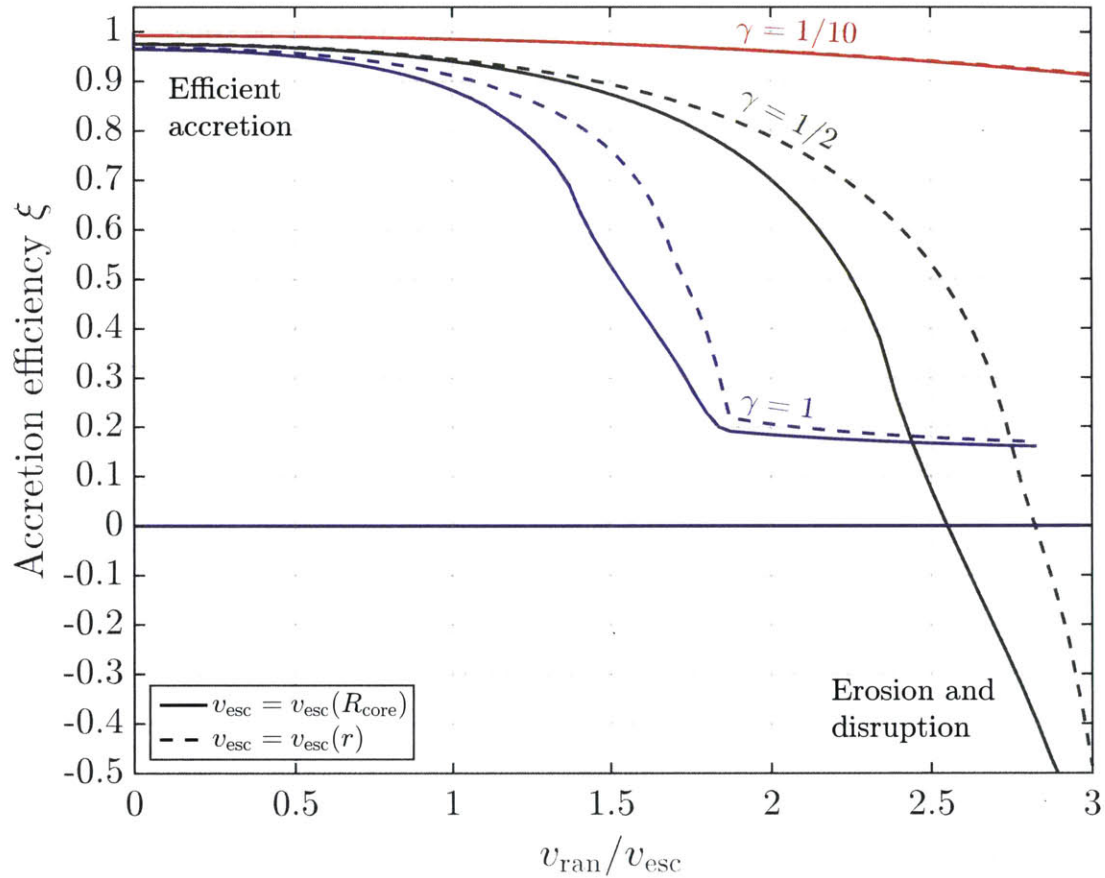
Figure 3-8: Accretion efficiency of rocky mass due to a collision. The accretion efficiency parameter of Asphaug (2010) is shown as a function of impact speed, with regimes of efficient accretion and disruption indicated. Three impactor to core mass ratios γ are shown, as are the cases in which the escape velocity is $\sqrt{2GM_{\mathrm{core}}/R_{\mathrm{core}}}$ (solid lines) and in which it varies as a function of r (dashed lines).

3.7 Summary

In this chapter we have presented a framework by which we can calculate the envelope mass loss due to an impact between two planetary mass bodies. Such an impact results in the generation and propagation of a shock wave throughout a planetary atmosphere, potentially leading to devolatization. We began by introducing the one-dimensional equations of fluid motion and the shock equations. We then presented a numerical scheme by which these equations can be solved. Our one-dimensional integration scheme allows us to calculate atmospheric mass loss at a given location on the core surface. In order to calculate global envelope mass loss due to an impact, we then presented a scheme consistent with SPH results that allows us to parameterize global ground motion (and hence envelope mass loss) in terms of impactor momentum. We closed by analytically calculating the amount of *solid* mass we expect to lose due to a giant impact by using a framework similar to that used to calculate envelope mass loss.

Giant impacts are a natural consequence of planet formation, and the construction of any self-consistent formation model requires accounting for them, as does the interpretation of observational data. In the following two chapters, we apply the results presented both in this chapter and in Chapter 2 to planet formation in the presence of the gas disc (Chapter 4) and after the gas disc has dissipated (Chapter 5) in order to better constrain the formation mechanisms underlying the observed exoplanet population.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

The Formation of

Super-Earths and

Mini-Neptunes with Giant

Impacts*

4.1 Introduction

In the previous two chapters, we have presented a framework for the
thermal evolution of gas envelopes and impact-induced devolatization
that can now be applied towards understanding specific formation path-
ways for exoplanets and exoplanetary systems. In this chapter and
Chapter 5, we use observational data and theoretical models in order

---

*This chapter is based on Inamdar & Schlichting (2015).

to investigate where and how exoplanets form. In this chapter, we focus on formation of super-Earths and mini-Neptunes in the presence of the gas disc.

As we have discussed in Chapter 1, it is now clear that close-in super-Earths and mini-Neptunes are amongst the most common planets in our Galaxy (Howard, Marcy, Johnson et al. 2010, Borucki, Koch, Basri et al. 2011, Batalha, Rowe, Bryson et al. 2013; see also Fig. 1-3). Based on their bulk density, it is also clear that many of these planets consist of rocky or icy cores surrounded by large H/He envelopes (Adams, Seager, & Elkins-Tanton 2008, Lopez & Fortney 2014; Fig. 1-4). In some cases, these envelopes must comprise a few percent or more of the overall planet mass and have radii commensurate with the core radius (by comparison, Earth's atmosphere makes up about 10^{-6} of its total mass and 10^{-3} its total radius).

In order to understand the origin of super-Earths and mini-Neptunes, recent work has considered the *in situ* accretion of gas envelopes by fully formed cores of several Earth masses. Lopez, Fortney, & Miller (2012) focused on modeling the thermal evolution of mini-Neptunes to determine the composition and structure of the Kepler-11 system. Lopez & Fortney (2014) investigated the thermal evolution and structure of mini-Neptunes in general, and calculated mass-radius relationships that were then used to determine core and envelope masses and radii for a number of observed exoplanets. Bodenheimer & Lissauer (2014) and Rogers, Bodenheimer, Lissauer et al. (2011) considered the accretion and evolu-

tion of mini-Neptunes over a variety of semimajor axes. In their models, the core luminosity is provided by accreting planetesimals [which do not appreciably change the core mass; see Eq. (2.27)], and they conclude that formation further out in the disc is likely. Ikoma & Hori (2012) likewise considered gas accretion by a fully formed core in the presence of a dissipating gas disc, assuming that the planetary cores first migrated inwards and accreted their envelopes *in situ*.

A common assumption of all these investigations is that the planetary cores have already formed before they proceed to accrete their gaseous envelopes. However, as we have discussed in Section 1.2, planet formation likely proceeds through several, distinct phases (e.g., Safronov 1972, Goldreich, Lithwick, & Sari 2004), and the core masses that can be attained in the inner disc by accreting all the solids locally available are typically much less than that of the Earth, so that *in situ* formation of super-Earths or mini-Neptunes requires a phase of giant impacts. Recall from Eq. (1.9) that, assuming a shear-dominated velocity dispersion, the isolation mass M_{iso} that can form in the absence of migration at a semimajor axis a is given by

$$M_{\mathrm{iso}} = \left[\frac{4\pi a^2 C \Sigma_s}{(3M_\star)^{1/3}} \right]^{3/2} , \tag{4.1}$$

where $M_\star$ is the mass of the host star, Σ_s is the surface density of solids in the disc, and C is a constant that describes the radial extent of the isolation mass's feeding zone in units of Hill radius. The Hill radius, which

defines the extent over which the gravitational force of the planet overcomes tidal interactions with the host star, is given by $r_H = a \left(M_{iso}/3M_\star\right)^{1/3}$.

While in the outer disc, at several tens of AU, M_{iso} is on the order of a Neptune mass for a MMSN-type disc (Hayashi 1981), in the inner disc, isolation masses are only a fraction of an Earth mass (Fig. 1-9), so that forming terrestrial planets or more massive cores in the inner disc requires an additional stage of assembly. In this giant impact phase (Chambers & Wetherill 1998, Agnor, Canup, & Levison 1999), collisions between protoplanets ultimately result in the formation of several terrestrial planet-mass bodies (Chambers 2001, Raymond, Quinn, & Lunine 2004). We note that while it may seem possible that an arbitrary increase in disc solid density can allow for isolation masses in the inner disc of several $M_\oplus$, it has been shown that for standard gas-to-dust ratios $Z \sim 100$, the corresponding gas disc is unstable for a significant fraction of observed exoplanets within 0.1 AU of their host star (Schlichting 2014).

The necessity of giant impacts to assemble sufficiently large masses in the inner disc has motivated the development of models in which close-in super-Earths and mini-Neptunes form *in situ* via giant impacts. In one proposed scenario, $50 - 100M_\oplus$ of rocky material is delivered to the inner part of the protoplanetary disc (Hansen & Murray 2012), while in another, the protoplanetary disc has surface densities enhanced relative to the MMSN (Chiang & Laughlin 2013). In both cases, the final assembly of the planets occurs via giant impacts. In this chapter, we test

the viability of such *in situ* formation scenarios by examining how much gaseous envelope could have been accreted by planetary embryos before giant impacts, how much could be retained throughout the giant impact phase, and how much could have subsequently been accreted from a depleted disc after the giant impact phase.

This chapter is organized as follows. In Section 4.2, we calculate the atmospheric masses accreted by isolation masses for various semimajor axes. In Section 4.3, we calculate both the local and global atmospheric mass loss as a function of impactor mass and velocity for planets with various atmosphere-to-core mass ratios. In Section 4.4, we construct a giant impact history and calculate atmospheric masses that can be retained throughout the giant impact phase using the results from Section 4.3. In Section 4.5, we explore the possibility that envelope accretion occurred after giant impacts in a depleted gas disc. Discussion and conclusions follow in Section 4.6.

4.2 Accretion of Envelopes by Isolation Masses

4.2.1 Isolation Masses

To determine M_{iso} for a typical close-in planet, we first calculate the minimum surface density of solids Σ_s in the disc needed to form the planet *in situ* with giant impacts. For a planet of mass M_p, $\Sigma_s = M_p/(2\pi a \Delta a)$, where for giant impacts, the annulus width $\Delta a \simeq 2v_{\mathrm{esc}}/\Omega$ (Schlichting 2014). Recall $v_{\mathrm{esc}} = \sqrt{2GM_p/R_p}$ is the escape velocity of the planet and

$\Omega = \sqrt{GM_\star/a^3}$ is the Keplerian frequency of its orbit. To calculate v_{esc}, we use in this chapter the approximate mass-radius relationship (Lissauer, Ragozzine, Fabrycky et al. 2011)

$$M_p = M_\oplus \left(\frac{R_p}{R_\oplus} \right)^{2.06} , \qquad (4.2)$$

often used for Earth- to Neptune-sized planets. Using Eq. (4.1), we then calculate the corresponding isolation mass M_{iso}.

In Fig. 4-1, we show the isolation masses for observed exoplanets. Assuming that the planets formed at their observed semimajor axis and that their isolation masses were later merged by giant impacts to yield the observed exoplanets, the isolation masses for close-in planets are between 0.1 and $5M_\oplus$. We also map the corresponding disc surface density Σ_s to an enhancement relative to that of the MMSN, which we denote $\Sigma/\Sigma_{\mathrm{MMSN}}$. For a typical close-in planet with an assembled mass of about $4.5M_\oplus$ at 0.1 AU, M_{iso} has a median value of $0.6M_\oplus$, corresponding to a disc enhanced relative to the MMSN by a factor of roughly $\Sigma/\Sigma_{\mathrm{MMSN}} = 20$ (blue cross).
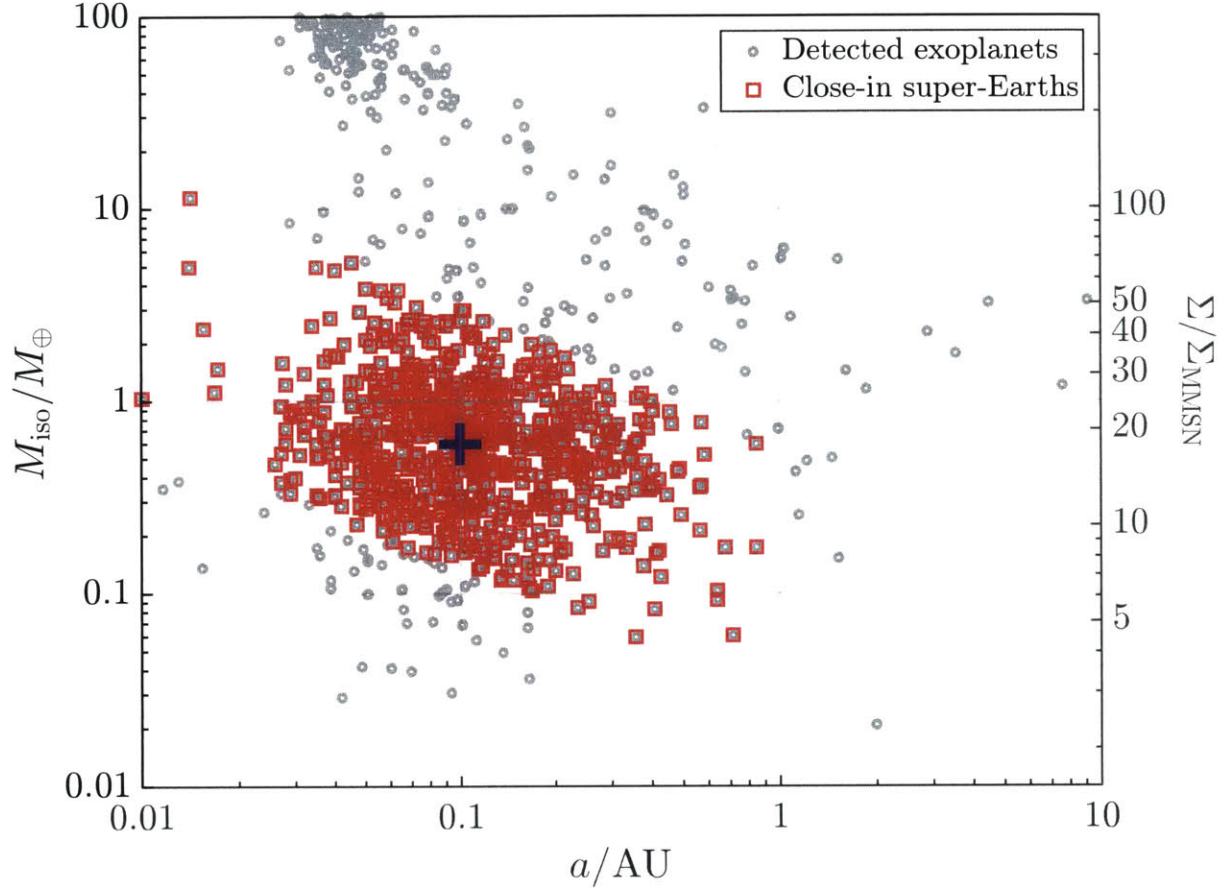
Figure 4-1: Isolation masses M_{iso} for detected exoplanets at their current orbital distances. Isolation masses M_{iso} for all detected exoplanets relative to an Earth mass, $M_{\oplus}$, are shown as grey circles. Close-in super-Earths and mini-Neptunes, exoplanets with semimajor axes less than 1 AU and with masses greater than an Earth mass but less than a Neptune mass, are highlighted as red squares. On the right-hand axis, we map Σ_s for isolation masses to a protoplanetary disc surface density enhancement relative to that of the minimum mass Solar nebula (denoted $\Sigma/\Sigma_{\mathrm{MMSN}}$). For a typical close-in planet with an assembled mass of about $4.5M_{\oplus}$ at 0.1 AU, M_{iso} has a median value of $0.6M_{\oplus}$ (blue cross). Exoplanet data taken from exoplanet.eu database (Schneider, Dedieu, Le Sidaner et al. 2011).

4.2.2 Atmospheric Accretion Model

The envelope mass M_{atm} is given by

$$M_{\text{atm}} = 4\pi \int_{R_{\text{core}}}^{r_{\text{min}}} \rho(r)r^2 dr, \tag{4.3}$$

where r is the distance measured outward from the centre of the core, R_{core} is the radius of the core, and ρ is the density profile of the envelope. The amount of gas that an isolation mass core can accrete also depends on the extent of its Hill or Bondi radius, the smaller of which is denoted r_{min}. The Bondi radius r_B describes the range over which the escape velocity of the core is greater than the sound speed c_s of the gas [Eq. (1.3)]. If G is the gravitational constant, then $r_B = 2GM_{\text{core}}/c_s^2$. The core mass M_{core}^* at which r_{min} transitions from the Bondi radius to the Hill radius is calculated by setting $r_B = r_H$, yielding

$$M_{\text{core}}^* = \left[\frac{a\gamma_a k_B T_{\text{disc}}}{2G\mu m_p \left(3M_\star\right)^{1/3}} \right]^{3/2}. \tag{4.4}$$

Here, T_{disc} is the local temperature of the gas disc, k_B is the Boltzmann constant, μ is the molecular mass of the gas (taken to be 2.34), m_p is the proton mass, and γ_a is the adiabatic index of the gas. Eq. (4.4) implies that if M_{core} is less than M_{core}^*, then r_{min} is given by r_B, and if M_{core} is greater than M_{core}^*, then r_{min} is given by r_H. We assume throughout that

the core radius, $R_{\rm core}$, is related to $M_{\rm core}$ as

$$R_{\rm core} \approx R_\oplus \left(\frac{M_{\rm core}}{M_\oplus} \right)^{1/4},$$ (4.5)

which is consistent with results from both observations (Lopez & Fortney 2014, Lissauer, Fabrycky, Ford et al. 2011) and planetary interior mass-radius models for planets composed primarily of rocky material (Seager, Kuchner, Hier-Majumder et al. 2007). Eq. (4.5) also implies that for some $M_{\rm core}$, $r_{\rm min}$ becomes smaller than $R_{\rm core}$, in which case the isolation mass cannot accrete any envelope. In the inner disc, where $M_{\rm core} < M_{\rm core}^*$, this occurs when

$$M_{\rm core} = \left(\frac{R_\oplus c_s^2}{2 G M_\oplus^{1/4}} \right)^{3/4}.$$ (4.6)

We assume that the surface density of solids in the disc is given by $\Sigma_s = \Sigma_{s,0} \left(a/{\rm AU} \right)^{-3/2}$, while the surface density of gas is given by $\Sigma_g = \Sigma_{g,0} \left(a/{\rm AU} \right)^{-3/2}$ [Eqs. (1.4)]. For the accretion of gaseous envelopes, we also assume throughout this chapter that gas surface densities are enhanced by a factor of 200 relative to the solid disc surface densities so that for a disc enhanced relative to the MMSN by a factor of 20, $\Sigma_{s,0} = 140$ g/cm^2 and $\Sigma_{g,0} = 28,000$ g/cm^2. In our disc model, the temperature $T_{\rm disc}$ is assumed to be uniform at 1500 K out to 0.1 AU, beyond which $T_{\rm disc}$ scales as $a^{-2/3}$ (D'Alessio, Calvet, & Hartmann 2001). The vertical structure of the disc is assumed locally isothermal, so that if Ω is the Keplerian angular velocity, the scale height $H = c_s/\Omega$ and $\rho_{\rm disc} \simeq \Sigma_g/\left(2H \right)$.

141

When calculating atmospheric structure, care must be taken to ensure $r_{min} < H$, since for $r_{min} > H$, the core will no longer accrete spherically. For our disc model, we have verified that for semimajor axes of typical close-in exoplanets (0.03 − 0.3 AU; see, e.g., Fig. 4-1), r_{min} is less than H for the vast majority of isolation masses shown in Fig. 4-1, such that our assumption of spherical accretion is valid. We find that at 0.03 AU, r_{min} becomes comparable to H *only* for isolation masses of $4M_\oplus$ and larger (Fig. 4-2).

Analytically Calculated Adiabatic Envelope Masses

To calculate the structure of the accreted envelope, we use the disc density, pressure, and temperature (ρ_{disc}, p_{disc}, and T_{disc}, respectively) as boundary conditions at r_{min} and integrate the equations of hydrostatic equilibrium inward to R_{core}. The accreted envelope will typically have an outer layer in radiative equilibrium with the disc, while the inner portion of the envelope is in adiabatic equilibrium. In the inner disc ($\lesssim 0.1$ AU), where disc gas densities and opacities are sufficiently high to significantly retard cooling, the radiative layer of the envelope contributes negligibly to the total atmospheric mass, and the envelope can be regarded as essentially adiabatic (Rafikov 2006). In this case, the density, pressure, and temperature of the envelope are given by the adi-
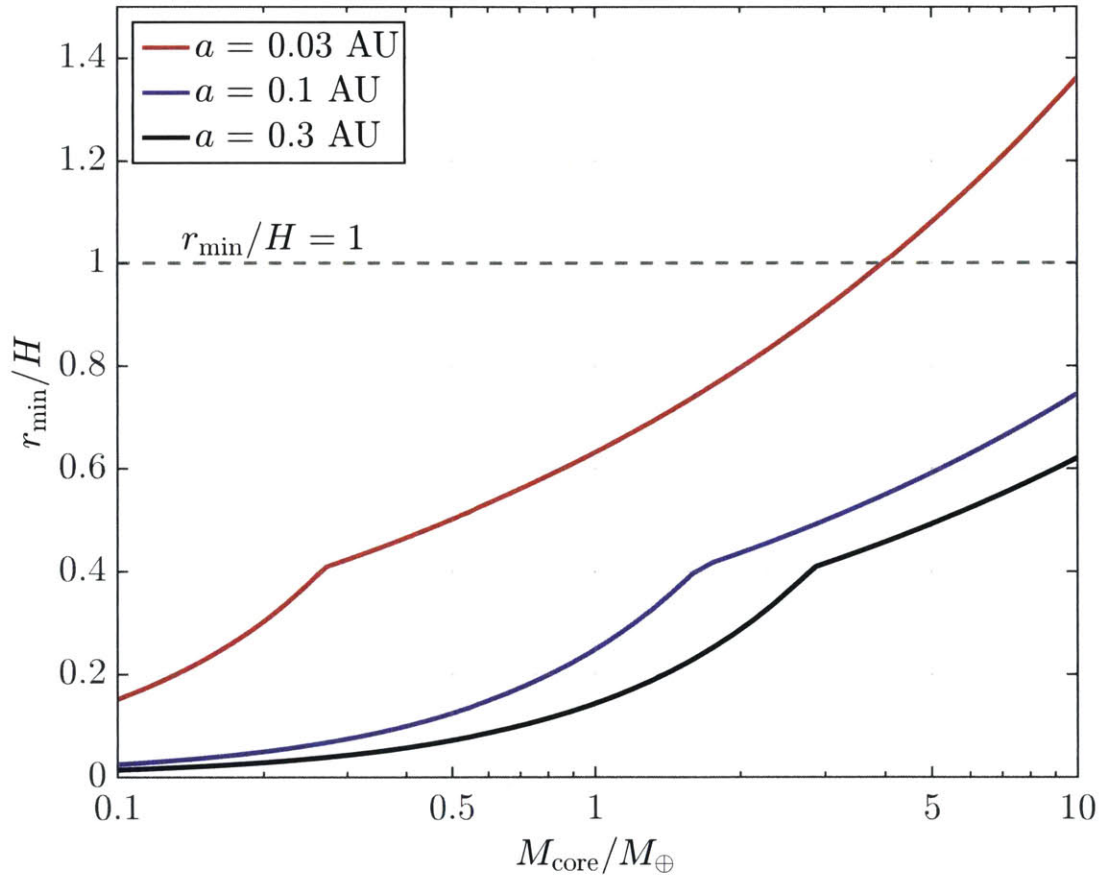
Figure 4-2: Comparison of scale height H and accretion radius $r_{\min}$. When $r_{\min}/H > 1$, accretion is no longer spherical. For our disc model, we find that $r_{\min}$ is comparable to H only for isolation masses greater than $4M_\oplus$ for semimajor axes smaller than 0.03 AU.

abatic equations (see Section 2.4.3)

$$\frac{\rho}{\rho_{\text{disc}}} = \left[\frac{\gamma_a - 1}{\gamma_a}\lambda\left(\frac{r_{\text{min}}}{r} - 1\right) + 1\right]^{1/(\gamma_a - 1)},\qquad(4.7\text{a})$$

$$\frac{p}{p_{\text{disc}}} = \left[\frac{\gamma_a - 1}{\gamma_a}\lambda\left(\frac{r_{\text{min}}}{r} - 1\right) + 1\right]^{\gamma_a/(\gamma_a - 1)},\qquad(4.7\text{b})$$

$$\frac{T}{T_{\text{disc}}} = \frac{\gamma_a - 1}{\gamma_a}\lambda\left(\frac{r_{\text{min}}}{r} - 1\right) + 1,\qquad(4.7\text{c})$$

where r is measured outward from the core centre and where $\lambda = GM_{\text{core}}\mu m_p/\left(r_{\text{min}}k_B T_{\text{disc}}\right)$ is the escape parameter at r_{min}. For an adiabatic envelope, to good accuracy, we can approximate the envelope mass M_{atm} as simply $4\pi\rho_{\text{disc}}r_{\text{min}}^3/3$. When $M_{\text{core}} \lesssim M_{\text{core}}^*$ (so that $r_{\text{min}} = r_B$), the envelope to core mass ratio can therefore be approximated by

$$\frac{M_{\text{atm}}}{M_{\text{core}}} \simeq 1.10 \times 10^{-2} \left(\frac{\Sigma_{g,0}}{28,000 \text{ g/cm}^2}\right)\left(\frac{T_{\text{disc}}}{1500 \text{ K}}\right)^{-7/2} \times$$
$$\left(\frac{a}{0.1 \text{ AU}}\right)^{-3}\left(\frac{M_{\text{core}}}{M_{\oplus}}\right)^2.\qquad(4.8)$$

On the other hand, if $M_{\text{core}} \gtrsim M_{\text{core}}^*$, so that $r_{\text{min}} = r_H$, we have

$$\frac{M_{\text{atm}}}{M_{\text{core}}} \simeq 3.08 \times 10^{-2} \left(\frac{\Sigma_{g,0}}{28,000 \text{ g/cm}^2}\right)\left(\frac{T_{\text{disc}}}{1500 \text{ K}}\right)^{-1/2}.\qquad(4.9)$$

Fig. 4-3 demonstrates the accuracy of Eqs. (4.7) compared to the more rigorous calculations we detail below.

144

Numerically Calculated Radiative-Convective Envelope Masses

Although in the inner disc the accreted envelope may be regarded as adiabatic, in the outer disc or in a partially depleted disc (see Section 4.5), the radiative component of the atmosphere contributes significantly to the total atmospheric mass. When this is the case, we must numerically integrate the equations of hydrostatic equilibrium and flux conservation. Since disc temperatures are sufficiently high that we expect to avoid metallic hydrogen forming at the envelope base, for computational ease, we assume in this chapter an ideal gas equation of state [Eq. (2.5)]. Then, the equations of hydrostatic equilibrium [Eq. (2.4)]

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\frac{GM_{\mathrm{core}}}{r^2}\rho, \tag{4.10}$$

and of energy transport [Eq. (2.9)]

$$\frac{\mathrm{d}T}{\mathrm{d}r} = -\frac{3\kappa_{\mathrm{R}}\rho}{16\sigma_{\mathrm{B}}T^3}\frac{L}{4\pi r^2}, \tag{4.11}$$

allow us to solve for the envelope structure. Again, κ_{R} is the Rosseland mean opacity of the gas in the envelope, σ_{B} is the Stefan-Boltzmann constant, and L is the luminosity in the envelope. When the lapse rate is such that the envelope is locally unstable to convection, according to Eq. (2.11), we replace Eq. (4.11) with $\mathrm{d}\log T/\,\mathrm{d}\log P = (\gamma_a - 1)/\gamma_a$, so that locally, adiabatic equilibrium holds, with equality of the two expressions occurring at the radiative-convective boundary (RCB). Our prescription

of opacities follows the discussion in Section 2.4.1. We note that grain opacity will tend to dominate in the cooler outer envelope, which is also the portion of the envelope in radiative equilibrium with the disc.

Accretion Luminosities

As discussed in Section 2.5, the luminosity $L \equiv L(\text{RCB})$ can arise from several possible sources, including planetesimal accretion and the intrinsic luminosity of the envelope itself due to contraction. To determine accreted envelope masses, we consider here two limiting cases.

Case 1. In the first case, we assume that the luminosity is due to core accretion. That is, in Eq. (2.26), we assume L is given only by $L_{\text{acc}} = GM_{\text{core}}\dot{M}_{\text{core}}/R_{\text{core}}$, where $\dot{M}_{\text{core}}$ is the core mass accretion rate. We take $\dot{M}_{\text{core}} = M_{\text{core}}/\tau_{\text{acc}}$, where, following Eq. (A2) of Rafikov (2006), the accretion timescale τ_{acc} appropriate for an isolation mass is given by $\tau_{\text{acc}} = 30 \text{ kyr} \times \left(M_{\text{core}}/M_\oplus\right)^{1/3} \left(a/0.1 \text{ AU}\right)^3$.[†] Using Eq. (4.5), we find

$$L_{\text{acc}} = 3.94 \times 10^{20} \text{ W} \times \left(\frac{M_{\text{core}}}{M_\oplus}\right)^{17/12} \left(\frac{a}{0.1 \text{ AU}}\right)^{-3}. \qquad (4.12)$$

For a typical close-in planet isolation mass of $0.6M_\oplus$ at 0.1 AU, $L_{\text{acc}} \approx 1.9 \times 10^{20}$ W. We consider this case an upper limit on the accretion luminosity because the accretion timescale is short relative to typical disc lifetimes. Using this limit essentially assumes that the start of planet

---

[†]The accretion timescale τ_{acc} can be calculated directly by taking the inverse of the interaction rate given by Eq. (1.6), assuming the random velocity is on the order of the escape velocity, and that the density profile of solids follows that of the MMSN.

formation is delayed and that accretion proceeds until the gas disc dissipates and giant impacts set in. When $L = L_{\mathrm{acc}}$, the cooling timescale τ_{cool} for the accreted envelope is on the order of $10^7 - 10^8$ yr (Rafikov 2006), while the disc lifetime is on the order of 10^6 yr (e.g., Hillenbrand 2005). Hydrostatic equilibrium is established on the order of the sound crossing time r_{min}/c_s (< 1 yr), so that in this case, we neglect the time evolution of the envelope structure and consider only hydrostatic equilibrium.

To verify our methods, we have compared our results to those reported by Bodenheimer & Lissauer (2014) and Rafikov (2006). Using the Bodenheimer & Lissauer (2014) accretion model, and the disc temperatures and surface densities used for their Runs 2, 1, and 0.5 (corresponding to accretion at semimajor axes of 2 AU, 1 AU, and 0.5 AU by cores of mass $2.15 M_\oplus$, $2.20 M_\oplus$, and $2.20 M_\oplus$, respectively), we find $M_{\mathrm{atm}}/M_{\mathrm{core}} = 0.031$, 0.010, and 0.034, respectively. Bodenheimer & Lissauer (2014) calculate $M_{\mathrm{atm}}/M_{\mathrm{core}} = 0.025$, 0.025, and 0.017, within a factor of $2 - 3$ of our results. It is likely that differences between the two models arise from different prescriptions of r_{min} and different opacity models, as well as the fact that here, for computational ease, we have assumed a simple ideal gas equation of state. We have also checked our calculations against those of Rafikov (2006). Using the disc, opacity, and accretion models described in Rafikov (2006), we find, for example, that for a $1 M_\oplus$ core, $M_{\mathrm{atm}} = 1.02 \times 10^{26}$ g at 0.1 AU assuming a slow accretion timescale ($\tau_{\mathrm{acc}} = 30$ kyr); $M_{\mathrm{atm}} = 7.17 \times 10^{24}$ g assuming a medium accretion timescale at 1 AU ($\tau_{\mathrm{acc}} = 140$ kyr); and $M_{\mathrm{atm}} = 2.50 \times 10^{24}$ g assuming

147

a fast accretion timescale at 10 AU ($\tau_{\mathrm{acc}} = 300$ kyr). All three results are indistinguishable from those presented in Fig. 6 of Rafikov (2006).

Case 2. In the second case, we assume that in Eq. (2.26), $L^* = L_{\mathrm{acc}} = 0$ and that the heat capacity of the core $c_{p,\mathrm{core}} = 0$. Instead, L evolves according to

$$L = \int_{R_{\mathrm{core}}}^{\mathrm{RCB}} 4\pi r^2 \rho \frac{\mathrm{d}S}{\mathrm{d}t}\,\mathrm{d}r. \tag{4.13}$$

In this case, L is dominated by the change in the internal and gravitational energy of the envelope interior to the radiative-convective boundary as the envelope contracts. Following Lee, Chiang, & Ormel (2014), we assume that the envelope is initially adiabatic after which it cools and the radiative-convective boundary evolves inwards. According to our numerical prescription, we solve the equations of hydrostatic equilibrium for a range of L and follow the time evolution of the envelope structure and mass using Eq. (4.13). It is necessary to cut off the envelope evolution and gas accretion at a time commensurate with the disc lifetime. For our pre-giant impact gas accretion calculations, we do this at $\tau_{\mathrm{disc}} \sim 2$ Myr, though we note that our calculations are insensitive to this cutoff time to within a few Myr.

This case corresponds to an upper bound for the atmospheric mass a core can accrete, since it represents the most efficient cooling scenario for the planet. As the envelope cools, the radiative-convective boundary moves inwards and the envelope accretes more gas. The rate of cooling will be regulated by several factors: the depth of the radiative-convective

boundary with respect to r_{min}, the opacity of the radiative layer, and the atmospheric mass. The contribution of an additional luminosity from the core, however, effectively delays this cooling, limiting the additional atmospheric mass that can be added over the lifetime of the gas disc.

4.2.3 Isolation Mass Gas Accretion Results

In Fig. 4-3, we show $M_{\text{atm}}/M_{\text{core}}$ calculated over a range of M_{core} consistent with close-in planet isolation masses, at semimajor axes of 0.03 AU, 0.1 AU, and 0.3 AU (see, e.g., Fig. 4-1). In thick dashed lines, we show atmospheric masses calculated assuming a fully adiabatic envelope using Eqs. (4.7). In thick solid lines, we show atmospheric masses calculated numerically assuming that entropy changes in the envelope are negligible and that $L = L_{\text{acc}}$, where the accretion luminosity L_{acc} is given by Eq. (4.12). In thin dashed lines with circular markers, we show atmospheric masses calculated numerically assuming $L_{\text{acc}} = 0$, with the evolution of L governed instead by Eq. (4.13). For a typical close-in planet isolation mass of $M_{\text{core}} = 0.6 M_{\oplus}$ at 0.1 AU, we find $M_{\text{atm}}/M_{\text{core}} \approx 4 \times 10^{-3}$.

Fig. 4-3 shows that in the inner disc, envelopes are mostly adiabatic with only a small outer radiative layer that contains negligible mass. This is due to the increased gas densities and hence opacities in the inner disc. It is only at larger semimajor axes (e.g., the $L_{\text{acc}} = 0$ case at 0.3 AU in Fig. 4-3) that we see a significant contribution to the total atmospheric mass from the outer radiative layer. Here, for cores that are sufficiently massive ($M_{\text{core}} \gtrsim 0.3 M_{\oplus}$), decreased opacities allow the

149

envelope to cool, contract, and accrete larger atmospheric masses.

For $a = 0.03$ AU and 0.1 AU, atmospheric masses drop off sharply at $M_{\text{core}} = 0.023 M_\oplus$. Here, Eq. (4.6) is satisfied, so that with our model parameters, no atmosphere can be accreted for $M_{\text{core}} \lesssim 0.023 M_\oplus$. The break in the 0.01 AU lines at about $M_{\text{core}} = 0.25 M_\oplus$, in the 0.1 AU lines at $M_{\text{core}} = 1.7 M_\oplus$, and in the 0.3 AU lines at $M_{\text{core}} = 2.9 M_\oplus$ indicate where $r_{\min}$ transitions from r_B to r_H.
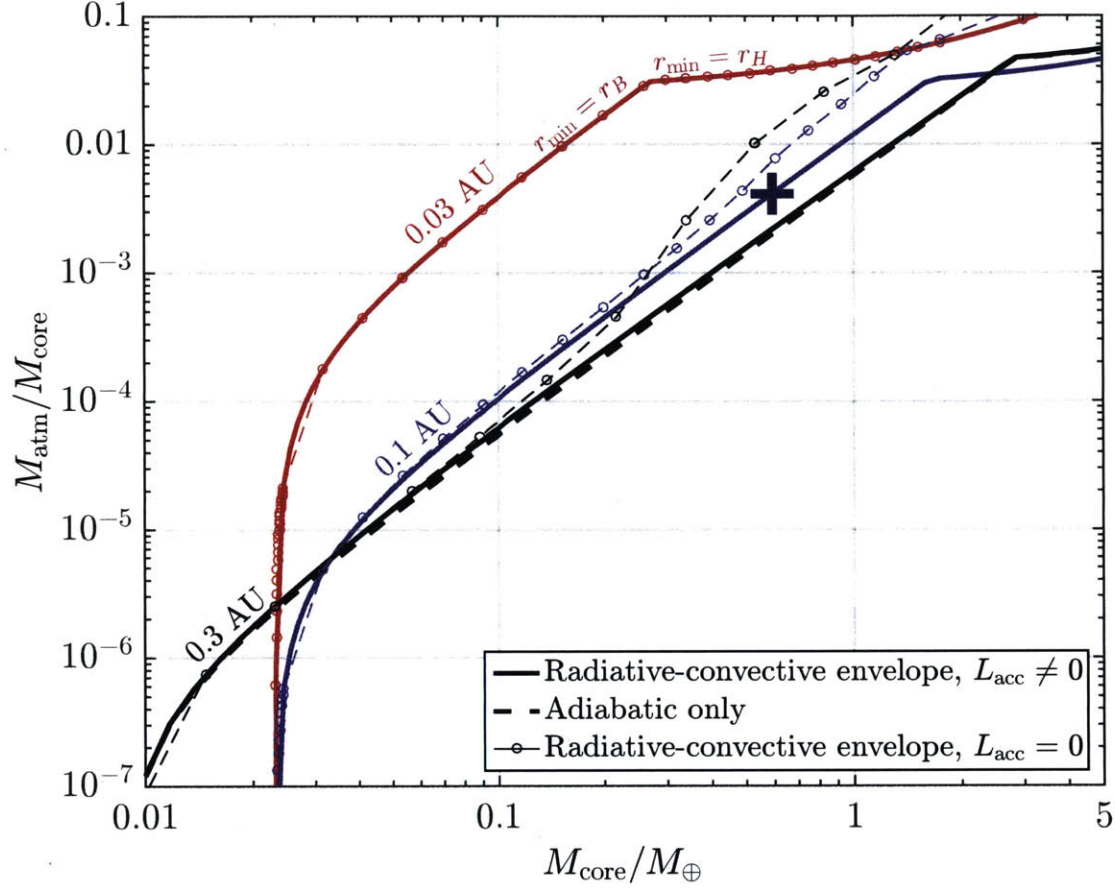
Figure 4-3: Atmospheric masses due to accretion of gas envelopes by isolation masses prior to the phase of giant impacts. Accreted envelope masses are shown for semimajor axes of 0.03 AU, 0.1 AU, and 0.3 AU. Atmospheric masses are calculated assuming a MMSN enhanced by a factor of 20 ($\Sigma_{g,0} = 28,000$ g/cm^2, $\Sigma_{s,0} = 140$ g/cm^2). In thick dashed lines, we show M_{atm}/M_{core} for the case in which we assume that the atmosphere is fully adiabatic [i.e., using the analytic expressions Eqs. (4.7)]. In thick solid lines, we show the results from numerical integrations assuming that the core accretion luminosity L_{acc} is nonzero and given by Eq. (4.12). In thin dashed lines with circular markers, we show M_{atm}/M_{core} for the case in which $L_{acc} = 0$ and L is governed by Eq. (4.13). For an isolation mass and semimajor axis typical of close-in planets ($M_{core} = 0.6 M_{\oplus}$, $a = 0.1$ AU, $\Sigma/\Sigma_{MMSN} = 20$; see Fig. 4-1), $M_{atm} \approx 4 \times 10^{-3} M_{core}$ (blue cross). The difference in atmospheric mass between all the modeled cases is small. This is due to increased densities and hence opacities at smaller semimajor axes, which yields almost adiabatic envelopes and inhibits significant cooling for isolation mass cores.

151

4.3 Giant Impact-Induced Hydrodynamic Escape

In Section 4.2, we calculated the atmospheric structure and mass that an isolation mass accretes in the presence of the full protoplanetary disc. For a typical close-in planet at 0.1 AU, M_{iso} is about $0.6 M_\oplus$ (Fig. 4-1). If a close-in planet forms *in situ*, in order to form a core of $M_{\mathrm{core}} \gtrsim M_\oplus$, a planetary embryo must undergo an additional stage of assembly—giant impacts—in order to achieve observed masses (Fig. 4-4).

Giant impacts between protoplanets, however, can lead to atmospheric mass loss by the planetary embryos. A collision between an impactor and a planetary embryo generates a shock wave that travels through the interior of the target core. This shock wave subsequently initiates ground motion over the surface of the core, which in turn launches a shock into the atmosphere. This can lead to ejection of all or part of the atmosphere (Fig. 3-1). Using the framework developed in Chapter 3, we calculate in Section 4.3.1 the local atmospheric mass loss fraction χ_{local} as a function of local ground velocity, and use these results to calculate the global atmospheric loss fraction χ_{global} (Section 4.3.2). Previous work has focused on atmospheric loss in thin or plane-parallel atmospheres (Genda & Abe 2003, Schlichting, Sari, & Yalinewich 2015). Here we consider the regime applicable to close-in exoplanets, in which the envelope radius is a substantial fraction of the core radius and in which curvature effects become non-negligible.
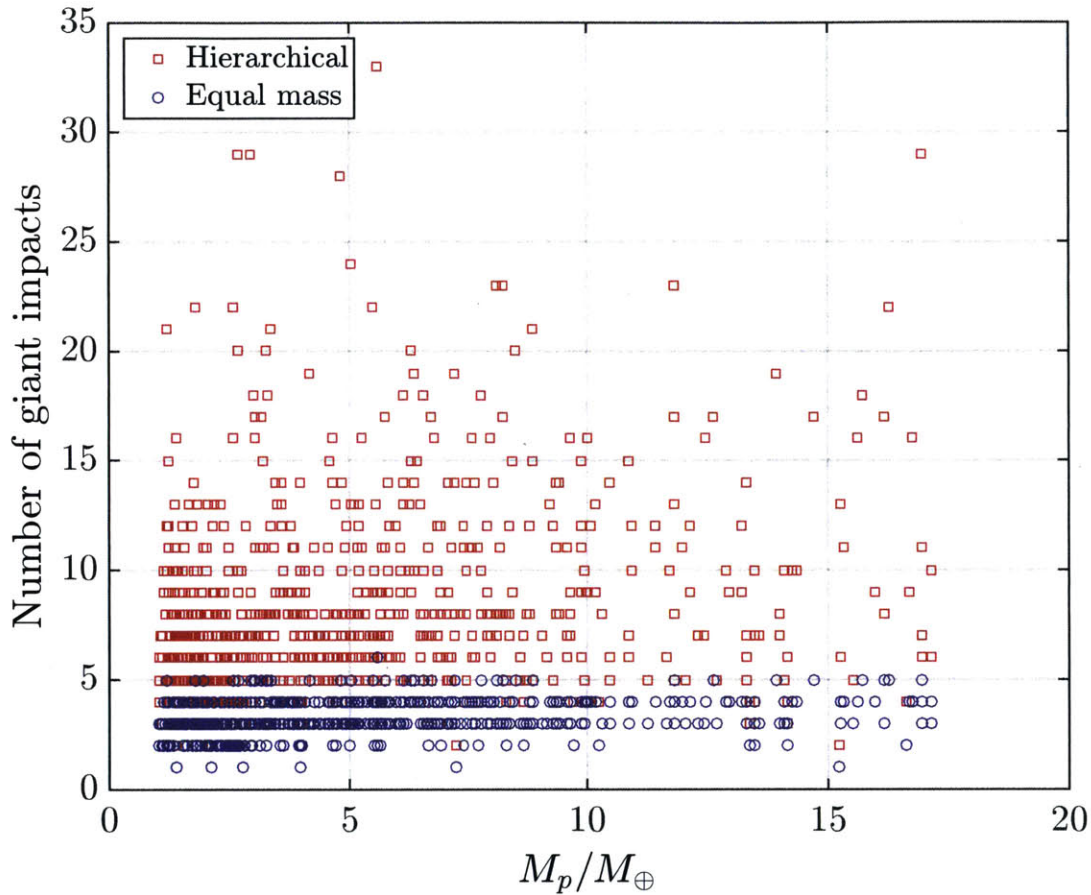
Figure 4-4: The number of giant impacts needed to assemble super-Earths and mini-Neptunes. We show the number of giant impacts needed to assemble observed exoplanets using two limiting scenarios. In the "hierarchical" scenario, assembly to final planet mass M_p occurs only by impactors whose mass is given by the planet's isolation mass (red squares). In the "equal mass" case, planets grow only through collisions between bodies of the same mass (blue circles). A typical planet requires 5-10 giant impacts to be fully assembled. Note that in the hierarchical scenario, the isolation masses for each planet have been calculated using the method given in Section 4.2.

4.3.1 Local Atmospheric Loss

In order to determine the mass loss fraction of an envelope as a function of ground velocity and atmospheric mass, we perform numerical simulations of ground motion-induced shock propagation through an atmosphere. We integrate the Lagrangian equations of motion Eqs. (3.10), assuming (as we did in Chapter 3) that the shock propagates through the envelope adiabatically. Hence, the equation of state is given by Eq. 3.11, and radiative losses are assumed to be negligible (see Section 3.5.2). The criterion for hydrodynamic escape is that the parcel at some $t \gg t_0$ has a velocity greater than its initial radius-dependent escape velocity at $t = t_0$, where t_0 is the time at which the shock is launched into the envelope. The boundary conditions assume a ballistic trajectory for the displaced ground mass [Genda & Abe 2003; see Eq. (3.12)].

In Section 4.2.3, we showed that in the inner-disc regime applicable to close-in planets ($\lesssim 0.1$ AU), the atmospheres can be regarded as adiabatic. For our atmospheric-loss calculations, we therefore assume an adiabatic profile [Eqs. (4.7)] as the initial condition for Eqs. (3.10) and (3.11). The adiabatic index γ_a in the shock equations is not necessarily constant during shock propagation. It can be different from the adiabatic index of the gas when hydrostatic equilibrium is assumed, and in particular, can depend on the dissociation or ionization state of the gas during shock propagation [see, e.g., Section 2.4.2]. However, we find that the effect of dissociation or ionization during propagation is relatively weak, consistent with results from the literature (Genda & Abe 2003, Schlicht-

ing, Sari, & Yalinewich 2015). We therefore take $\gamma_a = 7/5$ as a constant. If we assume an initial adiabatic profile for the atmosphere, subsequent nondimensionalization of the equations of motion shows that the solutions can be parametrized in terms of the envelope-to-core mass ratio $M_{\mathrm{atm}}/M_{\mathrm{core}}$ and the ratio of the ground velocity to the surface escape velocity v_g/v_{esc}.

As outlined in Sec. 3.3.1, we solve Eqs. (3.10) and (3.11) numerically using a finite-difference, staggered grid scheme with artificial viscosity. The radial dependence is discretized over 300 mesh points, and an adaptive time step is implemented to ensure numerical stability. Integrations are carried out over several hundred thousand time steps in order to ensure convergence (see Table 3.1). In Fig. 4-5, we show the local atmospheric mass loss fraction χ_{local} as a function of v_g/v_{esc} for $M_{\mathrm{atm}}/M_{\mathrm{core}}$ ratios spanning $10^{-6} - 10^{-1}$.
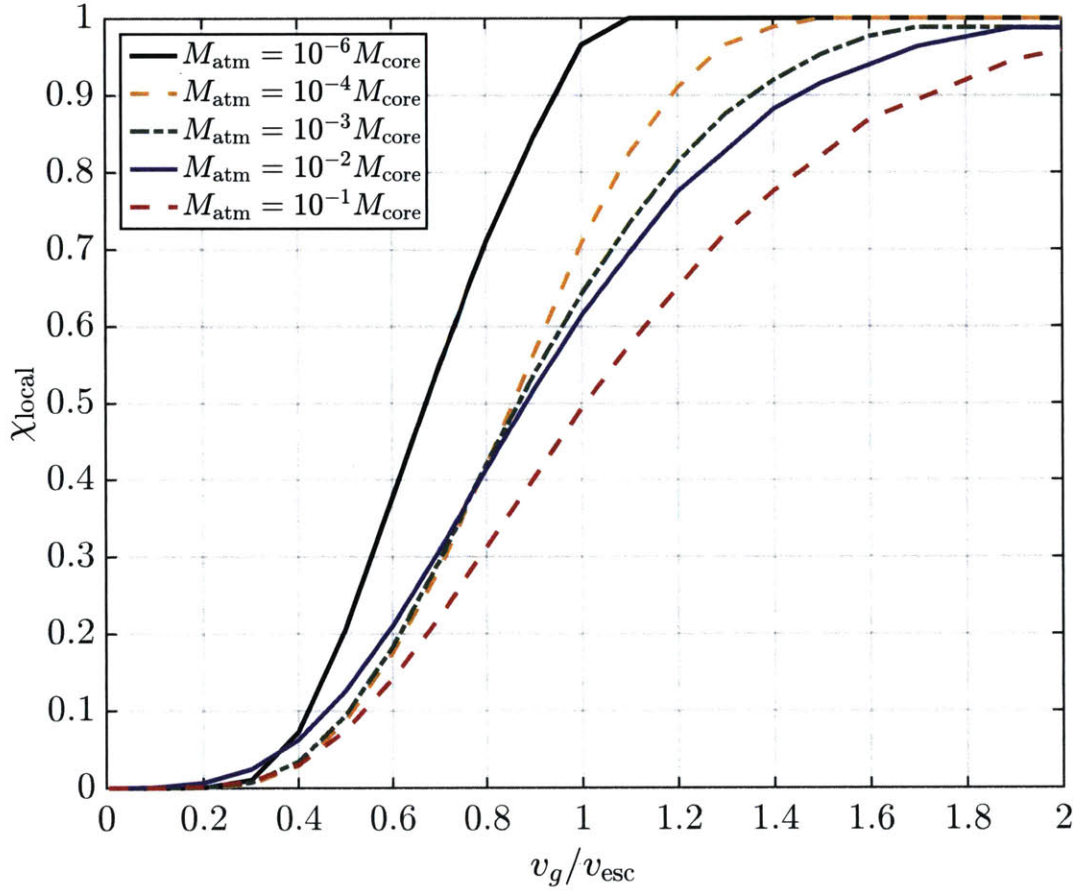
Figure 4-5: Local atmospheric mass loss fraction as a function of ground velocity for different envelope to core mass ratios. We show five cases corresponding to different atmosphere-to-core mass ratios M_{atm}/M_{core}: $M_{atm} = 10^{-6}M_{core}$, $M_{atm} = 10^{-4}M_{core}$, $M_{atm} = 10^{-3}M_{core}$, $M_{atm} = 10^{-2}M_{core}$, and $M_{atm} = 10^{-1}M_{core}$. For the thinnest envelope, $M_{atm} = 10^{-6}M_{core}$, our results agree with the plane-parallel simulations of Schlichting, Sari, & Yalinewich (2015) and Genda & Abe (2003).

156

4.3.2 Global Atmospheric Loss

To calculate the global mass loss due to an impact, we must determine how the ground velocity varies with location on the planetary surface. To do this, we use the framework presented in Section 3.4 that connects impactor momentum [see Eq. (3.23)] with the shock propagated through the core interior to the ground surface. Our model is based on the conservation of impactor momentum as the shock propagates through the core interior. Momentum conservation of the shock is consistent with results from smooth particle hydrodynamic simulations of catastrophic impacts (Love & Ahrens 1996, Benz & Asphaug 1999). In particular, suppose an impactor of mass m_{imp} collides with a protoplanet of mass M_{core} at a speed of v_{imp}. Then if $Q^* \equiv m_{\mathrm{imp}} v_{\mathrm{imp}}^2 / (2 M_{\mathrm{core}})$ is the specific energy of the impactor, a scaling consistent with constant impactor speed and momentum conservation of the shock yields $Q^* \propto R_{\mathrm{core}}$, close to the $Q^* \propto R_{\mathrm{core}}^{1.13}$ reported by Love & Ahrens (1996). (A similar scaling assuming energy conservation of the shock yields $Q^* \propto R_{\mathrm{core}}^2$.)

Referring to Section 3.4, we have that the ground velocity as a function of surface position is given by Eq. (3.26):

$$\frac{v_g}{v_{\mathrm{esc}}} = \left(\frac{v_{\mathrm{imp}}}{v_{\mathrm{esc}}} \right) \left(\frac{m_{\mathrm{imp}}}{M_{\mathrm{core}} + m_{\mathrm{imp}}} \right) \left(\frac{1 - \cos \phi}{2} \right) \left(4 - 3 \sqrt{\frac{1 - \cos \phi}{2}} \right), \quad (4.14)$$

where ϕ is the angle subtended by the surface point from the point of impact. Integrating the local mass loss fraction χ_{local} over the surface

gives the global mass loss fraction χ_{global} [Eq. (3.27)]:

$$\chi_{\text{global}} = \frac{1}{2} \int_0^\pi \chi_{\text{local}}(v_g/v_{\text{esc}}, \phi) \sin \phi \, d\phi. \qquad (4.15)$$

Combining the results in Fig. 4-5 with Eq. (3.27), we show in Fig. 4-6 χ_{global} as a function of the normalized impactor momentum $(v_{\text{imp}}/v_{\text{esc}})[m_{\text{imp}}/(m_{\text{imp}} + M_{\text{core}})]$.
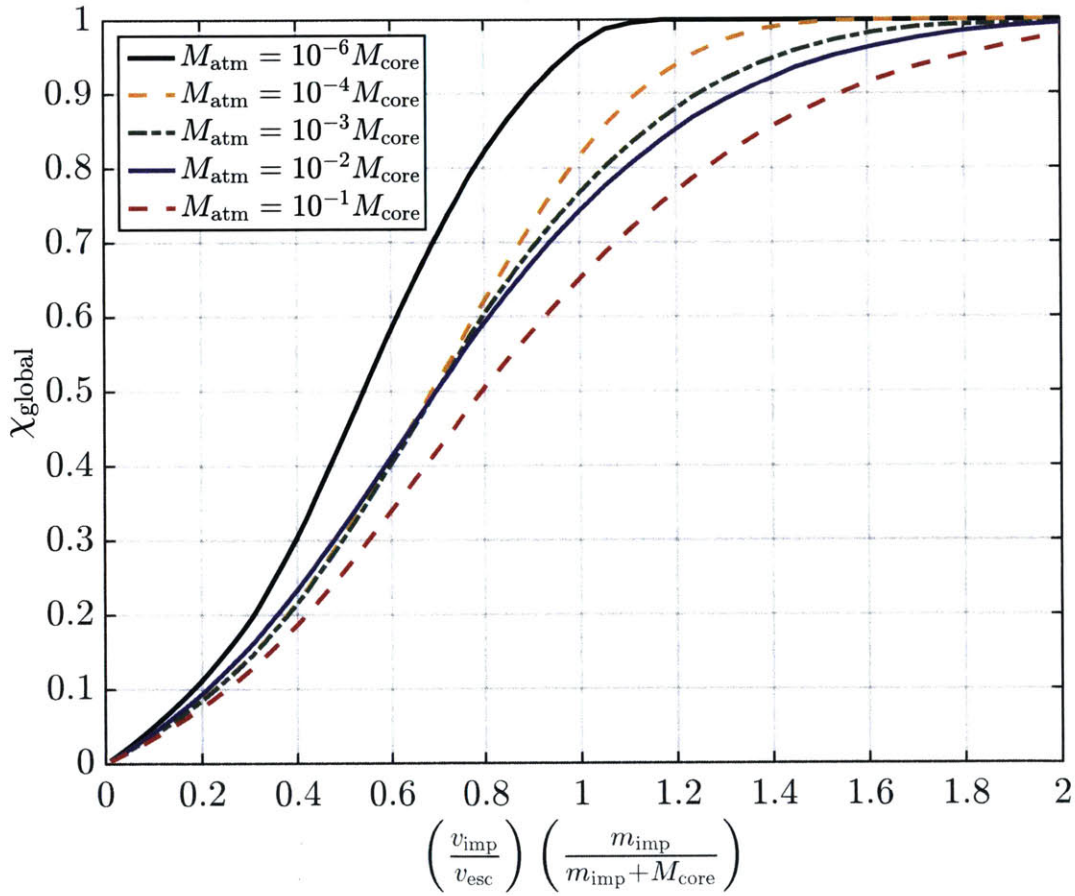
Figure 4-6: Global atmospheric mass loss fraction as a function of $(v_{\text{imp}}/v_{\text{esc}})[m_{\text{imp}}/(m_{\text{imp}} + M_{\text{core}})]$. Each curve represents a solution to Eq. (4.15) for given envelope mass to core mass ratio.

4.4 Atmospheric Erosion Through the Giant Impact Phase

To study the atmospheric mass loss during the giant impact phase, we construct hypothetical collision histories using the results from Section 4.3. Fig. 4-7 shows schematically two impact scenarios that we investigate. On the left side, a series of successive impacts with $0.5 M_\oplus$ embryos takes place, while on the right, a series of impacts occurs between equal-mass impactors. The impact histories that we investigate here represent two extremes: one in which the impactor mass is constant such that $M_{\rm core}/m_{\rm imp}$ always increases, and one in which equal-size impacts occur ($M_{\rm core}/m_{\rm imp} = 1$). The actual impact history is likely to be a combination of these two scenarios. We determine the evolution of the atmosphere-to-core mass ratio $M_{\rm atm}/M_{\rm core}$ with successive impacts. For impacts between disparate mass bodies, we assume a relative random velocity $v_{\rm ran} \sim v_{\rm esc}$, so that the impact velocity $v_{\rm imp} \sim \sqrt{2} v_{\rm esc}$. To calculate the global atmospheric mass loss of the target, we use the results from Section 4.3.2 to determine the atmospheric mass loss fraction for the larger body. For the smaller impactor, we assume that all its atmospheric mass is lost. For impacts between equal-sized bodies, we assume that impacts occur at a relative velocity of $v_{\rm imp} \sim \sqrt{2} v_{\rm esc}$, where v_{esc} is the escape velocity of a single body, and that a shock is launched into each body with a velocity about half the impact velocity, such that $(v_{\rm imp}/v_{\rm esc})[m_{\rm imp}/(m_{\rm imp} + M_{\rm core})] \approx 0.7$. We assume that the re-

sulting planet has an envelope with a mass equal to the sum of the remaining envelopes of the two individual impactors.

In Fig. 4-8, we show an example of the evolution of M_{atm}/M_{core} for each of the two impact scenarios. In this example, we assume that the initial core mass is $0.5M_\oplus$ with the initial atmosphere-to-core mass ratio $\left(M_{atm}/M_{core}\right)_0 = 2 \times 10^{-3}$ (see Fig. 4-3). Each marker indicates a separate impact event. The blue triangles indicate the history in which M_{core}/m_{imp} always increases. The orange squares represent the case in which all collisions are between equal-mass impactors. If no atmosphere is lost throughout the giant impact history, M_{atm}/M_{core} equals $\left(M_{atm}/M_{core}\right)_0$. For a final planet mass of $4.5M_\oplus$ and a collision history in which all impactors are $0.5M_\oplus$, our model yields a final atmosphere-to-core mass ratio of $\sim 10^{-2} \left(M_{atm}/M_{core}\right)_0 \sim 10^{-5}$. The case in which giant impacts occur between equal-mass impactors, on the other hand, yields a final $M_{atm}/M_{core} \sim 10^{-1} \left(M_{atm}/M_{core}\right)_0 \sim 10^{-4}$.

We have applied the collision histories shown in Fig. 4-7 to observed exoplanets. In Fig. 4-9, we show the atmospheric masses that we obtain for observed close-in exoplanets after a phase of giant impacts. For each exoplanet, we take its calculated isolation mass (see, e.g., Fig. 4-1) and calculate the initial atmospheric mass accreted by the isolation mass core given its semimajor axis and the corresponding enhancement in gas density relative to the MMSN (see Fig. 4-1). We then perform Monte Carlo simulations to determine the atmospheric mass loss due to giant impacts.

In Fig. 4-9, we show two cases. In the left panel, we assume all the collisions occur head-on. In this case, for each planet, we conduct 10 trials, with each small marker indicating the mean of 10 trials. In each trial, a core of initial mass M_{iso} successively undergoes a series of giant impacts with impactors either of its own mass ($M_{core}/m_{imp} = 1$) or with a mass equal to the initial isolation mass M_{iso} (M_{core}/m_{imp} increasing). For each collision, we randomly choose which type of impact occurs and assign a probability of $(M_{core}/M_{iso})/(M_{core}/M_{iso} + 1)$ that an impact with an embryo of mass M_{iso} occurs. This probability is chosen such that on average, a planet grows equally in mass by merging with equal-size bodies and smaller ones. For each trial, the core undergoes giant impacts until the final, observed mass is assembled. In the right panel, we allow the collisions to occur over a range of angles, to investigate the effect of off-axis impacts. We approximate this effect by reducing the incident momentum of the impactor by a factor of $\cos\theta$, where θ (sampled uniformly over $[-\pi/2, \pi/2]$) is the impact angle. The procedure here is the same as in the head-on case, but with 100 trials conducted instead, and with the markers shown in Fig. 4-9 representing the mean of those 100 trials.

In the head-on case, the median atmosphere-to-core mass ratio after the phase of giant impacts for observed close-in planets is 8×10^{-4} and the values for the lower and upper quartile range are 1×10^{-4} and 6×10^{-3}, respectively. These values are consistent with terrestrial planet atmospheres and exoplanets that have inferred rocky compositions, but are

161

typically smaller, by an order of magnitude, than atmospheric masses of $1 - 10$ percent inferred from observation for many close-in exoplanets. Accounting for the impact angle allows on average for the retention of roughly 70% more atmospheric mass (median $M_{\text{atm}}/M_{\text{core}} = 1 \times 10^{-3}$, with a lower and upper quartile range of 2×10^{-4} and 1×10^{-2}, respectively). Inferred atmosphere-to-core mass ratios for observed exoplanets are shown in large markers. *In situ* formation of close-in planets via giant impacts typically does not result in atmospheric masses that are $1 - 10$ percent or more of the core mass.
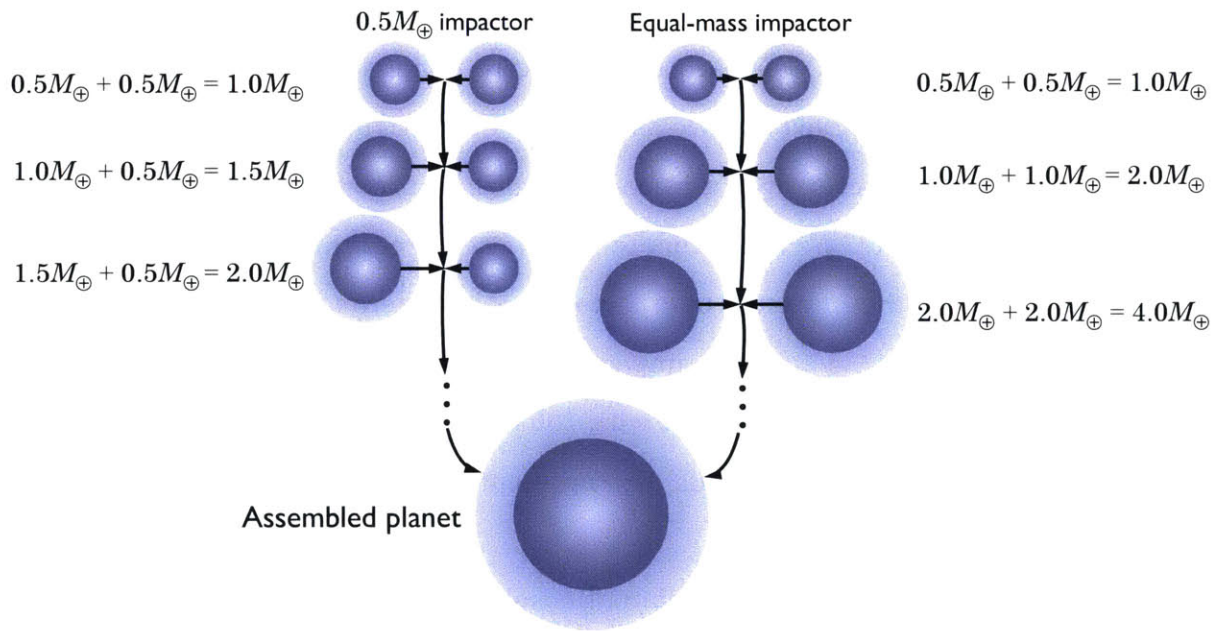
Figure 4-7: Example impact histories. On the left-hand side, each impactor has a mass of $0.5M_\oplus$, and the planet core grows gradually by successive addition of $0.5M_\oplus$ impactors. On the right-hand side, each impact consists of a collision between equal-mass impactors. Each giant impact history represents an extreme formation scenario, with a real giant impact history likely to be some combination of the two.
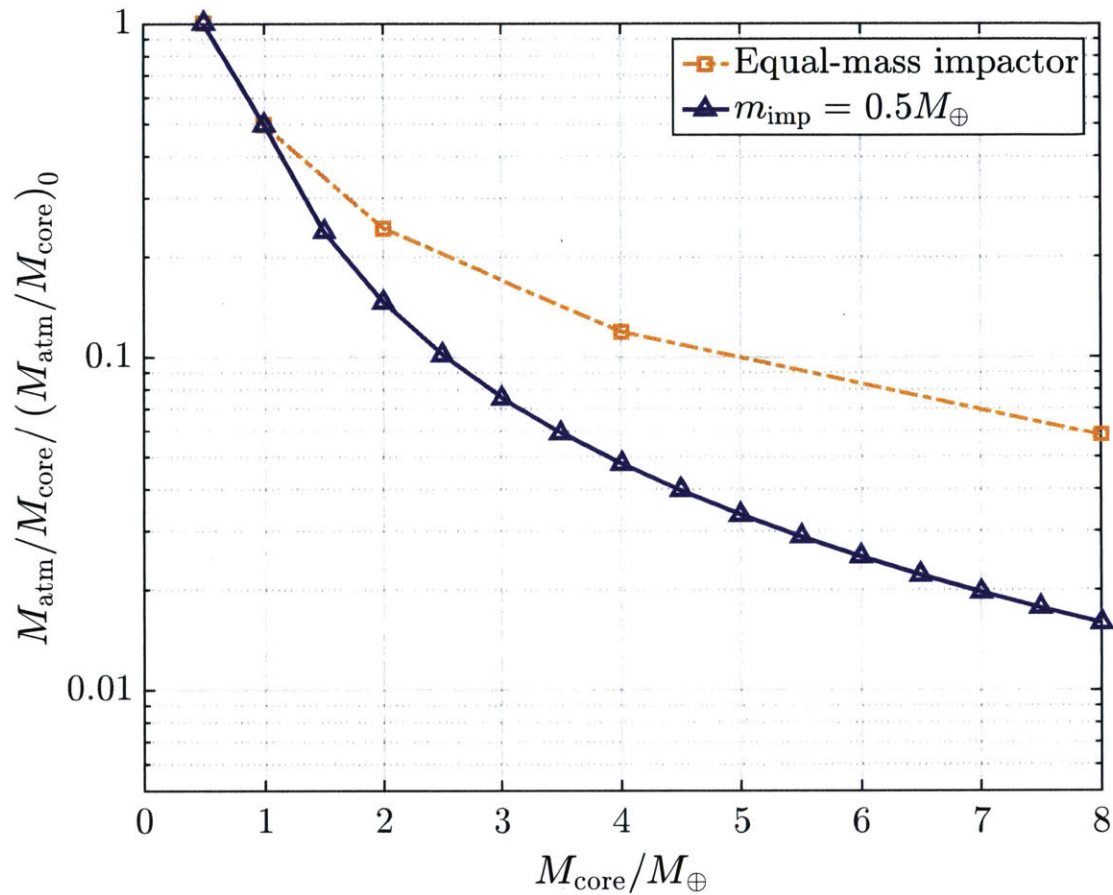
Figure 4-8: Evolution of atmosphere-to-core mass ratio as a function of impact history. Each marker indicates a different collision event. Blue triangles indicate a collision history in which the embryo grows gradually through impacts with $0.5M_\oplus$ impactors (left-hand side of Fig. 4-7). The orange squares are the case in which all impacts occur between equal-mass bodies (right side of Fig. 4-7). Atmosphere-to-core mass ratios are normalized by the initial isolation mass atmosphere-to-core mass ratio, $(M_{\mathrm{atm}}/M_{\mathrm{core}})_0$, which are calculated in Section 4.2. If there is no atmospheric loss throughout the giant impact history, then $M_{\mathrm{atm}}/M_{\mathrm{core}}/(M_{\mathrm{atm}}/M_{\mathrm{core}})_0 = 1$.
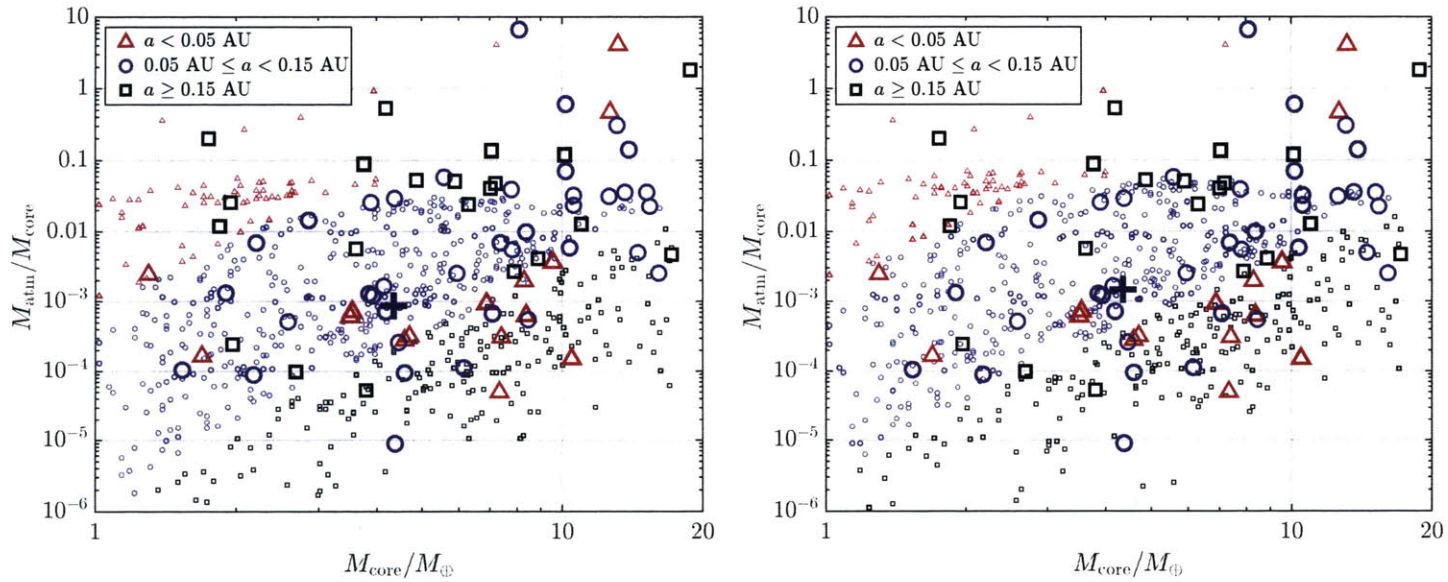
Figure 4-9: Atmosphere-to-core mass ratios for observed close-in exoplanets after a phase of giant impacts. Modeled planets are shown in the figure with small markers, while observed close-in planet atmospheric masses and corresponding semimajor axis ranges are shown with large markers. Observed planet envelope masses are taken from Lopez & Fortney (2014) and the results presented in Chapter 5. Initial atmospheric masses are calculated assuming observed close-in planets initially accreted gas envelopes *in situ* as isolation masses (see, e.g., Fig. 4-1) and were then assembled by giant impacts. In the left panel, we show results assuming head-on impacts, with one marker representing the mean of 10 simulations for each planet. In the right panel, we allow the impact angle to vary in each simulation, with each marker representing the mean of 100 simulations. For head-on impacts, the median atmosphere-to-core mass ratio after giant impacts for a close-in planet is 8×10^{-4} (blue cross, left panel), with a lower and upper quartile range of 1×10^{-4} and 6×10^{-3}, respectively. For angled impacts, the median atmosphere-to-core mass ratio is 1×10^{-3} (blue cross, right panel), with a lower and upper quartile range of 2×10^{-4} and 1×10^{-2}, respectively.

4.5 Post-Giant Impact Accretion

Since atmospheric masses containing $1 - 10$ percent or more of the total planet mass are difficult to achieve as a result of giant impacts, we investigate now the importance of gas accretion after assembly has taken place. In this case, the analysis presented in Section 4.2 still holds with slight modification. In order for giant impacts to proceed, the gas surface density in the disc will have had to decrease, so that $\Sigma_g \sim \Sigma_s$ (Goldreich, Lithwick, & Sari 2004). In the full disc, the excess of gas relative to solids can effectively limit giant impacts from proceeding by damping out the large eccentricities required for them. It is only after a sufficient amount of gas has been dissipated from the disc that giant impacts can proceed. Therefore, in order to calculate the masses of envelopes accreted after giant impacts, we take the gas surface density to be 200 times smaller than before the giant impact phase. For our opacity calculations, we assume that the gas-to-dust ratio is still 200.

In Fig. 4-10, we show $M_{\mathrm{atm}}/M_{\mathrm{core}}$ for gas accretion after giant impacts for two limiting cases. In thick solid lines, we show the case in which $L = L(\mathrm{RCB}) = L_{\mathrm{acc}}$. Here, we assume that the gravitational potential energy resulting from the last mass doubling of the planet by giant impacts is released over the disc dissipation timescale, so that $\dot{M}_{\mathrm{core}} = 0.5 M_{\mathrm{core}}/\tau_{\mathrm{diss}}$, where we take the disc dissipation timescale $\tau_{\mathrm{diss}} \sim 800$ kyr (Hillenbrand 2005). In thin dashed lines with circular markers, we show the case in which $L_{\mathrm{acc}} = 0$ and the evolution of L is governed by Eq. (4.13). In this case, we cut off gas accretion at τ_{diss}.

In contrast to envelope accretion from a full gas disc investigated in Section 4.2, in the case of a dissipating gas disc, the gas densities and opacities are now sufficiently low such that the envelope can cool, contract, and accrete more atmospheric mass over the disc dissipation timescale if $L_{acc} = 0$ and if we assume the core heat capacity $c_{p,core} = 0$. In this case, atmospheres containing several percent of the planets' total mass can be accreted. The $L_{acc} = 0$ case is an upper limit since planetesimal accretion very likely continued after the giant impact stage and since the giant impacts themselves will give rise to significant core luminosity. The magnitude of the core luminosity is highly uncertain since it depends on the viscosity of the core, which is unknown. If the gravitational potential energy resulting from the last mass doubling of the planet is released over the disc dissipation timescale, then the accreted envelope masses are reduced by about an order of magnitude compared to the $L_{acc} = 0$ case (see Fig. 4-10).

While the two limiting cases we explore here show a plausible range of atmospheric masses, atmospheric masses exceeding several percent seem difficult to accrete from a reduced gas disc after giant impacts. Post-giant impact accretion does not seem to be capable of producing atmospheric masses exceeding several percent of the core mass, and can only do so in the limiting case that there is no accretion luminosity and (perhaps more importantly) if the energy contribution of the core is ignored.
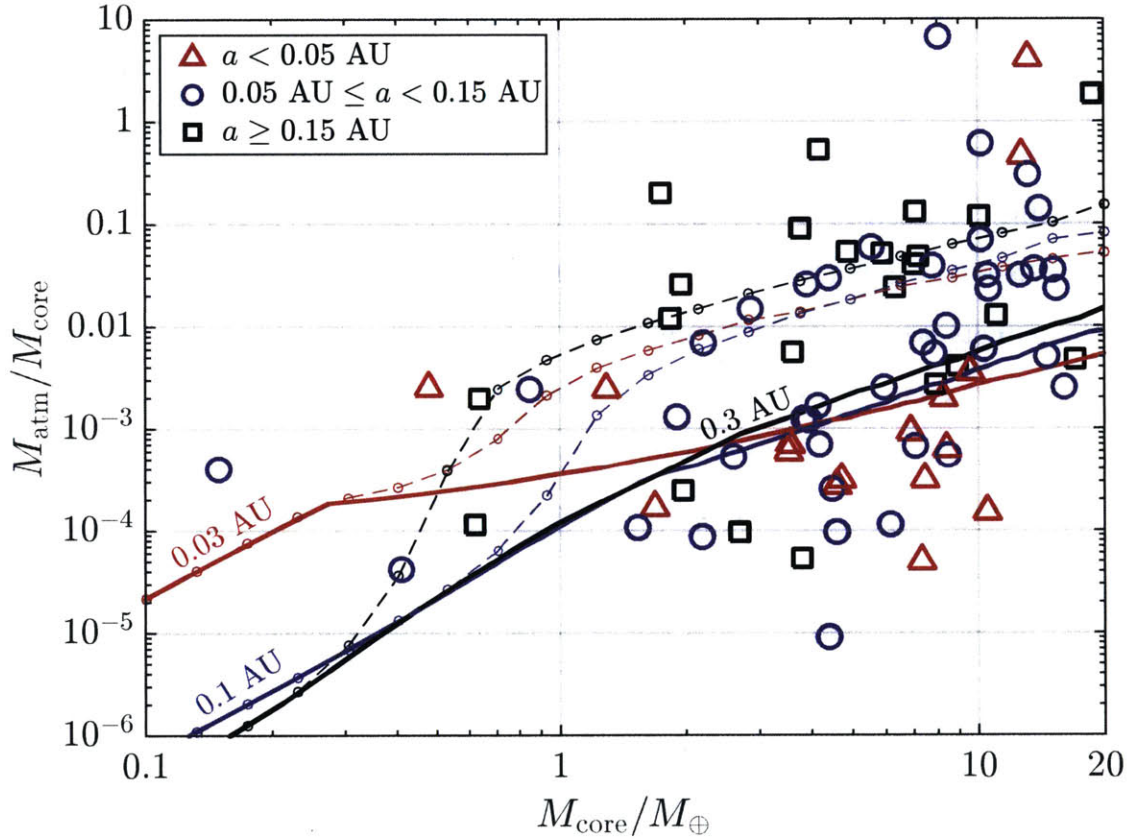
Figure 4-10: Atmospheric masses due to gas envelope accretion by assembled core masses after giant impacts. M_{atm}/M_{core} for observed close-in planets are shown with triangles (semimajor axes less than 0.05 AU), circles (0.05 − 0.15 AU), and squares (semimajor axes greater than 0.15 AU). M_{atm}/M_{core} ratios calculated for a range of core masses are shown at semimajor axes of 0.03 AU (red lines), 0.1 AU (blue lines), and 0.3 AU (black lines) for two limiting cases. The thick solid lines correspond to the case in which $L = L_{acc}$, where the accretion luminosity is due to the gravitational potential energy from a giant impact between two equal mass bodies of $0.5M_{core}$ that is released over 0.8 Myr. The thin dashed lines with circular markers correspond to the case in which $L_{acc} = 0$. M_{atm}/M_{core} for observed exoplanets taken from Lopez & Fortney (2014) and the results presented in Chapter 5.

4.6 Discussion and Conclusions

It has been proposed that close-in super-Earths and mini-Neptunes formed *in situ* either by delivery of $50 - 100 M_\oplus$ of rocky material into the inner disc (Hansen & Murray 2012), or in a disc enhanced relative to the MMSN (Chiang & Laughlin 2013). In both cases, the final assembly of the planets occurs via giant impacts.

In this chapter, we have investigated the atmospheric masses that close-in planets can achieve before, maintain during, and accrete after giant impacts and compared these atmospheric masses with observations. First, we have found that the atmospheres accreted by isolation masses are small. For a typical close-in planet isolation mass of $\sim 0.6 M_\oplus$ at 0.1 AU (Fig. 4-1), the typical atmosphere-to-core mass ratio $M_{\rm atm}/M_{\rm core}$ is about $10^{-3} - 10^{-2}$ (Fig. 4-3). This value is already less than atmospheric masses of up to $1 - 10$ percent inferred for a significant fraction of close-in super-Earths and mini-Neptunes. Additionally, the $M_{\rm atm}/M_{\rm core}$ ratio is reduced further by a factor of $\sim 10^{-2} - 10^{-1}$ due to giant impacts, leading to median atmosphere-to-core mass ratio after giant impacts of 8×10^{-4} with a lower and upper quartile range of 1×10^{-4} and 6×10^{-3}, respectively. Accounting for varying impact angle on average increases these values only by a factor order unity. Such values are consistent with terrestrial planet atmospheres but more than an order of magnitude below atmospheric masses of $1 - 10$ percent of the total planet mass inferred for many close-in exoplanets. Finally, we considered the accretion of gas envelopes by fully assembled cores af-

ter the phase of giant impacts. In this case, we have found that in the best-case scenario where there is no core luminosity from giant impacts, accreted atmospheric masses can only be at most a few percent (Fig. 4-10). If the gravitational potential energy resulting from the last mass doubling of the planet by giant impacts is released over the disc dissipation timescale as core luminosity, then the accreted envelope masses are reduced by about an order of magnitude compared to the $L_{\mathrm{acc}} = 0$ case. We note that the atmospheric masses we have calculated here should be regarded as upper limits since we have ignored other mass loss mechanisms such as photoevaporation (Lopez & Fortney 2013a, Sanz-Forcada, Micela, Ribas et al. 2011), tides (Gu, Lin, & Bodenheimer 2003), and Parker winds (Liu, Hori, Lin et al. 2015).

Another challenge with *in situ* formation is the timescale in the inner disc associated with radial drift of the isolation masses into their host stars due to gas drag. The timescale for radial drift τ_{drag} due to gas drag in the Stokes regime is

$$\tau_{\mathrm{drag}} = \frac{16}{\pi C_D n^2} \left(\frac{M_p}{R_p^2 \Sigma_g} \right) \left(\frac{a\Omega}{c_s} \right)^3 \frac{1}{\Omega},\qquad(4.16)$$

where $C_D \sim 1$ is the drag coefficient. Here, n is the power law index that describes the radial pressure profile in the disc, such that $P_{\mathrm{disc}} \propto a^{-n}$. Using the disc model we detail in Section 4.2.2, we find that $n = 3$ when $a \leq 0.1$ AU and that $n = 10/3$ when $a > 0.1$ AU. In Fig. 4-11, we show τ_{drag} for observed close-in exoplanet isolation masses at their

observed semimajor axes. To calculate τ_{drag}, we use the isolation masses and density enhancements relative to the MMSN shown in Fig. 4-1, while to calculate the radius of the isolation mass, we use Eq. (4.5). Fig. 4-11 demonstrates that a large fraction of isolation masses is expected to drift into their host stars for typical disc lifetimes ranging from $1-10$ Myr (shaded region). Thus, even without accounting for type I migration, this short timescale for inward radial drift due to gas drag presents a major challenge to *in situ* formation.

Lastly, while we have found that it is challenging to form close-in planets with atmospheric masses that are at least several percent of the their total mass *in situ* with giant impacts, it may be possible that such planets reached their fully assembled cores by planetesimal accretion facilitated by radial drift, thereby bypassing the giant impact phase altogether (Lambrechts & Johansen 2012). Recent work, however, suggests that planetesimal accretion results in oligarchic-type growth similar to that which also leads to the formation of isolation masses (Kretke & Levison 2014).

Given the challenges with *in situ* formation of close-in planets with massive atmospheres discussed above, we favor the likelihood that they formed further out in the disc and migrated inwards to their current location. We note that we find super-Earths with envelope mass fractions $M_{\mathrm{atm}}/M_{\mathrm{core}} \lesssim 0.1\%$ can form *in situ*. If this is the case, however, then the population of exoplanets with $M_{\mathrm{atm}}/M_{\mathrm{core}} \lesssim 0.1\%$ and those with $M_{\mathrm{atm}}/M_{\mathrm{core}} \gtrsim 1\%$ would likely have had to form via two different forma-

tion pathways, requiring the close-in exoplanet population to have two distinct origins. On the other hand, the formation of close-in planets at semimajor axes of a few AU bypasses the need for two different formation pathways, and does not require significantly enhanced disc masses compared to the MMSN (Schlichting 2014). In this case, the formation timescale, migration timescale, and disc lifetime are comparable (circumventing the need for extreme fine tuning), and Earth-mass isolation masses can be attained without the need for giant impacts. Whether there are indeed two distinct formation mechanisms or one will require further work, although there are some clues in the occurrence rate data hinting at the former, suggesting different formation mechanisms for close-in exoplanets with large and small radii (Howard, Marcy, Bryson et al. 2012).
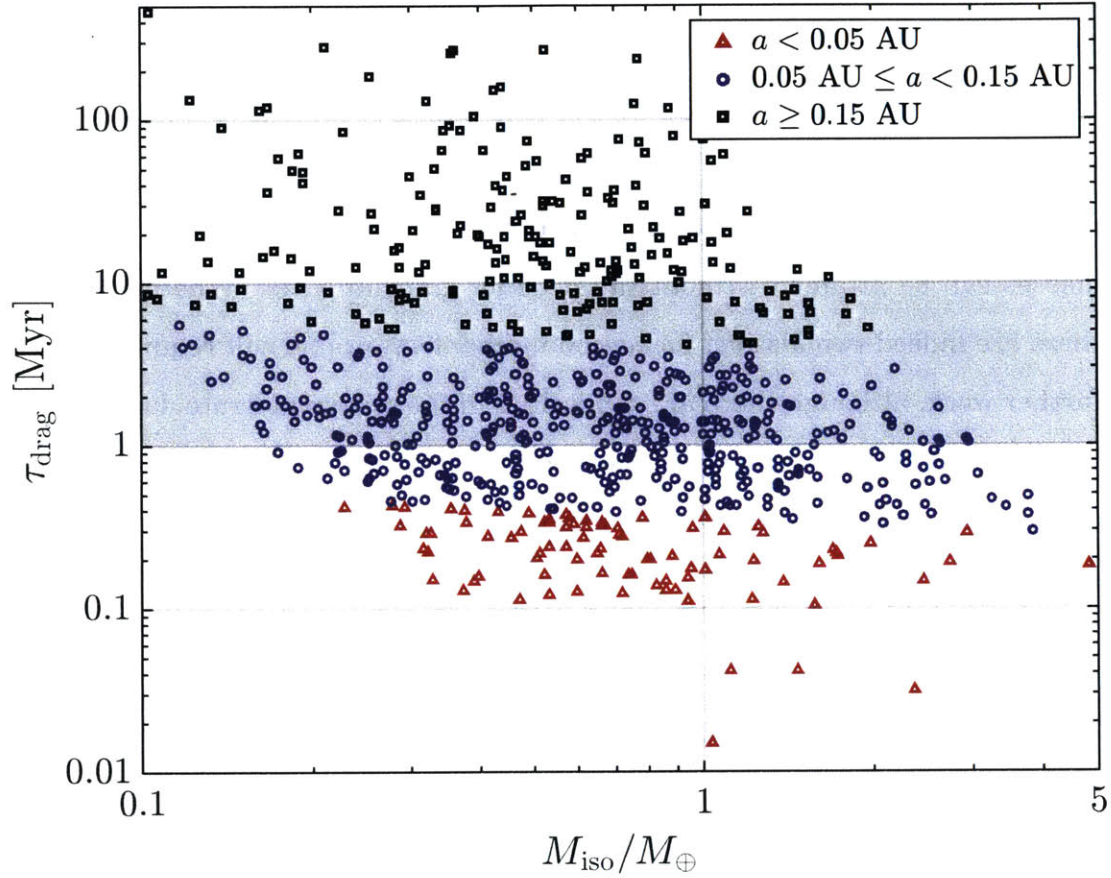
Figure 4-11: Radial drift timescales τ_{drag} due to gas drag for isolation masses calculated from close-in exoplanets (see Fig. 4-1). Markers are colored based on observed semimajor axis. For many systems, the radial drift timescale is short compared to typical disc timescales of $1 - 10$ Myr (shaded region), suggesting that these isolation masses will drift into their host star before giant impacts proceed.

Chapter 5

The Bulk Structural Diversity of Exoplanets[*]

5.1 Introduction

In considering the formation of close-in exoplanets in the presence of the gas disc, we constructed a self-consistent formation history that included core-nucleated gas accretion of envelopes by isolation masses, the assembly of final planet masses (and consequent envelope mass loss) through a phase of giant impacts, and finally, gas accretion from a depleted gas disc. While we found that super-Earths and mini-Neptunes with envelope mass fractions $\lesssim 0.1\%$ could have formed *in situ*, those with more massive envelopes likely had to have formed further out and migrated inwards to their observed semimajor axes. While it may be possible that

---

[*]This chapter is based in part on Inamdar & Schlichting (2016).

173

there are two distinct formation pathways and hence origins for close-in super-Earths and mini-Neptunes, we therefore found it more likely that most or all of them formed further out, at a few AU.

In this chapter, we consider the last stages of planet formation and evolution in the absence of the gas disc. As we have noted, super-Earths and mini-Neptunes are very common, with an occurrence rate per Sun-like star of more than 50%, with the planets possessing up to several percent of their total mass in their gas envelopes (Fressin, Torres, Charbonneau et al. 2013, Petigura, Howard, & Marcy 2013). The latter fact can be inferred either from mass-radius diagrams (Fig. 1-4) or, alternately, mass-density diagrams, in which the mean planet density $\bar{\rho}$ is plotted against planet mass M_p. An important consideration is that in Fig. 1-4, the mass-radius isocomposition curves drawn are cold: they are for degenerate matter and therefore depend little on temperature or incident stellar radiation. For a gaseous envelope in the absence of the gas disc, however, the incident stellar flux F_p is an important parameter, regulating in part the rate at which a gas envelope cools and evolves and (as we show below) in some cases leading to envelope mass loss through photoevaporation. For planets with non-negligible envelopes, "cold" mass-radius curves are therefore no longer applicable.

Fig. 5-1 displays the mean densities for exoplanets with measured masses and which have radii $R_p < 4R_\oplus$, where $R_\oplus$ is the radius of the Earth. The upper panel shows the mean densities as a function of stellar flux, F_p, received by the planets. A large spread in mean densities span-

174

ning about two orders of magnitude is apparent for fluxes of less than about $200 F_\oplus$, where $F_\oplus$ is the flux from our Sun at 1 AU. The deficit of low mean densities for $F_p > 200 F_\oplus$ is well explained by photoevaporation, which can strip significant fractions of the gaseous envelopes of highly irradiated planets (Lopez, Fortney, & Miller 2012, Lopez & Fortney 2013b; see also Section 5.3.2 below). And although it has been clearly shown that photoevaporation can lead to large changes in a planet's mean density and that this may explain the diverse densities in, e.g., the Kepler-11 system (Lopez, Fortney, & Miller 2012), it likely cannot account for the large spread in mean densities of planets with low stellar fluxes for which it should not be important. The lower panel in Fig. 5-1 displays the mean densities as a function of planet mass M_p normalized to an Earth mass $M_\oplus$. It illustrates clearly that super-Earths and mini-Neptunes of a given mass display a large range in mean densities. This is surprising because formation models would naïvely predict a single mass-radius relationship (Inamdar & Schlichting 2015, Lee & Chiang 2015, Ginzburg, Schlichting, & Sari 2015; see Fig. 4-10 in Chapter 4) and one would need to appeal to a diversity in formation environments to account for the large scatter (e.g., Dawson, Chiang, & Lee 2015). This is especially unsatisfactory for planets in multiple systems which display a large diversity in mean density [e.g., Kepler-20 (Fressin, Torres, Rowe et al. 2012), Kepler-36 (Carter, Agol, Chaplin et al. 2012), Kepler-48 (Marcy, Isaacson, Howard et al. 2014, Steffen, Fabrycky, Agol et al. 2013), and Kepler-68 (Gilliland, Marcy, Rowe et al. 2013)].

In this chapter we propose that the large range of observed mean densities may be caused by one or two giant impacts that occurred once the gas disc dissipated. This is in contrast to photoevaporation, which has been extensively investigated in the literature and been found to be incapable of reproducing the observed mean density variation, especially in multiplanet systems (Lopez, Fortney, & Miller 2012).[†] Such giant impacts are expected to be common because they are needed to provide long-term orbital stability of planetary systems and occur typically on timescales between 10 and 100 Myrs (Cossou, Raymond, Hersant et al. 2014). We calculate the planetary radii as a function of mass for ages of 10 to 100 Myrs and use these as input parameters in our one-dimensional hydrodynamical simulations in which we calculate the envelope fraction lost due to giant impacts for initial envelope mass fractions $M_{\mathrm{atm}}/M_{\mathrm{core}}$ of 1-10%. We demonstrate that a single collision between similarly sized exoplanets can easily reduce the envelope-to-core mass ratio by a factor of two. By following the planets' thermal evolution over several Gyrs, we show that this leads to a corresponding increase in mean densities by factors of at least two, and in fact by up to an order of magnitude. We also identify multiplanet systems in the observational data that are likely to have formed by giant impacts. We do this by calcu-

---

[†] It has been recognized that chromospherically active late-M dwarf stars may have significant flare activity in the form of UV rays and protons (Schmidt, Cruz, Bongiorno et al. 2007, Mohanty & Basri 2003), and that such flares may have an important impact on the habitability of close-in "Earth-like" planets (Cohen, Drake, Glocer et al. 2014, Segura, Walkowicz, Meadows et al. 2010). Given the short timescales (tens of minutes to an hour; Segura, Walkowicz, Meadows et al. 2010) and highly variable frequency and directionality of flare activity, they are a possible—although highly uncertain—source of some of the mean density variation for super-Earths and mini-Neptunes.

lating envelope mass fraction for the observed population of exoplanets and determining which systems exhibit the diversity characteristic of stochastic, giant impact-induced structural variety. Since photoevaporation is a variable method for sculpting planetary bulk structure, we construct photoevaporation models to determine parts of the phase space (M_{core} and $M_{\mathrm{atm}}/M_{\mathrm{core}}$) for which photoevaporation does not have an important effect. In this way, we can break the degeneracy between these two methods of envelope devolatization.

The chapter is structured as follows. In Section 5.2.1, we construct our own thermal evolution model to calculate the planet radii for a given core mass as a function of time. We show in Section 5.2.2 that giant impacts can significantly reduce the envelope-to-core mass ratio and demonstrate in Section 5.2.3 that this results in a large increase in a planet's mean density. In Section 5.3, we calculate the envelope mass fraction of observed exoplanets and detail our method for doing so. We discuss how degeneracies between our proposed mechanism for bulk density variation and photoevaporation can be broken in Section 5.3.2 and identify multiplanet systems that are good candidates for having formed by late-stage giant impacts. Our discussions and conclusions follow in Section 5.4.

Liu, Hori, Lin et al. (2015) independently proposed the idea that giant impacts may be responsible for the large diversity in exoplanet densities; their paper investigates the mass loss for two specific giant impacts using three-dimensional hydrodynamic simulations and their pa-

per was posted on the arXiv as we were preparing this manuscript for submission.

5.2 Planet Formation & Late Giant Impacts

Rogers (2015) has shown by modeling the composition of planets with measured radii and masses that the majority of planets larger than $\sim 1.6 R_\oplus$ have significant gaseous envelopes. This implies that these planets likely formed and interacted with the primordial gas disc. This interaction is expected to have resulted in migration and efficient damping of their eccentricities and inclinations leading to densely packed planetary systems regardless of their exact formation location. As the gas disc disappears on timescales of $1 - 10$ Myrs (Mamajek 2009, Haisch, Lada, & Lada 2001), secular excitation in densely packed planetary systems will lead to eccentricity growth culminating in one or two giant impacts producing planetary systems with long-term stability (Cossou, Raymond, Hersant et al. 2014, Deck, Payne, & Holman 2013). A large number of multiple planet systems discovered by *Kepler* may therefore have undergone one or two large collisions after the gas disc disappeared.

5.2.1 Initial Planetary Radii and Thermal Evolution

The radii of planets with significant gaseous envelopes will shrink with time as the planets cool and their envelopes contract. We evolve the con-

178
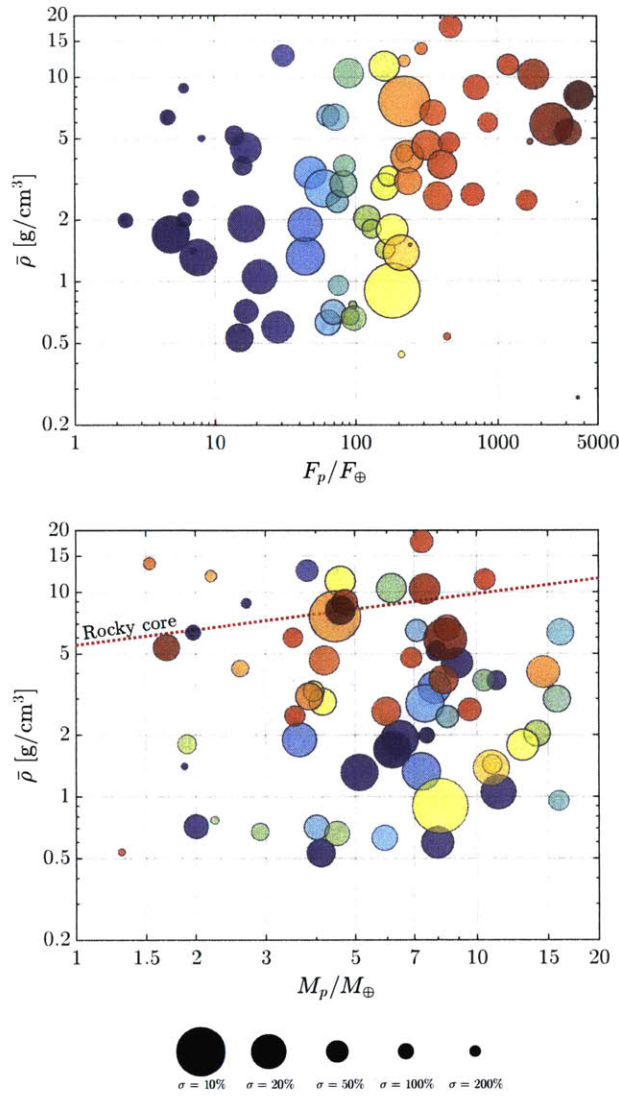
Figure 5-1: Densities of exoplanets with $R_p < 4R_\oplus$. The surface area of each data point is inversely proportional to the 1σ error of the density estimate, such that the most secure density measurements correspond to the largest points. The normalization of the error bars is shown at the bottom of the figure. The colors of the points represent the amount of flux received from the host star. The upper panel shows mean density as a function of flux, F_p, in units of the Earth flux, $F_\oplus$. The lower panel displays exoplanet densities as a function of planet mass in units of Earth masses, $M_\oplus$. Most data are taken from Weiss & Marcy (2014) and references therein. Additional data taken from Jontof-Hutter, Rowe, Lissauer et al. (2015) and Barros, Almenara, Demangeon et al. (2015). For reference, a mean density curve assuming a purely rocky planet (Seager, Kuchner, Hier-Majumder et al. 2007) is shown with a dotted red line.

traction of an envelope of fixed mass M_{atm} about a core with mass M_{core} over time using the method outlined in Chapter 2. In our models, we assume an initial hot start. Typically, for such hot start models, the entropy of the envelope when contraction begins is set at an initial, high value, and the planet is then allowed to cool (Lopez & Fortney 2014, Howe & Burrows 2015). Here we assume the envelope starts off with an arbitrary high intrinsic luminosity, and we solve the equations of hydrostatic equilibrium iteratively to determine the radius corresponding to a given $(M_{\mathrm{atm}}, M_{\mathrm{core}})$ pair. We do the same for a range of luminosities, so that by linking the radius change to a change in internal luminosity and the energy budget of the envelope, we can track the evolution of envelope radius with time via Eq. (2.26). We choose the initial luminosity to be 10^{25} W, although we find the results are generally insensitive to the exact choice of this initial condition for $M_{\mathrm{core}} \gtrsim 2M_{\oplus}$ (Lopez & Fortney 2014; see also Section 5.2.3 below).

For super-Earths and mini-Neptunes with large hydrogen-helium adiabats, the temperature and pressure at the core surface exceed several thousand kelvin and many kilobars, respectively (see, e.g., Figs. 2-9). As discussed in Section 2.5, these values imply that the rocky core should be partially or fully molten, enabling easy heat transfer between the rocky core and the gaseous envelope.

For the heat capacity of the core we assume $c_{p,\mathrm{core}} = 10^3$ J/kg/K (Alfè, Gillan, & Price 2002). We do not include heating from radioactive decay, which we found to have little impact on our results. We set our

outer pressure boundary condition to 20 mbar = 2000 Pa, suitable for the viewing geometry of optical transits (Lopez & Fortney 2014). Our outer temperature boundary condition is set by assuming a range of stellar fluxes from $1 - 5000 F_{\oplus}$. The equations of hydrostatic equilibrium are supplemented with those of energy transport. We assume that when the Schwarzschild instability criterion is satisfied, then energy transport is convective, and that when it is not, energy transport is due to radiative diffusion. In the latter case, energy transport is governed by the local optical depth. Our opacities are determined from OPAL opacity tables (Iglesias & Rogers 1996). We assume a metallicity of $Z = 0.02$ with Hydrogen and Helium mass fractions of $X = 0.80$ and $Y = 0.18$, respectively, yielding a mean molecular mass of 2.3 proton masses. We ignore for the moment mass loss due to photoevaporation. In Fig. 5-2, we show an example of the cooling history of a planet with a $4 M_{\oplus}$ core and different envelope mass fractions. Envelopes with larger masses contract over longer timescales as they have a larger energy budget and hence longer Kelvin-Helmholtz timescales. In Fig. 5-3, we show (assuming an envelope mass fraction of $M_{\mathrm{atm}}/M_{\mathrm{core}} = 1\%$) the effect of varying the incident flux on the planet's cooling track. In general, higher stellar fluxes delay the cooling of the envelope and result in larger radii, due to higher amounts of energy being deposited in the envelope. In Fig. 5-4, we show how cooling tracks vary if we now change the core mass ($2.4 M_{\oplus}$ and $8.5 M_{\oplus}$ in this case) and the envelope mass fraction ($M_{\mathrm{atm}}/M_{\mathrm{core}} = 0.5\%$ and 5%). In these cases, we clearly see there is a

competition between the energy sink provided by the gravitational potential of the core and the energy content of the envelope: more massive cores result in faster cooling, as do less massive envelopes.
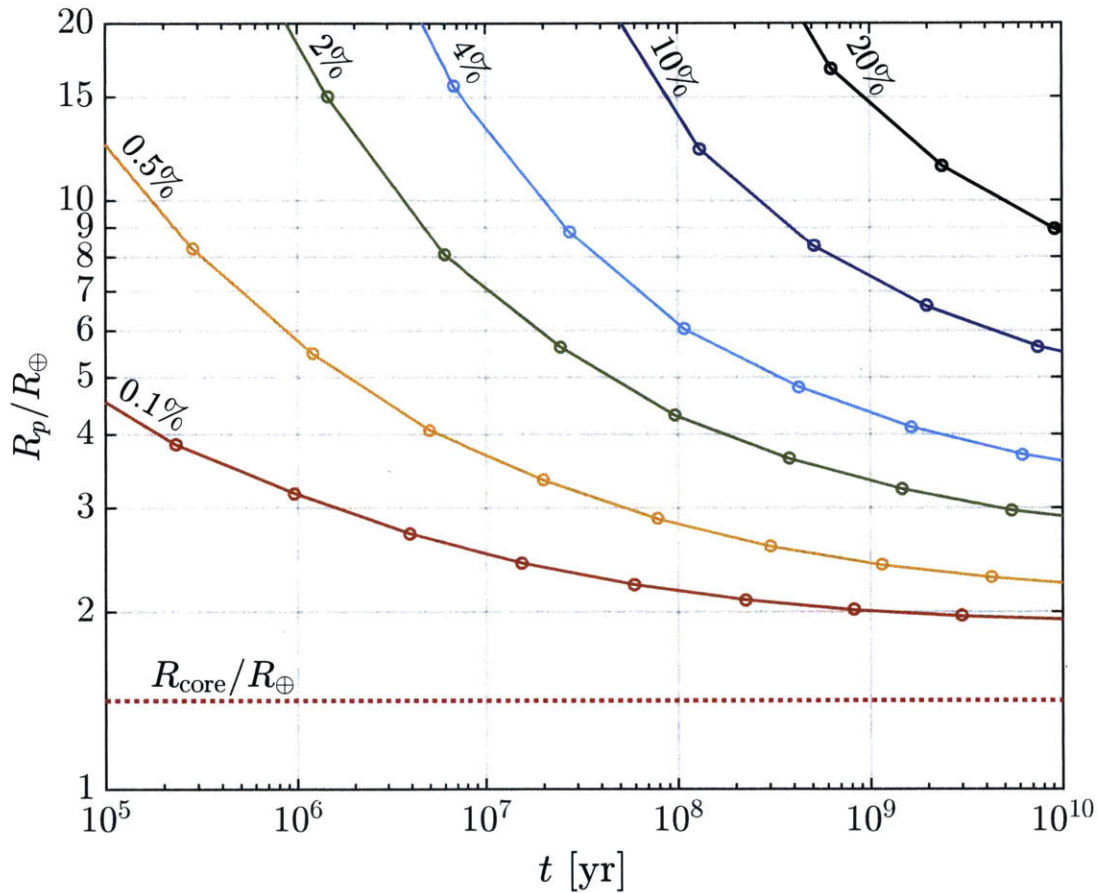
Figure 5-2: Radius evolution as a function of time for a planet with $M_{core} = 4M_\oplus$. The different colored lines correspond to different envelope-to-core mass ratios and the red dotted line to the core radius. We assume that the incident stellar flux $F_p = 100F_\oplus$ and that the core radius scales with the core mass as $R_{core}/R_\oplus = (M_{core}/M_\oplus)^{1/4}$ (Seager, Kuchner, Hier-Majumder et al. 2007).
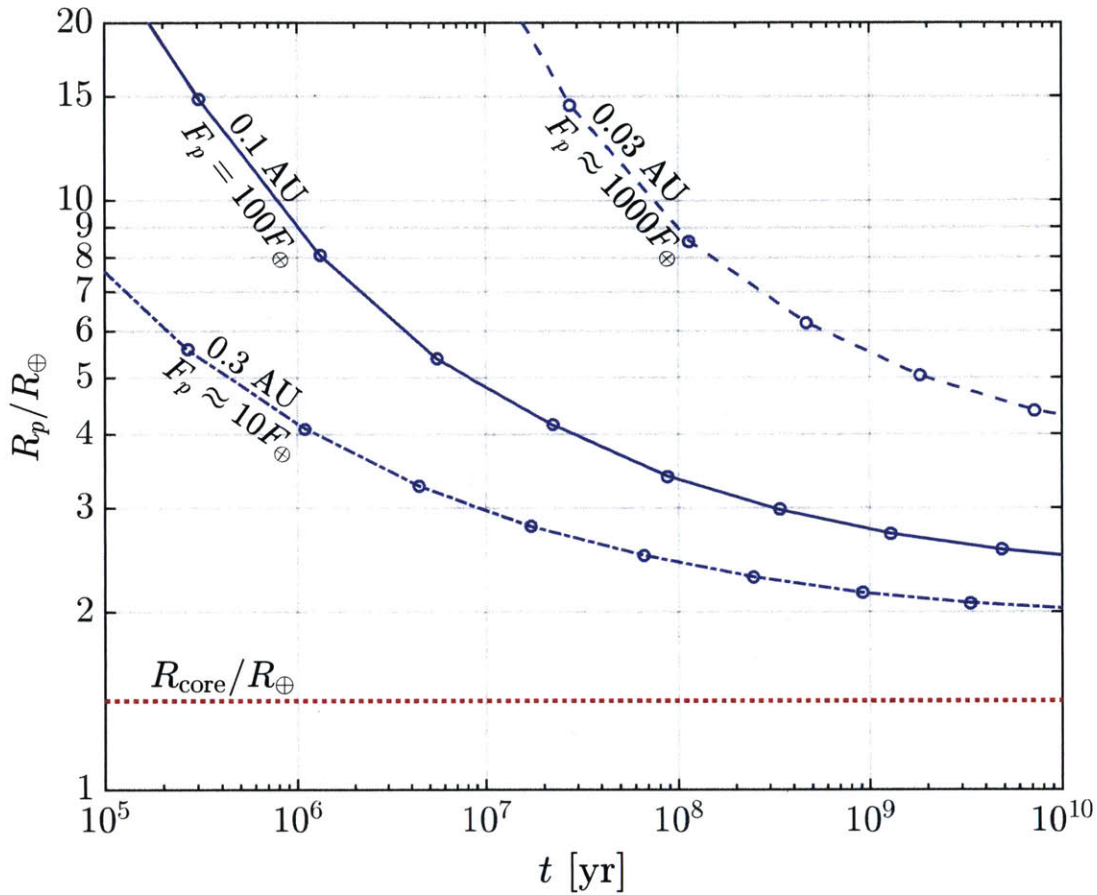
Figure 5-3: The effect of varying incident flux on the thermal evolution of an exoplanet. The planet shown here has a core mass of $4M_\oplus$ and an envelope mass fraction of $M_{\text{atm}}/M_{\text{core}} = 1\%$. Three incident fluxes (or semimajor axes) are shown: 0.03 AU (corresponding to $\approx 1000F_\oplus$ for a Sun-like star; dashed line), 0.1 AU ($100F_\oplus$; solid line), and 0.3 AU ($\approx 10F_\oplus$; dash-dotted line). The radius corresponding to the core is shown by the red dotted line.

183
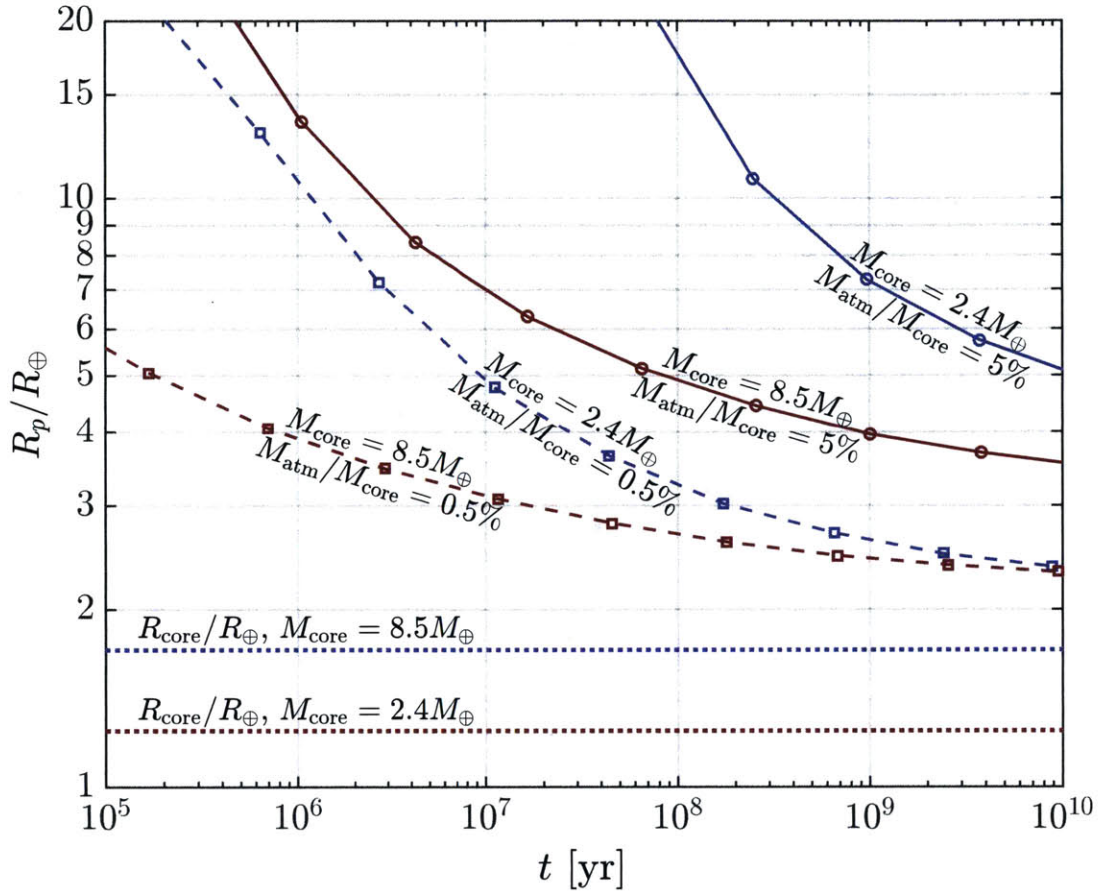
Figure 5-4: The effect of varying core mass or envelope mass fraction on the thermal evolution of an exoplanet. We show two core masses— $M_{\mathrm{core}} = 2.4 M_\oplus$ (red lines) and $8.5 M_\oplus$ (blue lines)—as well as two envelope mass fractions— $M_{\mathrm{atm}}/M_{\mathrm{core}} = 0.5\%$ (dashed lines) and 5% (solid lines). The dotted lines indicate the core radii for each mass.

5.2.2 Envelope Mass-Loss due to a Giant Impact

Using one-dimensional hydrodynamical simulations, we calculate the envelope mass-loss resulting from a giant impact. Our method follows

that outlined in Chapter 3 and applied in Chapter 4: modeling the adiabatic part of the envelope, we track the propagation of a shock launched into the envelope due to local ground motion by solving the hydrodynamic equations with a finite-difference, Lagrangian scheme. Each mass parcel is tracked, and if it reaches velocities greater than its initial, radius-dependent escape velocity from the planet, it is considered lost. We determine the global envelope mass-loss fraction by integrating the local mass loss over the entire surface of the planet where we account for the global distribution of the different ground velocities (Fig. 3-1). The ground velocities are calculated by relating the impactor mass m_{imp} and impact velocity v_{imp} to the resulting ground motion at the various locations of the planet by approximating the impacts as point-like explosions on a sphere and assuming momentum conservation of the impactor (Leinhardt & Stewart 2012). We have compared our envelope mass-loss results with those reported in Stewart, Lock, & Mukhopadhyay (2014b) and Leinhardt & Stewart (2012), who used three-dimensional impact simulations to determine the surface velocity field. For the parameters corresponding to the various Moon-forming scenarios investigated in their work, we find good agreement between their envelope mass-loss results and ours.

Since the impacts are triggered once the gas disc has disappeared, they typically happen when the planetary system is between 10-100 Myrs old. As shown in Figs. 5-2–5-4, the planetary radii at these early times are significantly more extended than at ages of a few Gyrs by when they

185

had time to cool and contract. When estimating the envelope mass-loss fraction due to a planetary collision we therefore determine the loss fraction for a range of planetary radii. We note here that a giant impact can significantly modify the radial profile of the envelope. However, since the envelope profile after thermal evolution over Gyr timescales is generally insensitive to the exact conditions during the first few tens of millions of years, the collision is not expected to leave any significant long-term signatures in the planet's envelope.

Fig. 5-5 shows the resulting mass loss as a function of impactor momentum from our hydrodynamical simulations for envelope mass fractions of 1% and 5%. For the mass loss results shown here, we assume throughout that the core mass is $4M_\oplus$, so that a typical, mass-doubling late-stage giant impact results in a core of mass $8M_\oplus$, roughly the median planet mass seen in Figs. 5-1. The planetary radii are chosen such that they correspond to systems that are 100 Myrs of age (see, e.g., Fig. 5-2). An adiabatic index of $\gamma_a = 1.1$ is used in this simulation because when examining the planet's thermal and density profiles during the accretion and cooling phase, it has been found that due to dissociation, $\gamma_a < 4/3$ (e.g., Lee & Chiang 2015; see also Section 2.4.2). This low value of γ_a is due to the dissociation of hydrogen, which is marginally important for super-Earths and mini-Neptunes. The value of γ_a is of interest because it determines how the mass is distributed inside the gaseous envelope. As we have noted in Section 3.3, for $\gamma_a < 4/3$, the mass in an adiabatic envelope is concentrated towards the core, whereas

for values of $\gamma_a > 4/3$, which applies for diatomic gas with five degrees of freedom ($\gamma_a = 7/5$) and monoatomic gas with three degrees of freedom ($\gamma_a = 5/3$), the mass is concentrated towards the radiative-convective boundary. How the mass distribution in the gaseous envelope affects the global mass loss is shown in Fig. 5-6, which displays the envelope mass-loss fraction for $\gamma_a = 7/5$ and $\gamma_a = 1.1$. For identical planet masses and radii more mass is lost for the $\gamma_a = 1.1$ than the $\gamma_a = 7/5$ case because in the former the mass of the envelope is concentrated towards the core such that the shock that is launched into the envelope from the core can impart a larger momentum onto the envelope. Fig. 5-7 displays the mass loss dependence on the envelope radius. For $\gamma_a = 7/5$ we find that for identical collision parameters less mass is lost for larger envelope radii. This result arises because for $\gamma_a = 7/5$ the envelope mass is concentrated towards the edge of the envelope and larger radii result in lower envelope densities at the core, which in turn implies that the shock travels with a smaller momentum into the envelope. In contrast, for the $\gamma_a = 1.1$ case, the mass loss dependence on envelope radius would be weaker, because most of the envelope mass is concentrated towards the core.

In Figs. 5-5–5-7, we assumed an envelope mass parcel was lost only if it was accelerated to its initial, height-dependent escape velocity. In Fig. 5-8 on the other hand, we show how much the envelope mass-loss could be increased for very close-in planets because of their small Hill radii, per the most general criterion given by Eq. (3.22). For extended

envelopes and small semimajor axes the planet's radius can become comparable to the Hill radius. In the example shown in Fig. 5-8, the planet's radius is chosen such that it is equal to its Hill radius.

Using the results presented in Figs. 5-5–5-8 we can read off the envelope mass-loss fraction χ_{global} for given impact parameters. For an impactor with mass m_{imp} and radius r_{imp} and target core mass M_{core} and radius R_{core} the impact velocity is given by $v_{imp} = \sqrt{v_\infty^2 + v_{esc}^2}$, where $v_{\mathrm{esc}} \equiv \sqrt{2G(M_{\mathrm{core}} + m_{\mathrm{imp}})/(R_{\mathrm{core}} + r_{\mathrm{imp}})}$ is the mutual escape velocity. We therefore find that a collision between comparable mass planets with $v_\infty \sim v_{\mathrm{esc}}$ that about half of the gaseous envelope of both target and impactor is lost. This yields a final planet with a core mass that is about twice the original mass and an envelope-to-core mass ratio that is reduced by a factor of two.
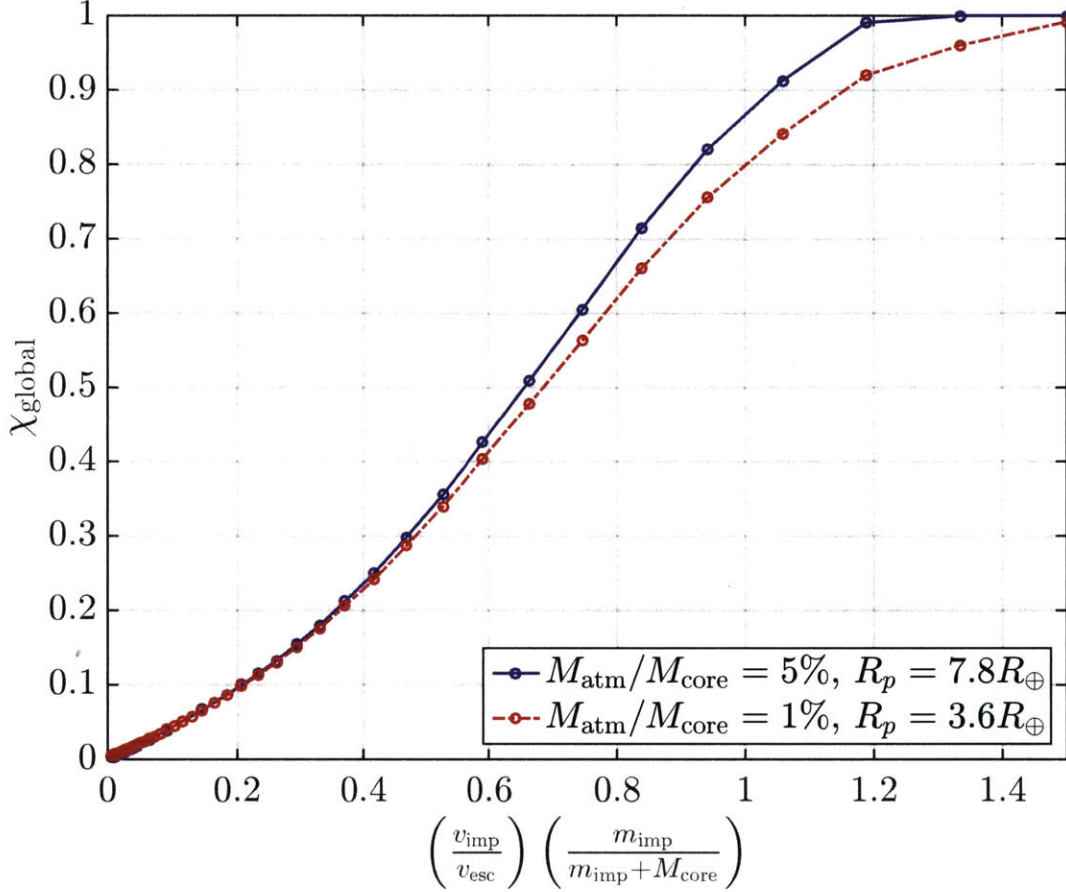
Figure 5-5: Global envelope mass-loss fraction χ_{global} as a function of impactor momentum: effect of varying $M_{\text{atm}}/M_{\text{core}}$. We show here the envelope fraction lost for envelope mass fractions of 1% and 5%, where the planetary radii were chosen such that they correspond to systems that are 100 Myrs of age. A $4M_{\oplus}$ core is assumed in each, although we have found that the envelope mass-loss fraction depends only very weakly on core mass. The adiabatic index of the gas is assumed to be $\gamma_a = 1.1$.
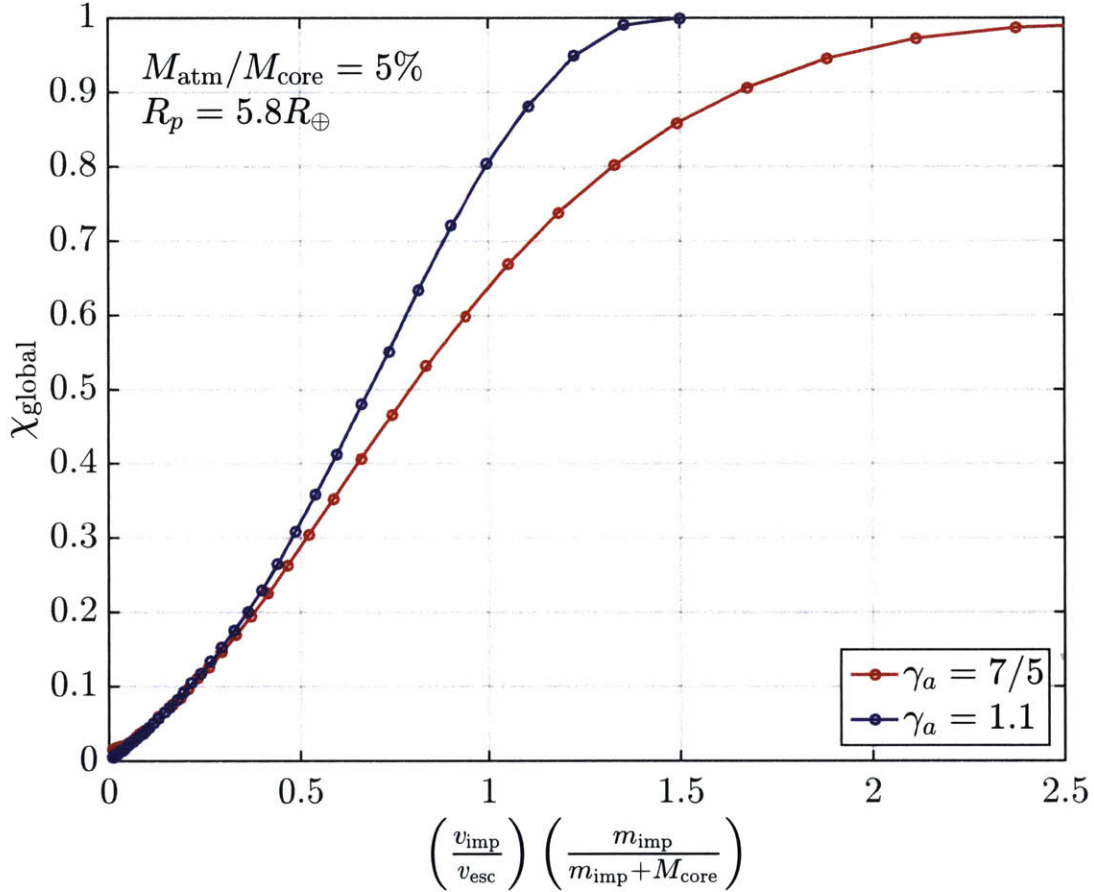
Figure 5-6: Global envelope mass-loss fraction, χ_{global}, as a function of impactor momentum: effect of varying γ_a. The adiabatic index γ_a drives the mass distribution in the envelope and hence the effectiveness by which momentum can be transferred from the core to the envelope. In general, a more compressed envelope ($\gamma_a = 1.1$, for instance, corresponding to a partially dissociated envelope) results in more efficient momentum transfer and thus more efficient envelope mass loss. We assume a $4M_\oplus$ core mass and envelope mass fraction of $M_{\text{atm}}/M_{\text{core}} = 5\%$.
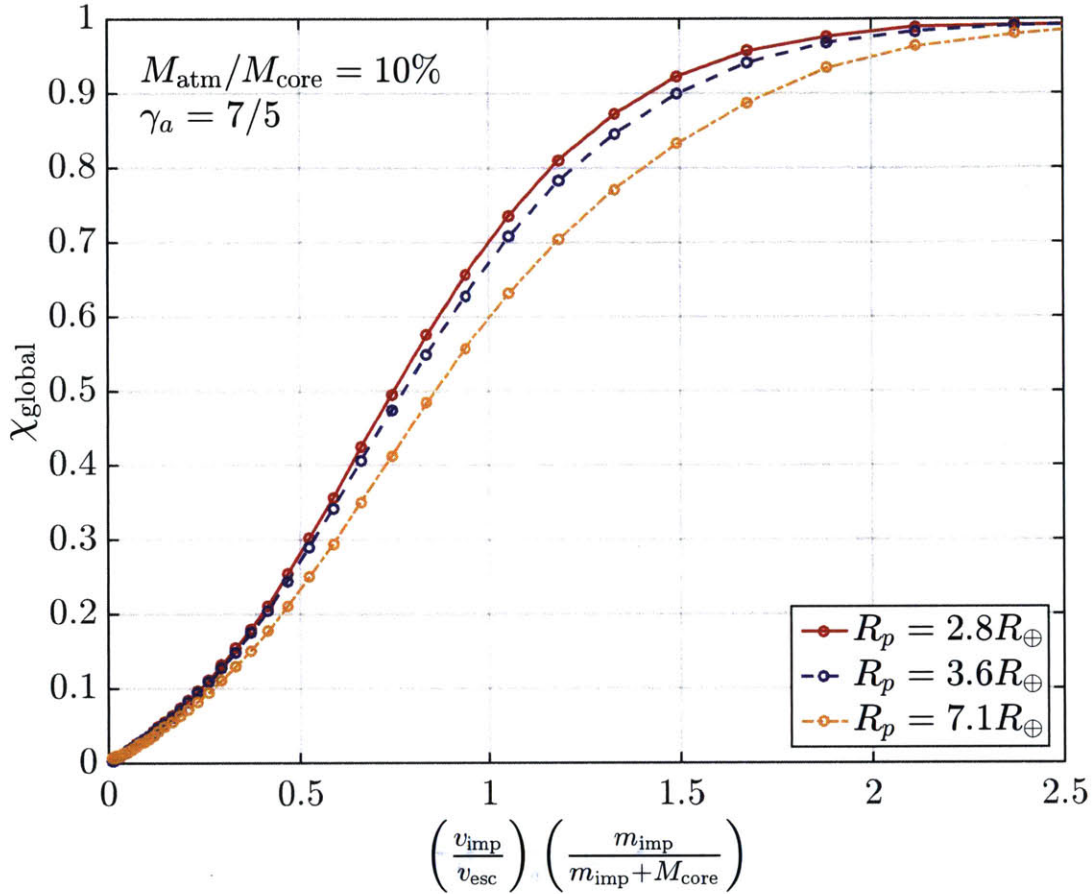
Figure 5-7: Global envelope mass-loss fraction, χ_{global}, as a function of impactor momentum: effect of varying R_p. Varying the planet radius while fixing the adiabatic index and envelope mass fraction has a similar effect to varying the adiabatic index in that a less extended envelope experiences greater envelope mass loss due to more efficient momentum transfer from the core to the envelope. We assume a $4M_{\oplus}$ core mass, an envelope mass fraction of $M_{\mathrm{atm}}/M_{\mathrm{core}} = 10\%$, and an adiabatic index of $7/5$.
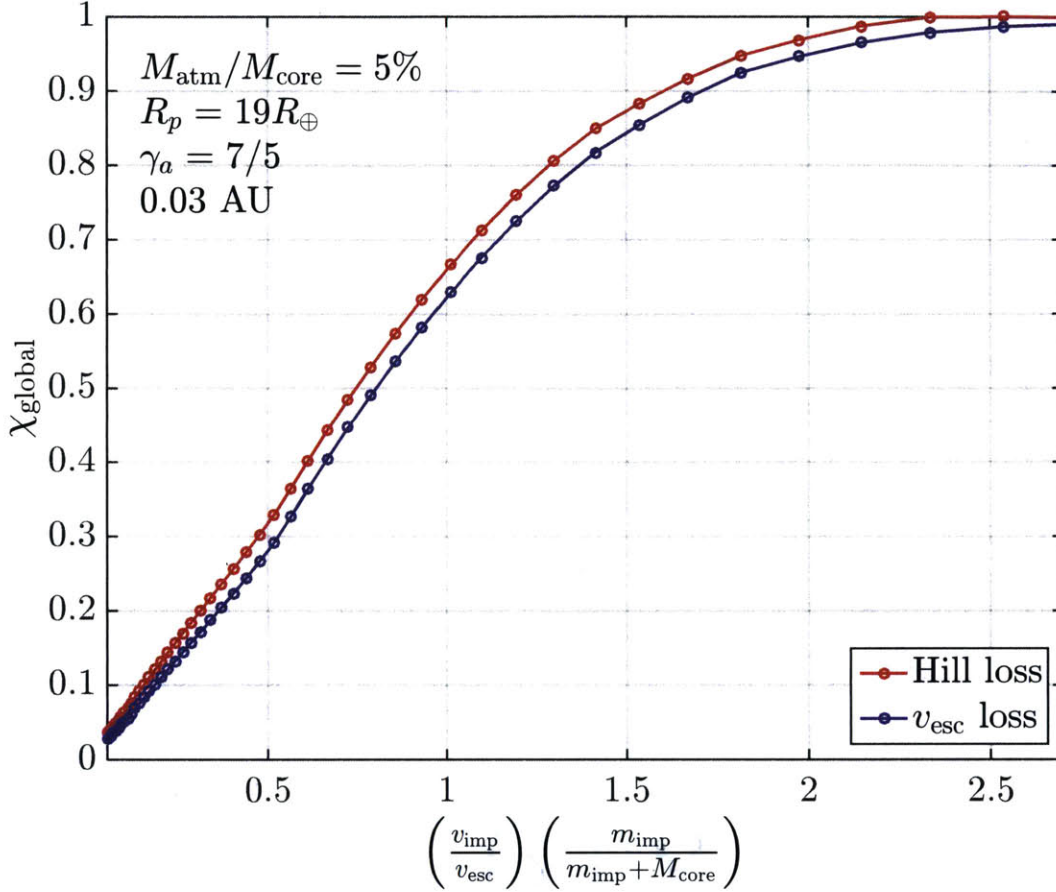
Figure 5-8: Global envelope mass-loss fraction, χ_{global}, as a function of impactor momentum: Hill loss vs. v_{esc} loss. We show here the enhanced atmospheric loss for planets that are very close to their host star due to their small Hill radii (see Section 5.2.2 for details). The planet radius here has been chosen so that it is equal to its Hill radius. We assume $4M_{\oplus}$ core mass and envelope mass fraction of $M_{\text{atm}}/M_{\text{core}} = 5\%$.

5.2.3 Relating Envelope Mass Fractions to Mean Densities

To relate this reduction in $M_{\mathrm{atm}}/M_{\mathrm{core}}$ to observed mean densities we obtain the mean density for planets of various masses for different $M_{\mathrm{atm}}/M_{\mathrm{core}}$. We do this by calculating the contraction of the planet's radius as a function of time and then obtaining for a given planet mass and $M_{\mathrm{atm}}/M_{\mathrm{core}}$ the corresponding radius at an age of 4 Gyr (see Section 5.2.1 for details).

Figs. 5-9 and 5-10 show the resulting exoplanet densities as a function of planet mass and $M_{\mathrm{atm}}/M_{\mathrm{core}}$ from our thermal evolution models. In Fig. 5-10, we take the planets shown in Figs. 5-1 and place them into three flux bins, spaced equally in logarithmic flux space. The bins are centered on $F_p = 10F_\oplus$, $100F_\oplus$, and $1000F_\oplus$. The $1000F_\oplus$ case shows clear evidence for photoevaporative mass loss, which we explore further in Section 5.3.2 and Fig. 5-14 below. On the other hand, the $10F_\oplus$ and $100F_\oplus$ cases show a significant variety in mean density, beyond that which might be inferred from the narrow mass-radius or mass-density results shown in Fig. 4-10.

To demonstrate the effect of a giant impact on $\bar{\rho}$, consider the following scenario. Suppose we have a planet of mass $4M_\oplus$, with an envelope mass fraction $M_{\mathrm{atm}}/M_{\mathrm{core}} = 5\%$ (corresponding to the optimistic, maximum cooling case shown in Fig. 4-10) at a semimajor axis $a = 0.1$ AU. The density in the absence of a giant impact would be about 0.3 g/cm^3. Now suppose the planet undergoes a mass-doubling giant impact that increases the core mass to $8M_\oplus$, but which ejects half the planet's en-

velope mass. The increase in core mass and decrease in envelope mass leads to an increase in the mean density to $\bar{\rho} \approx 2$ g/cm^3—an increase of nearly an order of magnitude. A similar collision, but starting with an envelope mass fraction $M_{\mathrm{atm}}/M_{\mathrm{core}} = 1\%$, results in a density change from about 1.5 g/cm^3 to 3.5 g/cm^3, an increase by a factor of about 2, demonstrating that giant impacts can indeed lead to a broad spread in mean densities. The mean densities we find after 4 Gyr of cooling agree with those of Lopez & Fortney (2014) typically to within about 10-25% for core masses $\gtrsim 2M_\oplus$. At lower core masses $\lesssim 2M_\oplus$, we find, similar to Howe & Burrows (2015), somewhat larger planetary radii than reported by Lopez & Fortney (2014). Specifically, we find radii that are up to 50% larger than those calculated by Lopez & Fortney (2014) and which result in lower mean densities for small planets. We suspect that these discrepancies are likely due to the different initial conditions used [fixed luminosity in this work, fixed entropy in Lopez & Fortney (2014), and fixed radii in Howe & Burrows (2015)] all of which have the greatest impact on the thermal evolution of low mass planets $\lesssim 2M_\oplus$. We note that our density results for Neptune-mass planets with low mean densities suggest that none of the planets possesses more than $\sim 20\%$ of its mass in its gaseous envelopes. This is consistent with results from calculations investigating core-nucleated runaway gas accretion (Rafikov 2006, Piso & Youdin 2014): since planets with more massive envelopes are expected to become gas giants, we do not expect any sub-Neptune planets to have more massive envelopes.
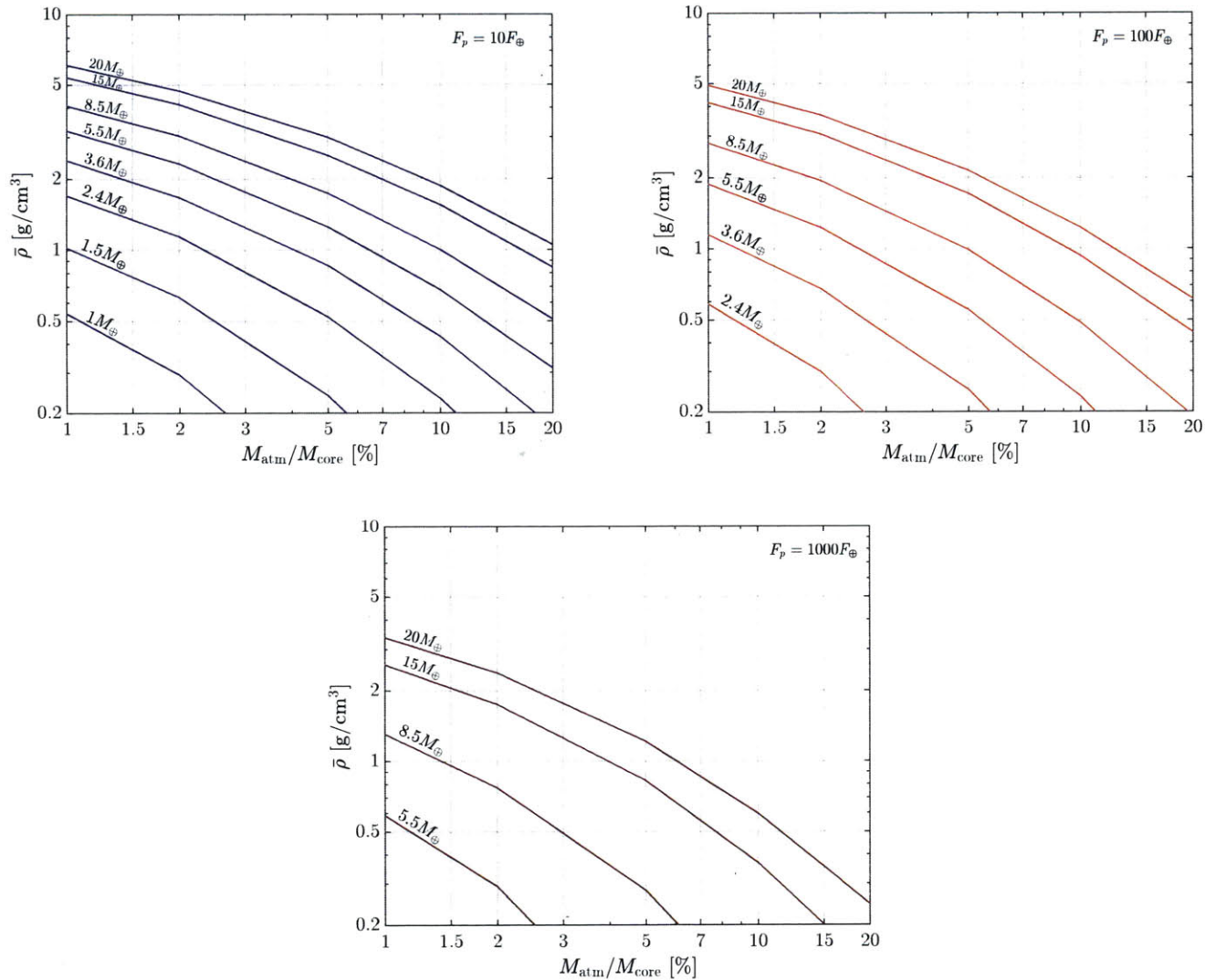
Figure 5-9: Exoplanet mean densities, $\bar{\rho}$, as a function of envelope-to-core mass fraction, M_{atm}/M_{core} for various core masses M_{core}. In each panel, we show results from our thermal evolution models for different incident stellar fluxes: $10F_{\oplus}$ (upper left panel), $100F_{\oplus}$ (upper right panel), and $1000F_{\oplus}$ (lower panel). Each line corresponds to the planet density of the indicated core mass. In each case, the planets have thermally evolved for 4 Gyr.
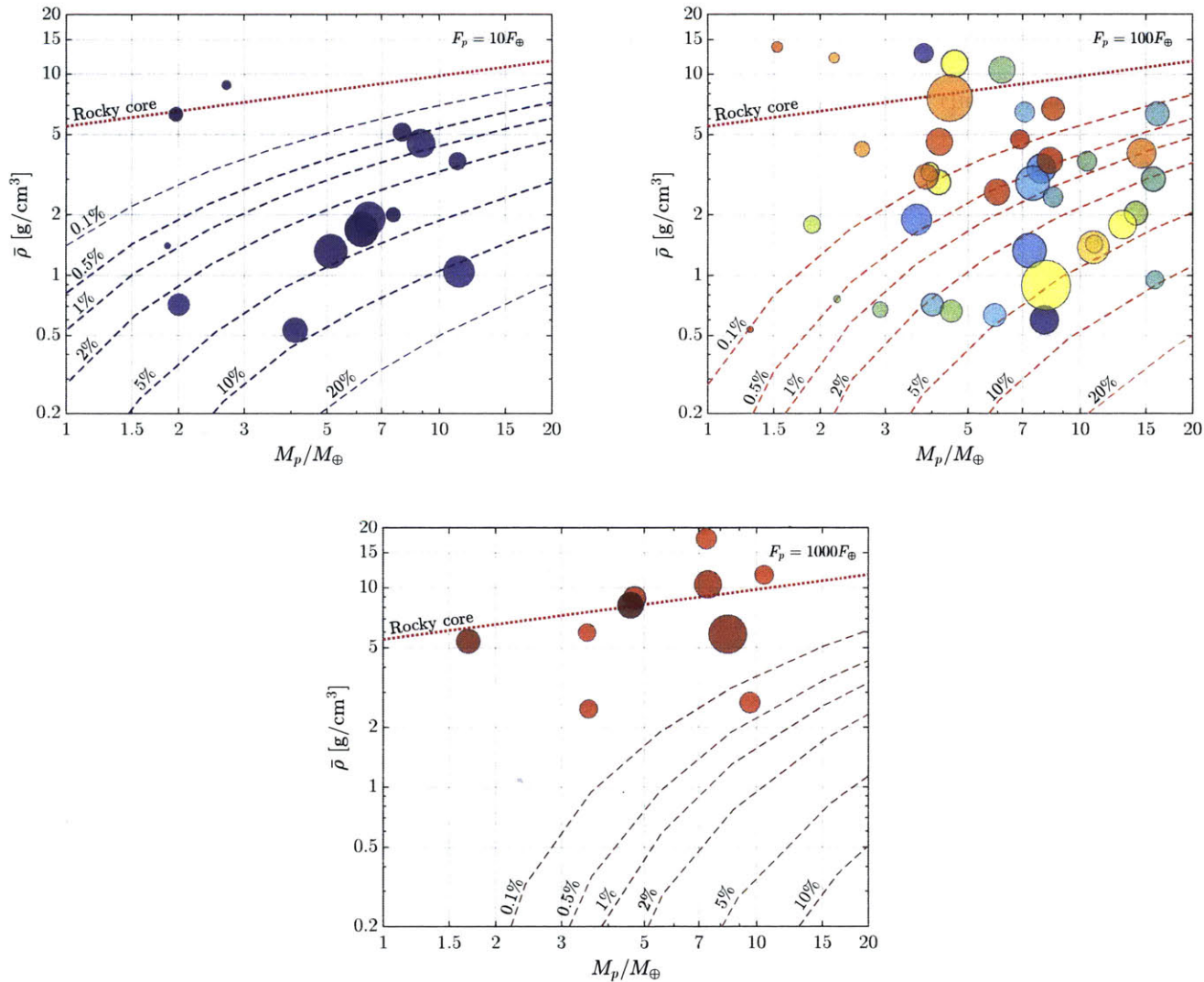
Figure 5-10: Exoplanet mean densities, $\bar{\rho}$, **as a function of planet mass** M_p **for various envelope-to-core mass fractions,** M_{atm}/M_{core}. **Each panel shows planet densities** $\bar{\rho}$ **as a function of planet mass** M_p **for a variety of labeled envelope mass fractions (dashed lines). We have binned the observed planets shown in Fig. 5-1 by flux, into (logarithmically) equal-spaced bins centered on** $10F_\oplus$ **(upper left panel),** $100F_\oplus$ **(upper right panel), and** $1000F_\oplus$ **(lower panel). In each case, the lines of constant** M_{atm}/M_{core} **are after** 4 **Gyr of evolution. The dotted red lines indicate a** $(M_p, \bar{\rho})$ **relationship corresponding to a rocky planet.**

5.3 The Diversity of the Exoplanet Population and Identification of Formation Pathways for Individual Systems

In the previous sections, we have shown that late-stage giant impacts—which are an expected outcome of planet formation, and which bring about orbital stability in closely packed systems—are capable of leading to significant diversity in planet mean density and bulk structure. In this section we now use our thermal evolution models and observations of planet mass and radius to identify individual systems whose characteristics are consistent with late-stage giant impacts as opposed to photoevaporative mass loss. We do this by first calculating the envelope mass fractions of observed exoplanets. We then examine the effect of photoevaporative mass loss and determine those regimes in the phase space that allow us to break the degeneracies between photoevaporation and giant impact-induced devolatization. We then identify those multiplanet systems whose formation is consistent with late-stage giant impacts.

5.3.1 Inferring the Bulk Composition of Observed Exoplanets

In order to determine the bulk composition (i.e., the envelope mass fraction $M_{\mathrm{atm}}/M_{\mathrm{core}}$) of exoplanets, we must use the observed envelope mass

and radius pair (M_p, R_p), as well as the system age and incident flux. In this section, we detail the technique we use for inferring the envelope mass fraction. The outlined technique is then used to derive envelope mass fractions for all reported sub-Neptune radius planets reported in the literature to date, which we report in Table 5.1. Our results here are a considerable expansion beyond those reported in both Lopez & Fortney (2014) and Howe, Burrows, & Verne (2014), amounting to 76 planets in total. We note that we have already used some of our results here in Chapter 4, to supplement the envelope mass fraction values reported in Lopez & Fortney (2014). For the demonstration that follows, we refer to Figs. 5-11, where we apply the method to the specific planet Kepler-11d.

Our post-disc dispersal thermal evolution models are generated on a grid of fixed core mass $M_{\text{core}} = \{1, 1.5, 2.4, 3.6, 5.5, 8.5, 15, 20\}$ $M_\oplus$ and fixed atmosphere-to-core mass fractions $M_{\text{atm}}/M_{\text{core}} = \{0.1, 0.5, 1, 2, 5, 10, 20\}$ %. As a result, the space of planet masses we cover is given by $M_{\text{core}}(1 + M_{\text{atm}}/M_{\text{core}})$, which is not uniform in planet mass (dotted red lines in Fig. 5-11(a)). Given an observed (M_p, R_p) pair [red cross in Fig. 5-11(a)], we must develop a way to interpolate between each irregular quadrilateral defined by the simulated grid points [circles in Fig. 5-11(a)].

Our atmospheric mass inference routine therefore takes several steps in order to account for the numerical setup. First, we generate thermal evolution models for a range of incident stellar fluxes ($\{F\} = \{1, 5, 10, 100, 1000, 5000\}$ $F_\oplus$) over the core mass and envelope mass fraction ranges given above. For a planet in a given system, we take the system age reported in the liter-

ature and determine the mass-radius relationship at that time for each atmospheric mass fraction $M_\mathrm{atm}/M_\mathrm{core}$ [solid curves in Fig. 5-11(a)].

We then determine for each $\{F\}$ the atmospheric mass fraction corresponding to an observed planet's (M_p, R_p). To do this, we generate a mass space centered around M_p, called (say) m_p. m_p is used as the independent variable for a second-order polynomial fit to the model mass-radius curves. The purpose of doing this is essentially to generate an interpolated value of the reported planet mass for each mass-radius curve. In this way, we overcome the limitation placed by a non-uniform grid. In Fig. 5-11(a), the polynomial fits are shown as thin dashed lines. The small, colored crosses show the values of the polynomial fits at each envelope mass fraction. In Fig. 5-11(b), we show how the envelope mass fraction is inferred from the cross-section in the mass-radius diagram with the projection in the right panel. Once a $M_\mathrm{atm}/M_\mathrm{core}$ value is inferred over all the $\{F\}$ values, we interpolate logarithmically over flux-envelope mass fraction space to determine $M_\mathrm{atm}/M_\mathrm{core}$ based on the planet's observed F_p. We note that, while we have shown here the interpolation routine taking place in (M_p, R_p) space, we have also performed the envelope mass inversion in $(M_p, \bar{\rho})$ space and found (as expected) no difference in our results.

The results of performing our routine on exoplanets reported in the literature are given in Table 5.1. Most of the mass-radius data are taken from Weiss & Marcy (2014) and references therein, with additional data taken from Jontof-Hutter, Rowe, Lissauer et al. (2015) and Barros, Al-

menara, Demangeon et al. (2015). For systems without published ages, we assume the age of the system is 4.5 Gyr. In our calculations, we have used the error estimates for the planet masses M_p (which for many of the reported planets can be large, given their derivation from transit timing variations) to calculate the appropriate bounds on $M_{\mathrm{atm}}/M_{\mathrm{core}}$.

Roughly a dozen of our modeled planets are in the sample of Lopez & Fortney (2014). In all cases, the agreement between our results and theirs are within a factor of one for more massive envelopes ($M_{\mathrm{atm}}/M_{\mathrm{core}} \gtrsim$ 1%), while smaller envelope mass fractions are within a factor of a few. For instance, for the Kepler-11 planets, Lopez & Fortney (2014) report their envelope mass fractions as $M_{\mathrm{atm}}/(M_{\mathrm{atm}} + M_{\mathrm{core}}) = 0.51\%$, 0.7%, 4.57%, 15%, and 2.49% for Kepler-11b,c,d,e, and f, respectively.[‡] Our average values of $M_{\mathrm{atm}}/(M_{\mathrm{atm}} + M_{\mathrm{core}})$ for the same planets are 0.13%, 1.45%, 3.95%, 11.89%, and 2.52%. Except for the lowest mass envelopes [which likely suffer from insufficient grid density in both our models and those of Lopez & Fortney (2014)], agreement is good. Differences that do exist, however, are likely due to the different opacity models employed and slightly different initial conditions.

In Fig. 5-13, we collect our results visually, showing the calculated $M_{\mathrm{atm}}/M_{\mathrm{core}}$ values as a function of planet mass with incident stellar flux indicated. We also show a two-dimensional histogram of the envelope mass fractions in both envelope mass fraction and incident stellar flux.

[‡]The original mass fraction for Kepler-11c is given in Lopez & Fortney (2014) as 6.1%, based upon a now-updated mass measurement. The value of $M_{\mathrm{atm}}/(M_{\mathrm{atm}} + M_{\mathrm{core}}) \approx 0.7\%$ quoted above is therefore taken from Howe & Burrows (2015), who use an updated planet mass value.

Most observed envelope masses that can be calculated (i.e., that are not for completely rocky or bare planets above the red line in Figs. 5-10) are between $0.1 - 10\%$. There does not appear to be any obvious dependence of the occurrence rate of $M_{\mathrm{atm}}/M_{\mathrm{core}}$ on incident stellar flux for non-rocky planets.
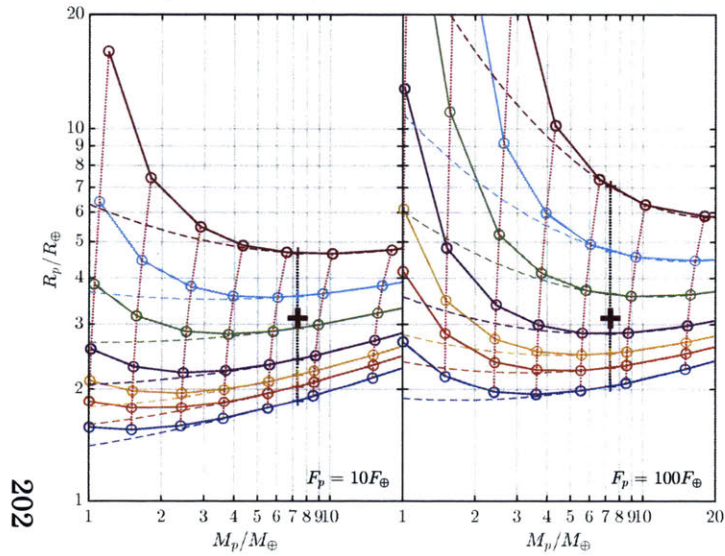
(a)
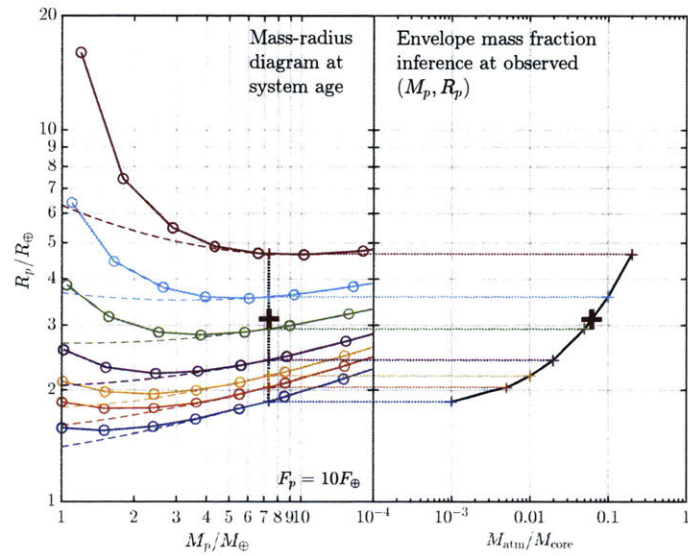
(b)

Figure 5-11: Demonstration of interpolation technique for Kepler-11d. In 5-11(a), we see snapshots at the system age ($t = 8.0$ Gyr) for stellar fluxes bounding those of Kepler-11d ($F_p = 44F_\oplus$). In 5-11(a), we see snapshots of the mass-radius diagram fixed at the observed system age. In each case, the dotted crimson lines indicate lines of constant core mass on our thermal evolution grid. The thin dashed lines are the second-order polynomial fits about the reported planet mass M_p. Each thick, solid line represents a different envelope mass fraction $M_{atm}/M_{core} = \{0.1, 0.5, 1, 2, 5, 10, 20\}\%$ from lowest to highest. In 5-11(b), we show how the observed (M_p, R_p) pair is then mapped explicitly to an envelope mass fraction. This procedure is carried out for all fluxes in the flux space $\{F\}$. The envelope mass fraction M_{atm}/M_{core} that we take for the planet is found by interpolating logarithmically amongst all the $\{F\}$ and M_{atm}/M_{core} values with the observed incident stellar flux F_p.
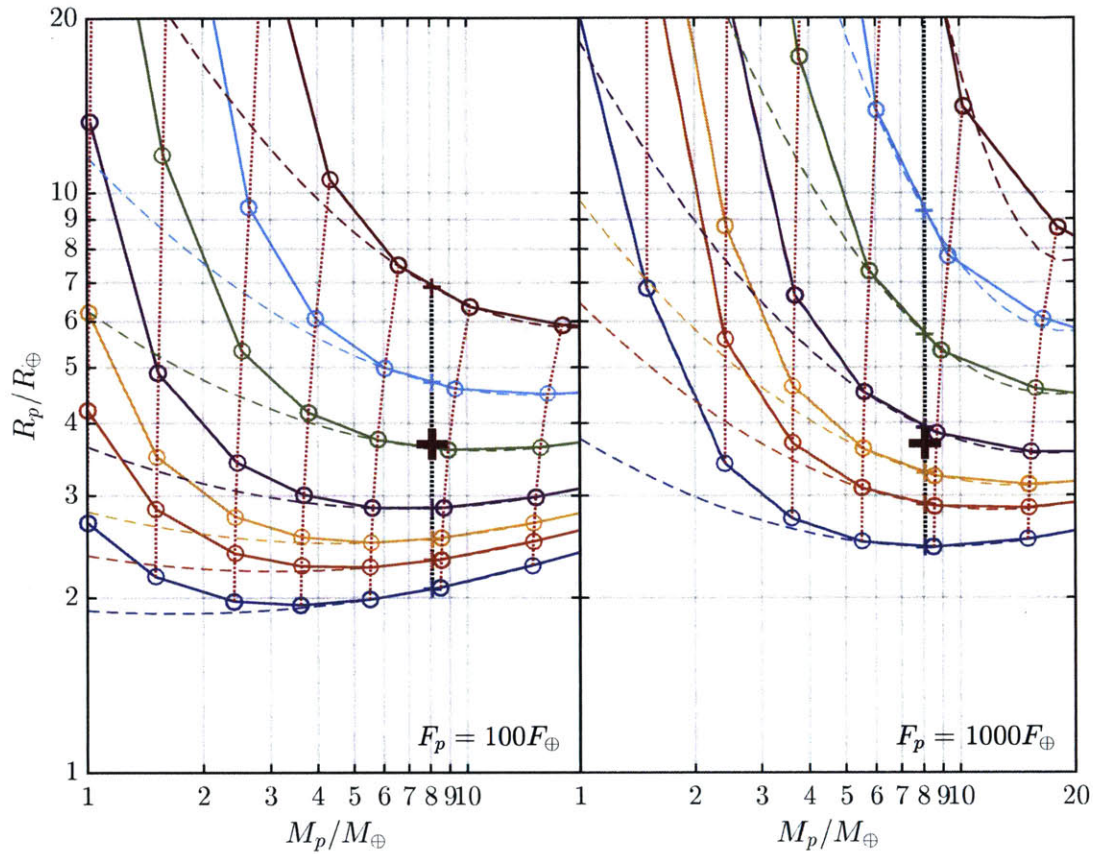
Figure 5-12: Demonstration of interpolation technique for Kepler-36c. The incident stellar flux $F_p = 180F_\oplus$ for Kepler-36c. As a result, the planet at system age ($t = 6.92$ Gyr) will be substantially more inflated than, e.g., Kepler-11d [Fig. 5-11(a)]. Due to the slightly younger age of Kepler-36c compared to Kepler-11c, the $100F_\oplus$ snapshots for the former are slightly more inflated [left panel above and right panel in Fig. 5-11(a)].

Table 5.1: Mass and radius data for observed exoplanets.

Planet	$M_p/M_\oplus$	$R_p/R_\oplus$	$\bar{\rho}$ [g/cm^3]	M_{atm}/M_{core} [%]	$F_p/F_\oplus$	t [Gyr]
55Cnce	8.4 ± 0.39	2 ± 0.08	5.9	$0.066^{+0.011}_{-0.0026}$	2400	4
CoRoT-7b	7.4 ± 1.2	1.6 ± 0.1	10	$0.04^{+0.005}_{-0.0094}$	1800	1.5
GJ1214b	6.5 ± 0.91	2.6 ± 0.09	1.9	$2.6^{+0.47}_{-0.16}$	17	4
HD97658b	7.9 ± 0.73	2.3 ± 0.16	3.4	$0.59^{+0.26}_{-0.04}$	48	4
Kepler-10b	4.6 ± 1.3	1.5 ± 0.02	8.1	$0.027^{+0.00036}_{-0.011}$	3700	11
Kepler-11b	1.9 ± 1.2	1.8 ± 0.04	1.8	$0.13^{+0.0023}_{-0.028}$	130	8
Kepler-11c	2.9 ± 2.2	2.9 ± 0.06	0.68	$1.5^{+0.78}_{-0.79}$	91	8
Kepler-11d	7.3 ± 1.1	3.1 ± 0.07	1.3	$4.1^{+3.5}_{-0.011}$	44	8
Kepler-11e	8 ± 1.8	4.2 ± 0.08	0.6	$13^{+6.2}_{-0.55}$	28	8
Kepler-11f	2 ± 0.8	2.5 ± 0.06	0.71	$2.5^{+0.049}_{-0.75}$	17	8
Kepler-18b	6.9 ± 3.5	2 ± 0.1	4.8	$0.14^{+0.01}_{-0.052}$	460	10
Kepler-20b	8.5 ± 2.1	1.9 ± 0.16	6.7	$0.081^{+0.013}_{-0.027}$	350	8.8
Kepler-20c	16 ± 3.3	3.1 ± 0.25	3	$2.3^{+0.36}_{-0.98}$	82	8.8
Kepler-20d	7.5 ± 7.2	2.8 ± 0.23	2	$4.9^{+2.8}_{-2.3}$	6	8.8
Kepler-30b	11 ± 1.4	3.9 ± 0.2	1.1	$12^{+3.1}_{-0.056}$	21	4.5
Kepler-36b	4.5 ± 0.3	1.5 ± 0.03	7.6	$0.029^{+0.0047}_{-0.0028}$	220	6.9
Kepler-36c	8.1 ± 0.53	3.7 ± 0.05	0.9	$4^{+3.8}_{-0.15}$	180	6.9
Kepler-68b	8.3 ± 2.3	2.3 ± 0.03	3.7	$0.23^{+0.14}_{-0.035}$	410	6.3
Kepler-68c	4.4 ± 2.8	0.95 ± 0.04	28	$0.0009^{+0.0088}_{-0.00076}$	190	6.3
Kepler-78b	1.7 ± 0.41	1.2 ± 0.09	5.4	$0.017^{+0.044}_{-0.015}$	3100	4.5
Kepler-100c	0.85 ± 4	2.2 ± 0.05	0.44	$0.24^{+0.063}_{-0}$	210	6.5
Kepler-100b	7.3 ± 3.2	1.3 ± 0.04	18	$0.0055^{+0.016}_{-0.00056}$	470	6.5
Kepler-93b	2.6 ± 2	1.5 ± 0.03	4.2	$0.052^{+0.038}_{-0.015}$	220	6.6
Kepler-102e	8.9 ± 2	2.2 ± 0.07	4.5	$0.41^{+0.14}_{-0.031}$	17	1.4
Kepler-102d	3.8 ± 1.8	1.2 ± 0.04	13	$0.0053^{+0.013}_{-0.0028}$	31	1.4
Kepler-102f	0.62 ± 3.3	0.88 ± 0.03	5	< 0.011	8	1.4
Kepler-102b	0.41 ± 1.6	0.47 ± 0.02	22	$0.0042^{+0}_{-0.0036}$	78	1.4
Kepler-94b	11 ± 1.4	3.5 ± 0.15	1.4	$2.3^{+1.2}_{-0.15}$	210	1.4

(Continued on next page)

Table 5.1 (continued from previous page)

Planet	$M_p/M_\oplus$	$R_p/R_\oplus$	$\bar{\rho}$ [g/cm^3]	$M_{\mathrm{atm}}/M_{\mathrm{core}}$ [%]	$F_p/F_\oplus$	t [Gyr]
Kepler-103b	14 ± 4.7	3.4 ± 0.09	2	$3.6^{+0.07}_{-1.4}$	120	6.7
Kepler-106c	10 ± 3.2	2.5 ± 0.32	3.7	$0.61^{+0.12}_{-0.02}$	84	4.8
Kepler-106e	11 ± 5.8	2.6 ± 0.33	3.7	$1.3^{+0.99}_{-0.62}$	16	4.8
Kepler-106b	0.15 ± 2.8	0.82 ± 0.11	1.5	$0.04^{+0}_{-0.038}$	240	4.8
Kepler-95b	13 ± 2.9	3.4 ± 0.09	1.8	$3.1^{+0.36}_{-0.13}$	180	5.6
Kepler-109b	1.3 ± 5.4	2.4 ± 0.07	0.54	$0.25^{+0.17}_{-0}$	440	6.3
Kepler-109c	2.2 ± 7.8	2.5 ± 0.07	0.76	$0.69^{+0.65}_{-0}$	95	6.3
Kepler-48b	3.9 ± 2.1	1.9 ± 0.1	3.3	$0.14^{+0.013}_{-0.019}$	170	3.1
Kepler-48c	15 ± 2.3	2.7 ± 0.14	4	$0.5^{+0.063}_{-0.061}$	230	3.1
Kepler-48d	7.9 ± 4.6	2 ± 0.11	5.1	$0.27^{+0.41}_{-0.16}$	14	3.1
Kepler-79b	11 ± 6.7	3.5 ± 0.07	1.4	$2.8^{+0.39}_{-0.8}$	160	3.4
Kepler-79c	5.9 ± 2.1	3.7 ± 0.08	0.63	$5.9^{+5.8}_{-1.2}$	63	3.4
Kepler-79e	4.1 ± 1.1	3.5 ± 0.14	0.53	$9^{+0.97}_{-0.42}$	15	3.4
Kepler-113b	7.1 ± 3.3	1.8 ± 0.05	6.5	$0.067^{+0.07}_{-0.033}$	64	6.9
Kepler-25b	9.6 ± 4.2	2.7 ± 0.05	2.7	$0.36^{+0.18}_{-0.022}$	670	4.5
Kepler-37d	1.9 ± 9.1	1.9 ± 0.06	1.4	$1.2^{+0}_{-1.1}$	7	6
Kepler-37c	3.4 ± 4	0.75 ± 0.03	44	$< 1.4 \times 10^{-5}$	16	6
Kepler-37b	2.8 ± 3.7	0.32 ± 0.02	470	$< 8 \times 10^{-11}$	37	6
Kepler-68b	6 ± 1.7	2.3 ± 0.02	2.6	$0.27^{+0.18}_{-0.018}$	380	6.3
Kepler-68c	2.2 ± 3.5	1 ± 0.02	12	$0.0088^{+0}_{-0.0075}$	220	6.3
Kepler-96b	8.5 ± 3.4	2.7 ± 0.22	2.5	$1.1^{+0.35}_{-0.056}$	74	2.3
Kepler-131b	16 ± 3.5	2.4 ± 0.2	6.4	$0.25^{+0.11}_{-0.12}$	72	3.7
Kepler-131c	8.3 ± 5.9	0.84 ± 0.07	77	< 0.00052	29	3.7
Kepler-97b	3.5 ± 1.9	1.5 ± 0.13	6	$0.06^{+0.039}_{-0.0088}$	850	4
Kepler-98b	3.5 ± 1.6	2 ± 0.22	2.5	$0.071^{+0.0049}_{-0.00055}$	1600	4
Kepler-99b	6.2 ± 1.3	1.5 ± 0.08	10	$0.011^{+0.0069}_{-0.0014}$	90	4
Kepler-406b	4.7 ± 1.7	1.4 ± 0.03	8.9	$0.044^{+0.012}_{-0.013}$	710	5.8

Table 5.1 (continued from previous page)

Planet	$M_p/M_\oplus$	$R_p/R_\oplus$	$\bar{\rho}$ [g/cm^3]	$M_{\mathrm{atm}}/M_{\mathrm{core}}$ [%]	$F_p/F_\oplus$	t [Gyr]
Kepler-406c	1.5 ± 2.3	0.85 ± 0.03	14	$0.011^{+0}_{-0.0087}$	290	5.8
Kepler-407b	0.06 ± 1.2	1.1 ± 0.02	0.27	> 1.02	3600	4
Kepler-409b	2.7 ± 6.2	1.2 ± 0.03	8.8	$0.0098^{+0}_{-0.0097}$	6	7
KOI-94b	11 ± 4.6	1.7 ± 0.16	12	$0.02^{+0.03}_{-0.0051}$	1200	3.9
KOI-1612$_1$	0.48 ± 3.2	0.82 ± 0.03	4.8	$0.25^{+0}_{-0.24}$	1700	4
KOI-314b	0.06 ± 1.2	1.1 ± 0.02	0.27	> 1.01	3600	4.5
KOI-314c	0.06 ± 1.2	1.1 ± 0.02	0.27	> 1.01	3600	4.5
Kepler-138b	0.066 ± 0.059	0.52 ± 0.032	2.6	$0.0082^{+0.0066}_{-0.004}$	7	4
Kepler-138c	2 ± 1.9	1.2 ± 0.07	6.3	$0.024^{+0.34}_{-0.02}$	5	4
Kepler-138d	0.64 ± 0.67	1.2 ± 0.075	2	$0.2^{+0}_{-0.14}$	2	4
K2-19c	16 ± 7.7	4.5 ± 0.47	0.96	$13^{+1.1}_{-7.1}$	76	8
Kepler-26b	5.1 ± 0.65	2.8 ± 0.11	1.3	$5.6^{+0.0076}_{-0.27}$	8	4.5
Kepler-26c	6.2 ± 0.65	2.7 ± 0.12	1.7	$5.4^{+0.096}_{-0.26}$	5	4.5
Kepler-29b	4.5 ± 1.5	3.4 ± 0.22	0.66	$3^{+2.1}_{-0.62}$	96	4.5
Kepler-29c	4 ± 1.3	3.1 ± 0.2	0.71	$2.6^{+0.96}_{-0.43}$	69	4.5
Kepler-60b	4.2 ± 0.56	1.7 ± 0.13	4.6	$0.076^{+0.004}_{-0.0052}$	318	4.5
Kepler-60c	3.9 ± 0.81	1.9 ± 0.15	3.1	$0.13^{+0.0088}_{-0.0054}$	236	4.5
Kepler-60d	4.2 ± 0.84	2 ± 0.16	2.9	$0.18^{+0.016}_{-0.015}$	161	4.5
Kepler-105c	4.6 ± 0.92	1.3 ± 0.07	11	$0.011^{+0.0036}_{-0.0015}$	161	4.5
Kepler-307b	7.4 ± 0.91	2.4 ± 0.09	2.9	$0.74^{+0.41}_{-0.045}$	60	4.5
Kepler-307c	3.6 ± 0.65	2.2 ± 0.07	1.9	$0.58^{+0.019}_{-0.0081}$	44	4.5
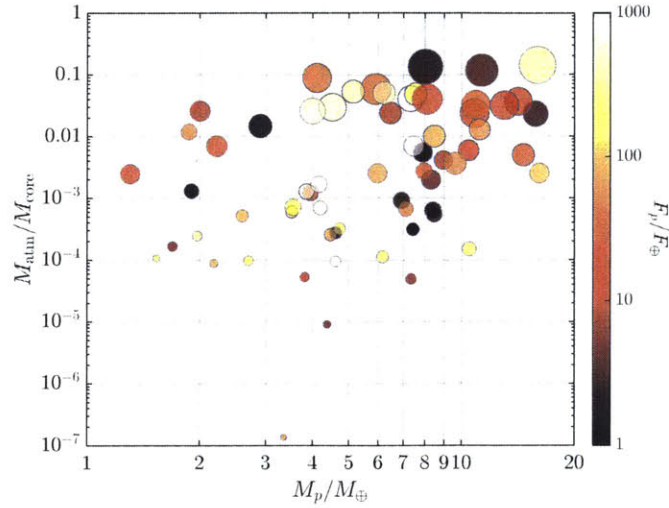
(a)
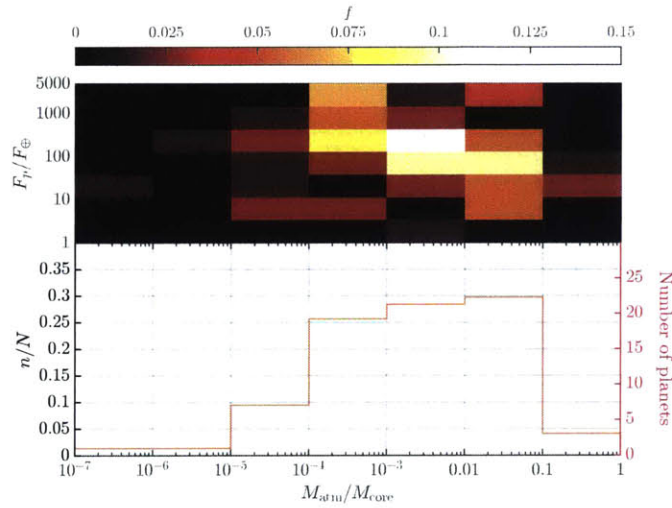
(b)

Figure 5-13: Envelope mass fractions of observed exoplanets. In 5-13(a), we show envelope mass as a function of planet mass for all of those envelope masses reported in Table 5.1. Points are colored according to incident stellar flux, while marker size is proportional to the planet radius. In the upper panel of 5-13(b), we show a two-dimensional histogram of planet occurrence rates f as a function of incident stellar flux F_p and envelope mass fraction M_{atm}/M_{core}. In the lower panel, we show these occurrences integrated over flux to show the fraction n/N of observed sub-Neptunes in a given envelope mass fraction bin, where N is the sample size $N = 76$ and n is the total number of planets in the M_{atm}/M_{core} bin. The envelope mass fraction and flux bins are spaced logarithmically. Most planets have envelope mass fractions between $0.1 - 10\%$, with no clear correlation between envelope mass fraction and incident stellar flux evident in both 5-13(a) and 5-13(b).

207

5.3.2 Breaking Degeneracies

Having now determined the envelope mass fractions for observed exo-planetary systems, we are in a position to identify potential pathways for individual systems. To do so, however, we must break the degeneracy that exists between photoevaporative mass loss and giant impacts. Indeed, photoevaporation is commonly invoked to explain the stark difference in mean densities for observed exoplanetary systems. However, given the highly uncertain nature of photoevaporation models, we seek instead to look at only end-member cases. That is, we should identify planets that should lose either all or none of their envelopes, with regimes in between taken as highly uncertain. To do so, we now use a thermal evolution model that accounts for photoevaporative mass loss.

The most-commonly used method for describing photoevaporative envelope mass loss is the so-called energy-limited mass loss model, in which the energy flux from high energy photons (either extreme ultraviolet light or X-rays) is assumed to eject envelope molecules with some efficiency ϵ. If F_{XUV} is the high energy flux as a function of time, then this gives an envelope mass loss rate $\dot{M}_{\text{atm}}$ of (Watson, Donahue, & Walker 1981)

$$\dot{M}_{\text{atm}} \sim \epsilon \frac{\pi R_p^2(t) F_{\text{XUV}}(t)}{GM_{\text{core}}/R_p(t)} = \epsilon \frac{\pi R_p^3(t) F_{\text{XUV}}(t)}{GM_{\text{core}}}. \tag{5.1}$$

Usually, the efficiency factor ϵ is taken to be 0.1 (Howe & Burrows 2015). Since we are interested in only in the extreme cases of photoevaporation

(either all or no envelope mass is lost), we assume for simplicity $\epsilon = 0.1$, since an order unity parameter change will have negligible effect on our models. The high energy flux for early times is highly uncertain, but for definiteness, we use the results of Lammer, Güdel, Kulikov et al. (2012):

$$
F_{\text{XUV}} = \begin{cases} 5.09 \times 10^{-1} \, \dfrac{\text{W}}{\text{m}^2} & t < 100 \text{ Myr} \\ 2.97 \times 10^{-2} \, \dfrac{\text{W}}{\text{m}^2} \times \left(\dfrac{t}{1 \text{ Gyr}}\right)^{-1.23} & \text{else.} \end{cases} \tag{5.2}
$$

In Fig. 5-14, we show the results of a suite of numerical simulations over a range of core masses and initial envelope masses, and the fraction of envelope mass that is lost due to photoevaporation according to Eqs. (5.1) and (5.2) with $\epsilon = 0.1$. We carry out the calculation for three levels of flux (10, 100, and 1000 times the Earth's flux), corresponding to semimajor axes of roughly 0.3 AU, 0.1 AU, and 0.03 AU, respectively.

The results shown in Fig. 5-14 indicate that there are regimes in which either all of the envelope mass should be retained or lost. If we identify multiplanet systems for which both or neither of these are the case, then we may be able to disentangle whether bulk density variation occurs due to photoevaporation or dynamical interactions.
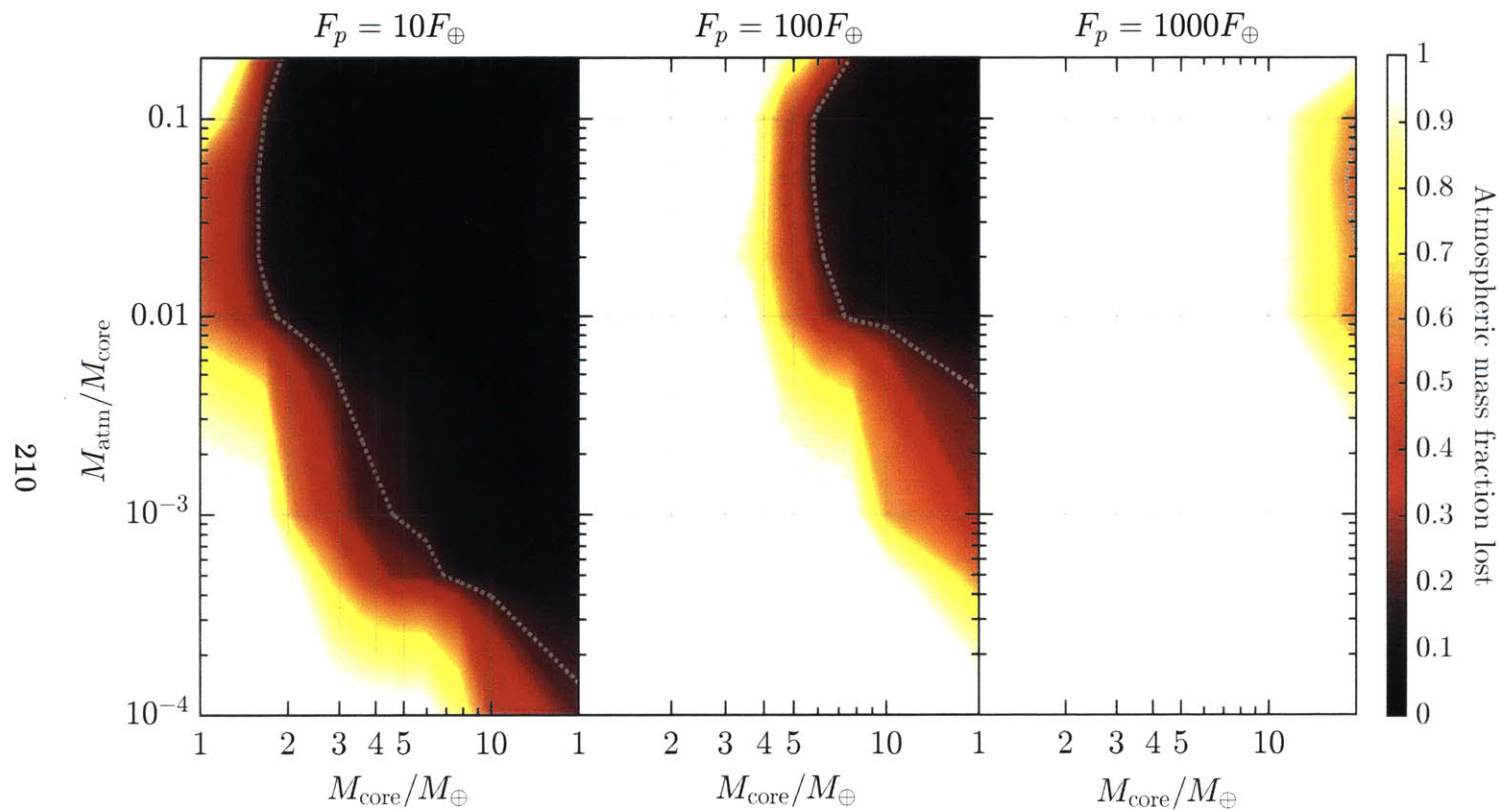
Figure 5-14: Atmospheric mass loss due to photoevaporation. For a range of core masses and initial envelope mass fractions $M_{\mathrm{atm}}/M_{\mathrm{core}}$, we determine the fraction of envelope mass lost due to photoevaporation. The simulations were carried out for three levels of incident stellar flux: 10, 100, and $1000 F_\oplus$.

In Figs. 5-15 and 5-16, we show all the multiplanet systems for which we have reliable mass and radius measurements. For each planet, we have fit atmospheric masses based on system age and reported mass and radius. We have identified, too, planets for which we expect photoevaporation to have a negligible effect. There are a number of multiplanet systems that present evidence of giant impact-induced mass loss, including Kepler-100, 102, 11, 138, 20, 36, 48, 60, and 68; these systems are thus strong candidates for formation by late-stage giant impacts. Characteristics of systems showing evidence of giant impacts are systems in which the more massive planet has a smaller envelope mass fraction (especially if it lies at a larger semimajor axis than smaller companions), and tightly packed systems with planets of similar mass having dissimilar envelope mass fractions. Other systems, on the other hand (such as Kepler-109, 26, 29, 307, 48, and 78) appear to have characteristics relatively inconsistent with either strong dynamical interactions between the planets or with significant photoevaporative erosion. These systems may have formed further out and migrated inwards, but achieved stable configurations without impacts.
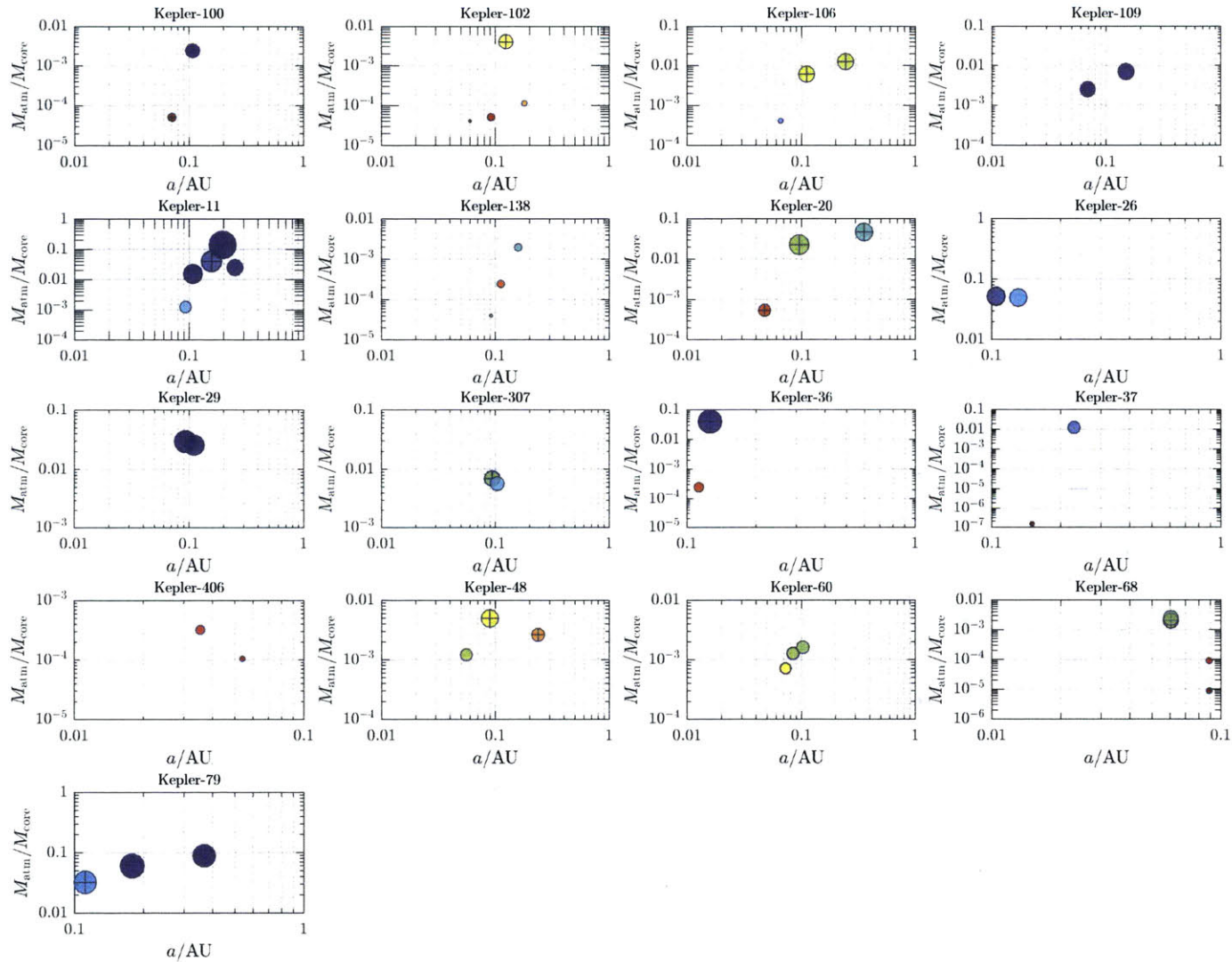
Figure 5-15: Atmospheric mass fits for multiplanet systems. Here, we show envelope mass as a function of semimajor axis. Each marker represents a planet in the indicated system, with marker size proportional to planet radius and marker color indicative of bulk density (redder indicates denser). Planets marked with a "+" are planets for which photoevaporation is not expected to have an effect on eroding envelope mass. For each system, incident stellar flux (in units of Earth flux) is indicated, as is system age.
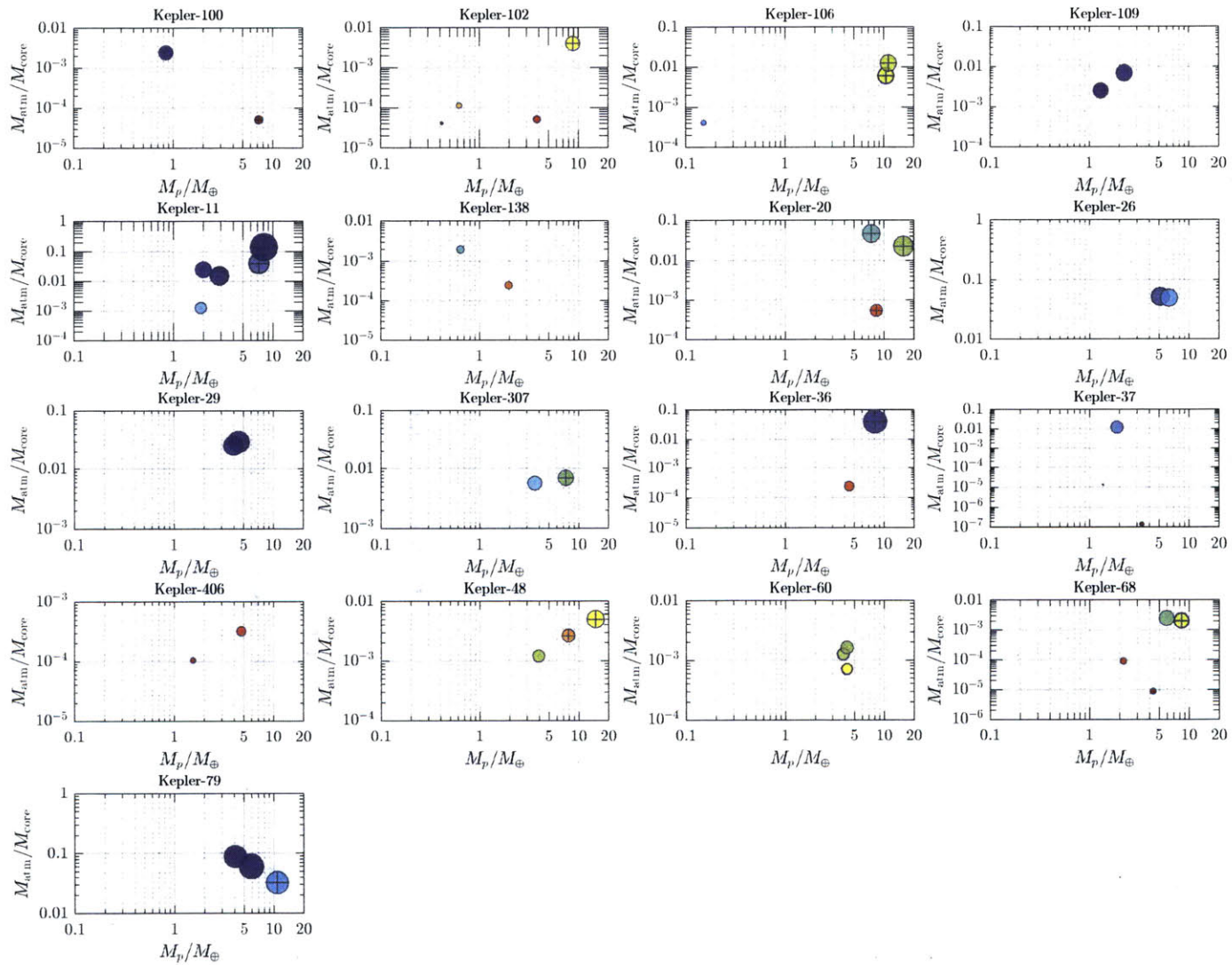
Figure 5-16: Atmospheric mass fits for multiplanet systems. Here, we show envelope mass as a function of core mass. The remaining details are the same as in Fig. 5-15.

5.4 Discussion and Conclusions

We have shown in this chapter that giant impacts between similarly sized planets can easily reduce the envelope-to-core mass ratio by factors of at least two and that this leads to an increase in mean density by factors of two or more. Since late giant impacts are frequently needed to achieve long-term orbital stability in multiple planet systems once the gas disc has disappeared (Cossou, Raymond, Hersant et al. 2014), we propose here that a small number of giant impacts may have given rise to the large spread observed in mean densities. Furthermore, giant impacts naturally yield a large diversity in densities in a given planetary system, because they are stochastic in nature and because typically only a small number of giant impacts is needed to achieve long-term stability.

In order to identify planetary systems that show evidence of such late-stage giant impacts, we then used our thermal evolution models in order to use reported masses, radii, incident stellar fluxes, and system ages for sub-Neptunes to calculate the envelope mass fraction of the ensemble. Since photoevaporation can also lead to envelope mass loss, we constructed photoevaporation models to look for end members of planet mass and envelope mass fraction in the phase space that allows us to rule out photoevaporation as the mechanism for envelope mass loss. Combining these results here, we identified a number of multiplanet systems that favor formation via giant impacts, including Kepler-100, 102, 11, 138, 20, 36, 48, 60, and 68.

The envelope mass loss we have calculated in Section 5.2.2 can be en-

hanced in several ways, including by a Parker wind (Liu, Hori, Lin et al. 2015). While our work in this chapter has increased the sample size of envelope mass fractions for observed exoplanets considerably, future observations (especially of multiplanet systems) and more precise mass measurements of exoplanets will allow us to continue to disentangle the various physical mechanisms that produce the observed diversity in the exoplanet population. In this way, we can continue to identity the most likely pathways for planet formation in order to shed light on (amongst other things) the uniqueness of our own Solar System.

Part II

Probing the Formation
History of the Solar System
via X-ray Spectroscopy

Chapter 6

Modeling the Expected Performance of the REgolith X-ray Imaging Spectrometer*

6.1 Introduction

In September 2016, NASA is scheduled to launch OSIRIS-REx ("Origins Spectral Interpretation Resource Identification Security Regolith Explorer"), a mission whose goal is to characterize and ultimately return a sample of the near-Earth asteroid (101955) Bennu (formerly 1999 RQ_{36} and hereafter Bennu; Lauretta & OSIRIS-REx Team 2012). Bennu was chosen as the target asteroid for OSIRIS-REx for several reasons. Spectral similarities in different near-infrared bands to B-type asteroids 24 Themis

---

*This chapter is based on Inamdar, Binzel, Hong et al. (2014).

and 2 Pallas raise the intriguing possibility that Bennu is a transitional object between the two. Furthermore, Bennu's reflectance spectra suggest that it may be related to a CI or CM carbonaceous chondrite analogue meteorite class (Clark, Binzel, Howell et al. 2011). Carbonaceous chondrites are believed to be amongst the most primitive material in the Solar System, undifferentiated and with refractory elemental abundances very similar to the Sun's. The discovery of water on the surface of 24 Themis provides additional scientific motivation for studying Bennu.

Bennu belongs to a class of asteroids known as near-Earth asteroids (NEA). Its semimajor axis is roughly 1 AU, and its orbit crosses Earth's (Campins, Morbidelli, Tsiganis et al. 2010). While this makes Bennu a particularly accessible target for exploration, it also makes Bennu a non-negligible impact risk to Earth. Calculations of Bennu's orbital elements suggest an impact probability of $\sim 10^{-4} - 10^{-3}$ by the year 2182, which coupled with its relatively large size (mean radius ~ 250 m) makes Bennu one of the most hazardous asteroids known (Milani, Chesley, Sansaturio et al. 2009).

Taken together, these unique features make Bennu an attractive target for future study. In order to better characterize Bennu's composition and physical state, OSIRIS-REx is equipped with a suite of instruments, amongst which is REXIS. REXIS, a student experiment aboard OSIRIS-REx, is an X-ray imaging spectrometer ("**RE**golith **X**-ray **I**maging **S**pectrometer") whose purpose is to reconstruct elemental abundance ratios of Bennu's regolith by measuring X-rays fluoresced by Bennu in response to Solar

X-rays (Fig. 6-1; Allen, Grindlay, Hong et al. 2013, Jones, Smith, & Masterson 2014). More details regarding REXIS's systems-level organization and operation can be found in Jones, Smith, & Masterson (2014).

In the previous chapters, we have focused on planet formation from the perspective of exoplanets. Nevertheless, it is clear that there are a number of outstanding questions of importance in our own Solar System, and understanding an intriguing body like Bennu will shed light on planet formation and the physics and chemistry of the early Solar System. Whether or not Bennu's composition is relatively uniform over its surface or markedly heterogeneous, what Bennu's thermal inertia is, and whether it is a rubble pile or not are amongst the questions that the OSIRIS-REx mission seeks to answer, and which will give hints to Bennu's past. In this chapter, we focus on the ability of REXIS to carry out its scientific mission of characterizing Bennu spectrally. Characterizing Bennu chemically and placing it within an analogue meteorite class may help identify potential parent bodies of meteorites and may open the door towards identifying more transitional bodies amongst the asteroid population.

6.1.1   Description of REXIS

REXIS is comprised of two distinct, complementary instruments. The first is the primary spectrometer. Measuring approximately 37 cm high and 20 cm wide, it is mounted on the main instrument deck of OSIRIS-REx and houses 4 charged coupled devices (CCDs) that measure X-rays
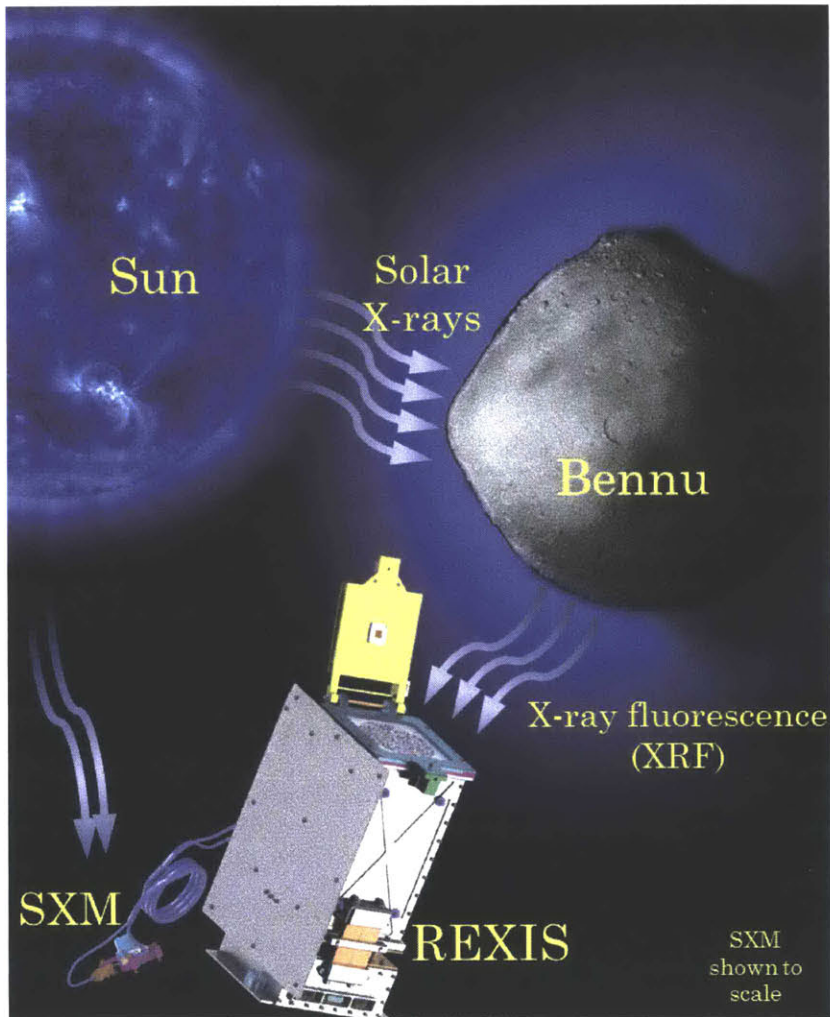
Figure 6-1: REXIS principle of operation demonstrated schematically. Except for the Solar X-ray Monitor, which is shown to scale relative to REXIS, the rest of the figure is not to scale. X-rays from the Sun impinge the regolith of Bennu, giving rise to X-ray fluorescence. These X-rays enter REXIS, where they are collected by CCDs. The radiation cover (shown in yellow) serves as a shade to prevent Solar radiation from entering REXIS. At the same time, the Solar X-ray Monitor (SXM), which is mounted on a different surface of OSIRIS-REx, collects Solar X-rays directly in order to understand the time variance of the Solar X-ray spectrum. The OSIRIS-REx spacecraft is not shown.

emitted by the Bennu's regolith (Fig. 6-2). REXIS images X-rays by means of a coded aperture mask mounted atop the spectrometer tower. The X-ray shadow pattern cast by the mask on the detector plane and knowledge of the mask pattern allows for a reprojection of the measured X-rays back onto the asteroid, so that localized enhancements in the X-ray signal on roughly 50 m scales can be identified on Bennu's surface. During the mission cruise phase, a radiation cover protects the CCDs from bombardment by nonionizing radiation (such as Solar protons) that can create charge traps in the CCDs and subsequently degrade the detector resolution (Prigozhin, Kissel, Bautz et al. 2000). This radiation cover is opened prior to calibration and asteroid observations (see below).

REXIS will observe Bennu for an overall observation period of $\sim$ 400 hr hours. During this time, OSIRIS-REx will be in a roughly circular orbit along the asteroid's terminator with respect to the Sun and about 1 km from the asteroid barycenter. REXIS will also collect calibration data. Since OSIRIS-REx orbits Bennu at approximately 1 km from the asteroid barycenter, and thus has a field of view that extends beyond the asteroid limb, cosmic sources of X-rays are a potential source of noise. Therefore, prior to asteroid observation, REXIS will observe the cosmic X-ray background (CXB) for a total of 3 hours. Furthermore, a period of 112 hours will be devoted to internal calibration to determine sources of X-ray noise intrinsic to the instrument itself. Throughout the operational lifetime of REXIS, a set of internal ^{55}Fe radiation sources (which

decay via electron capture to ^{55}Mn with a primary intensity centered at 5.89 keV) will be used to calibrate the CCD gain.

The asteroid X-ray spectrum measured by REXIS depends on both the elemental abundances of the asteroid regolith and the Solar state at the time of measurement. In order to remove this degeneracy, a secondary instrument is required to measure Solar activity. The Solar X-ray Monitor (SXM), which is mounted on the Sun-facing side of REXIS, is used to measure Solar activity and ultimately remove this degeneracy. The SXM contains a silicon drift diode (SDD) detecting element manufactured by Amptek and generates, over each 32 s observational cadence, a histogram of the Solar X-ray spectrum. The Solar X-rays collected by the SXM allow for a time-varying reconstruction of the Solar state, so that, in principle, the only unknowns during interpretation of the asteroid spectrum are the regolith elemental abundances. The elemental abundances that we infer from the collected spectra are then used to map Bennu back to an analogue meteorite class.

During the REXIS observation period, X-rays emitted by Bennu are collected on board by CCDs (CCID-41s manufactured by MIT Lincoln Laboratory). The spectra that are generated from these data are then used to interpret the elemental abundance makeup of the asteroid. The baseline CCD data flow in a single stream and REXIS data are processed in three distinct "modes" (Fig. 6-3):

Spectral Mode. Only the overall accumulation of spectral CCD data over the instrument's observational period is considered. No at-

tempt is made at producing local elemental abundance or abundance ratio maps. Instead, the data are used to determine the average composition of the asteroid from the spectral data collected in order to correlate Bennu to a meteorite class of similar composition.

Collimator Mode. Coarse spatially resolved measurements of elemental abundances on the surface of Bennu are carried out in collimator mode using time resolved spectral measurements combined with the instrument attitude history and field of view (FOV) response function. The FOV response function is uniquely determined by the instrument focal length as well as the diameter and open fraction of the coded aperture mask.

Imaging Mode. Higher spatial resolution spectral features on the asteroid surface are identified by applying coded aperture imaging (Caroli, Stephen, Di Cocco et al. 1987). In each time step, the data are the same as in collimator mode, though the distribution of counts on the detector plane is reprojected (using the known mask pattern and an appropriate deconvolution technique) onto the asteroid surface.

Science processing modes occur on the ground. Here, we are concerned with the performance of REXIS in Spectral Mode; discussion of REXIS's performance in imaging and collimator mode may be found in Allen, Grindlay, Hong et al. (2013).

6.1.2 Placing Bennu Within an Analogue Meteorite Class

One of the goals of REXIS is to place Bennu within an analogue meteorite class. Meteorites of similar class can often be grouped based on chemical or isotopic similarity. In particular, it has been recognized that major chondritic and achondritic meteorite groups can be distinguished on the basis of elemental abundance ratios, as can various subchondritic types (Nittler, McCoy, Clark et al. 2004). In Fig. 6-4, we show how various meteorite classes can be grouped on the basis of elemental abundance ratios of Fe/Si, Mg/Si, and S/Si. REXIS therefore collects X-rays between energies of 0.5 and 7.5 keV, within which prominent Fe, Mg, S, and Si emission features are found. The particular X-ray energies associated with these elements are summarized in Table 6.1. Consistent with the measurement of the X-ray signatures of these elements, REXIS has two high-level requirements associated with its performance in Spectral Mode. These are

- REX-3: REXIS shall be able to measure the global ratios of Mg/Si, Fe/Si, and S/Si of Bennu within 25% for that of a CI chondrite illuminated by a 4 MK, A3.3 Sun.
- REX-6: REXIS shall meet performance requirements given no less than 420 hours of observation time of Bennu.

The first reflects the fact that REXIS must measure the stated elemental abundance ratios to within 25% those of a typical CI chondrite during the quiet Sun. 25% error is sufficient to distinguish between achondritic

and chondritic types, as well as amongst various chondrite types, as indicated in Fig. 6-4 by the dashed line ellipses. The second requirement reflects the fact that REXIS must attain its science objectives within its allotted observation period.

Table 6.1: Summary of spectral lines of interest and their energies (Thompson & Vaughan 2001). In some cases, due to the close proximity of spectral features to one another, they are combined with one another in the analysis below.

Line Designation	Energy center [eV]	Notes
Fe-Lα	705.0	Due to proximity, combined with Fe-Lβ
Fe-Lβ	718.5	Due to proximity, combined with Fe-Lα
Mg-Kα_1/Kα_2	1,253.60	Due to proximity, combined with Mg-Kβ
Mg-Kβ	1,302.2	Due to proximity, combined with Mg-Kα_1/Kα_2
Si-Kα_1/Kα_2	1,739.98/1,739.38	—
S-Kα_1/Kα_2	2,307.84/2,306.64	—

Our purpose in this chapter is to determine, for a given asteroid regolith composition, Solar state, and instrument characteristics, the spectrum that we expect to collect from the asteroid, and how data collection and processing impact the eventual reconstruction of the hypothetical elemental abundances of Bennu. We model the expected performance of REXIS in its Spectral Mode and place bounds on its ability to place Bennu within an analogue meteorite class. We will accomplish this in several steps. First, we model the ideal X-ray spectra that we expect to be generated by Bennu and the Sun. We then model the instrument

response for both the spectrometer and the SXM, accounting for factors such as total throughput, detector active area, quantum efficiency, and spectral broadening. We then model the data processing. Here, we combine the instrument response-convolved spectra from both Bennu and the Sun to determine how well we can reconstruct Bennu's elemental abundances and place the asteroid within an analogue meteorite class. We show that REXIS can accomplish its required objectives with sufficient margin.

6.2 Methodology

Our overall methodology in simulating the expected performance of REXIS is summarized in Fig. 6-5. Our basic procedure is to first simulate physical observables—in our case, asteroid and Solar spectra—under expected conditions. We then simulate the process of data collection for both the spectrometer and the SXM. Finally, we simulate the interpretation of the data and assess our ability to reconstruct the original observables using our processed data. In order to assess our expected performance, throughout the entire modeling process, we keep track of all simulated quantities, including those that would be unknowns during the mission lifetime, such as the actual Solar and asteroid spectrum.
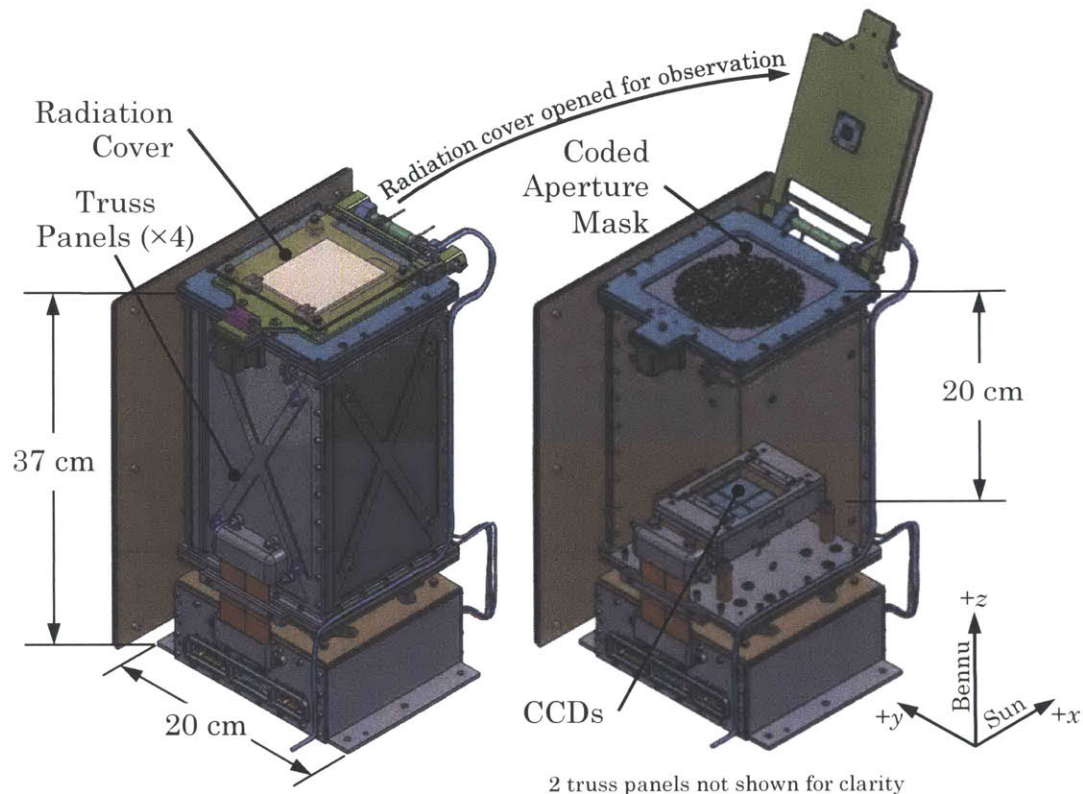
Figure 6-2: REXIS geometry. On the left, we show the instrument while it is in its stowed position, with the radiation cover closed in order to protect the CCDs from radiation damage. On the right, we show the instrument in its observing state, with the radiation cover opened to allow the CCDs to collect X-rays. The coded aperture mask is mounted above the truss structure and allows for fine angular resolution detection of local elemental abundances in Imaging Mode. In Collimator Mode, however, the telescopic "focal length" of 20 cm and the diameter of the coded aperture mask (9.84 cm) provide the coarse angular resolution. In Spectral Mode, which we focus on in this chapter, the surface area of the CCDs, the diameter of the coded aperture mask, and the focal length drive the instrument grasp (see Sec. 6.2.2).

227

REXIS Science Data Stream

Data Product #1
Spectral Mode
Globally averaged elemental
abundances of Bennu and its
analog meteorite classification

Data Product #2
Collimator Mode
Coarse angular resolution
~100m scales of possible
elemental abundance
concentrations

Data Product #3
Imaging Mode
Fine angular resolution <50m
scales of possible elemental
abundance concentrations

Increasing Spatial Resolution

Figure 6-3: REXIS data processing modes. All data processing modes
rely on the same data set, indicated schematically above. In this chap-
ter, we focus on modeling the baseline performance of Spectral Mode, in
which the globally averaged spectrum of Bennu is used without regard
to spatial variations of spectral features on the surface in order to place
Bennu within an analogue meteorite class.
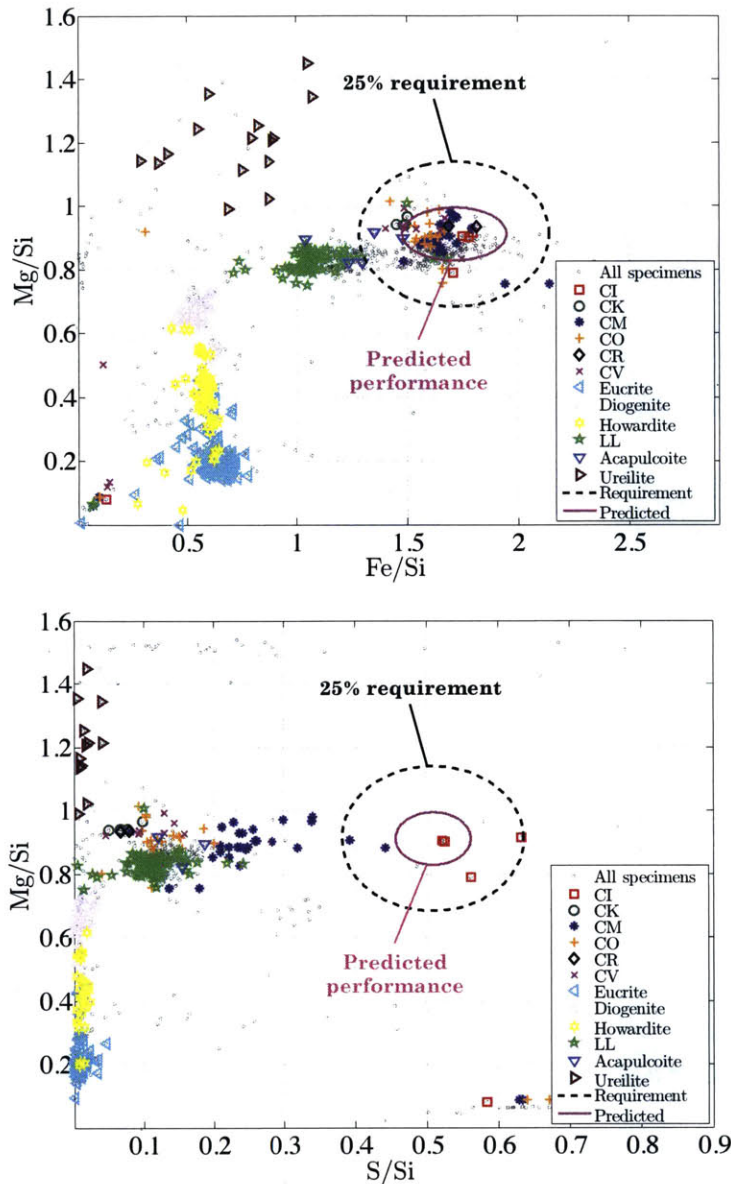
Figure 6-4: Trends in meteorite classification as a function of elemental abundance ratios. In the upper panel, we see how achondritic and chondritic meteorite specimens can be distinguished on the basis of their Mg/Si and Fe/Si elemental abundance ratios (Nittler, McCoy, Clark et al. 2004). To differentiate between various chondrite subtypes, we rely on the Mg/Si and S/Si ratios (lower panel). In both panels, we show the REXIS requirement of 25% error centered around a CI chondrite-like baseline. The expected REXIS performance under nominal conditions is indicated in the magenta ellipses. These error ellipses represent systematic error; further consideration of statistical error places the confidence in these calculations of systematic error at 3.5σ (see Sec. 6.3). Composition data for meteorites are from Nittler, *et al.* (Nittler, McCoy, Clark et al. 2004).
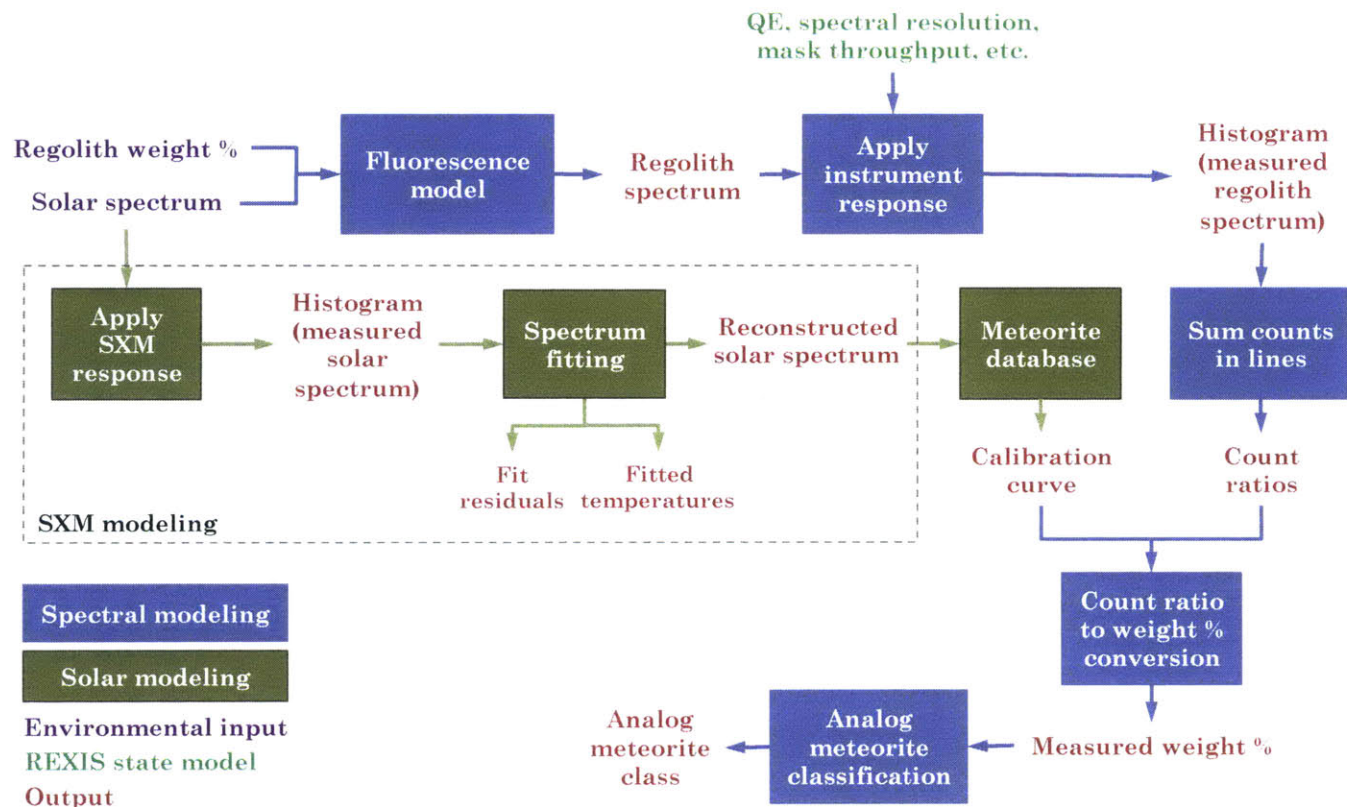
Figure 6-5: Spectral mode simulation pipeline. In this flow diagram, we indicate the various processes contained in our simulations. The primary inputs may be broadly characterized as environmental inputs (regolith composition, Solar spectrum; purple text) and REXIS state inputs, which includes geometric factors and detector characteristics accounting for the overall instrument response function (green text). Environmental inputs are used to create a synthetic regolith spectrum, to which the instrument response is then applied. Data processing is then simulated. In the SXM simulation, the "actual", differential emission measure-integrated quiet Solar spectrum is convolved with the SXM response function. We use the resulting histogram, which simulates the actual SXM data product, to reconstruct a best fit isothermal Solar spectrum. The best fit Solar spectrum, which is not detector response-convolved, is then used to generate calibration curves using a meteorite database. These calibration curves map observed photon count ratios to asteroid regolith weight ratios, which we then use to identify Bennu amongst analogue meteorite classes.

6.2.1   Simulating Observables

The baseline observables for the spectrometer and the SXM are the asteroid and Solar X-ray spectra, respectively. For the discussion that follows, we denote the asteroid spectrum $I_B(E)$ and the Solar spectrum $I_\odot(E)$. The cosmic X-ray background spectrum, which we must also consider, is denoted $I_{CXB}(E)$. In each case, the spectrum is a function of energy E and has units of $\mathrm{photons/cm^2/s/Sr/keV}$. Based on ground observations, the expected asteroid spectrum I_B is that from a CI-like asteroid regolith. Since the OSIRIS-REx mission occurs during the Solar minimum, the expected Solar spectrum is that from a quiet Sun.

Asteroid Spectrum

Asteroid spectra are calculated using the standard fluorescence equation for the intensity of the fluorescent lines (Jenkins 1995). We also include contributions from coherent scattering.[†] The contribution from incoherent scattering is at least an order of magnitude less than that from coherent scattering and is ignored here (Lim & Nittler 2009). We

---

[†]All X-ray data, including fluorescent line energies, fluorescence yields, jump ratios, relative intensities, photoabsorption cross-sections, and scattering cross-sections are derived from the compilations of Elam, Ravel, & Sieber (2002) and Kissel (2000). The Kissel scattering cross-section data may be found at the following URLs:

- `http://ftp.esrf.eu/pub/scisoft/xop2.3/DabaxFiles/f0_rf_Kissel.dat`
- `http://ftp.esrf.eu/pub/scisoft/xop2.3/DabaxFiles/f0_mf_Kissel.dat`
- `http://ftp.esrf.eu/pub/scisoft/xop2.3/DabaxFiles/f1f2_asf_Kissel.dat`
- `http://ftp.esrf.eu/pub/scisoft/xop2.3/DabaxFiles/f0_EPDL97.dat`

assume the asteroid, which is modeled as a sphere of radius 280 m, is viewed in a circular terminator orbit 1 km from the asteroid center. From the point of view of REXIS, half of the asteroid is illuminated while the other half is dark. Furthermore, the asteroid is not uniformly bright on its Sun-facing side, and the energy-integrated flux peaks at a point offset from the asteroid nadir. The effect of these angles is taken into account when generating the asteroid spectrum (for more details, see Appendix A). The asteroid spectrum itself is a function of the Solar spectrum. It is also, to a much lesser extent, a function of the CXB, which is significantly lower in intensity than the incident Solar radiation, and which is only effective at inducing fluorescence at energies much higher than we are concerned with. In generating $I_B(E)$, we use $I_\odot(E)$, as discussed below in Sec. 6.2.1.

Solar Spectrum

We calculate Solar X-ray spectra $I_\odot(E)$ using the CHIANTI atomic database (Dere, Landi, Mason et al. 1997, Landi, Young, Dere et al. 2013) and So-larSoftWare package (Freeland & Handy 1998). The Solar spectrum is that generated by the Solar corona, the primary source of X-rays from the Sun. Since REXIS will be observing Bennu during the Solar minimum, we model the expected Solar spectrum by using the quiet Sun differential emission measure (DEM) derived from the quiet Sun data of Dupree, Huher, Noyes et al. (1973) and elemental abundances of Meyer (1985) and Anders & Grevesse (1989).

The DEM is a quantity that encodes the plasma temperature dependence of the contribution function and hence intensity of the radiation (Lim & Nittler 2009, Golub & Pasachoff 1997, Landi & Drago 2008). The DEM can be derived from observations, and for the quiet Sun it tends to peak at a single temperature (in the range of about $3 - 6$ MK), so that to first order, the Solar corona can be approximated as comprising an isothermal plasma. In general, however, the actual Solar X-ray spectrum will require an integration of the DEM over all temperatures present in the plasma along the observer's line of sight (see Appendix B). For higher coronal temperatures, access to higher energy states leads to a so-called hardening of the Solar spectrum (Lim & Nittler 2009), an effect which is most pronounced during a Solar flare. In this case, the DEM peaks at more than one temperature. Since we expect the majority of our observations to take place while the Sun is relatively inactive, during data processing, we take advantage of the fact that the corona can be approximated as isothermal (for more details, see Appendix B). Finally, we note that the Solar X-ray spectrum depends on the elemental abundances of the Solar corona, for which several models are available (Lim & Nittler 2009). However, our results are relatively insensitive to the coronal elemental abundance model employed.

CXB Spectrum

The CXB spectrum that we use in our models is calculated following Lumb, Warwick, Page et al. (2002). In this model, $I_{\mathrm{CXB}}(E)$ is calculated

by assuming that the CXB is comprised of two optically thin components (Mewe, Gronenschild, & Van Den Oord 1985) and a power law component (Zombeck 1990). In general, the CXB flux becomes comparable to the asteroid flux at $\sim$ 2 keV, near the S-K complex (see Fig. 6-6). Measurement of sulfur is critical, since it enables us to differentiate amongst different chondritic varieties (Fig. 6-4). Hence, we ultimately find that measurement of the S/Si ratio is most sensitive to this particular source of noise and requires the longest amount of measurement time to achieve statistical significance (see Sec. 6.3.2).

Internal Background

Fluorescence from the REXIS instrument itself can be present in the signal we measure. Incident X-rays primarily from Bennu (but also from the CXB) can strike the inner portions of the instrument and induce fluorescence. Ideally, a ray-tracing simulation would be carried out to determine the extent of this internal noise. For our work, however, we use data from the Chandra ACIS instrument that has been suitably scaled down to match the detector area of the REXIS CCDs (Chandra X-ray Center, Chandra Project Science MSFC, & Chandra IPI Teams 2014). A comparison of Bennu's spectrum with that of the CXB and the internal background is shown in Fig. 6-6.
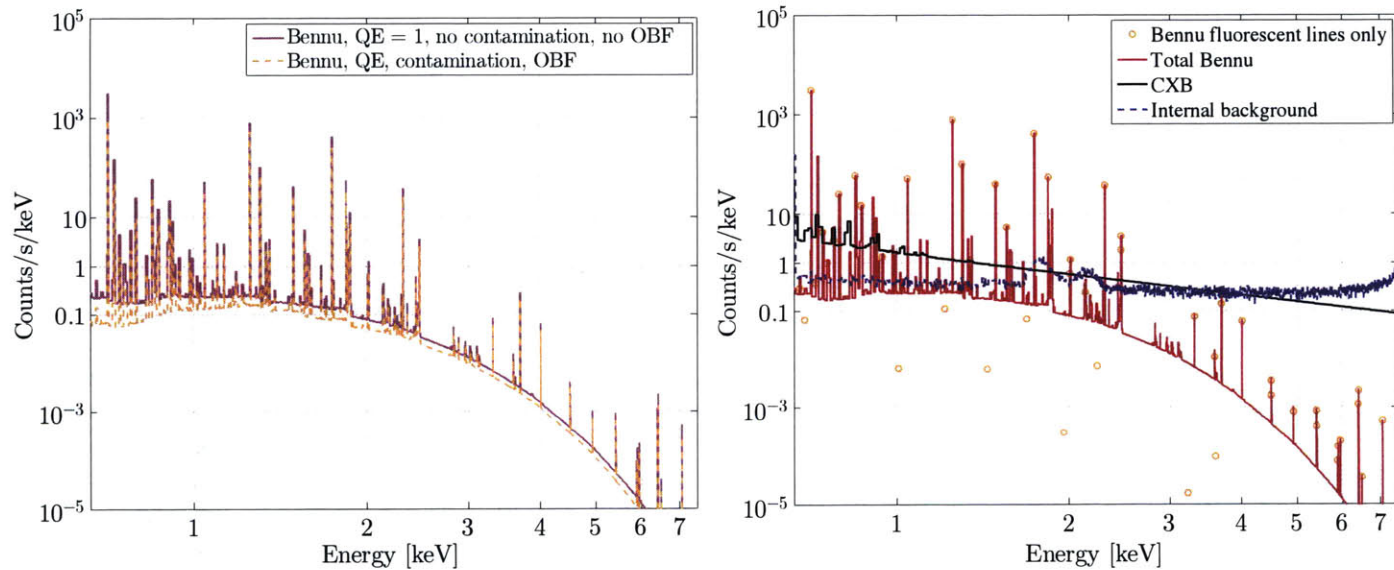
Figure 6-6: Comparison of spectra of interest. On the left panel, we show an estimate for our model of the spectrum from Bennu as we would expect to be emitted directly from the asteroid (solid magenta line). We also show Bennu's spectrum after detector quantum efficiency, molecular contamination, and the optical blocking filter (OBF; Fig. 6-7) are taken into account. On the right panel, we show Bennu's spectrum compared to sources of noise. Fluorescent lines from Bennu are shown as orange markers, while the complete Bennu spectrum, including scattering, is shown in solid red. The cosmic X-ray background (CXB) is shown in black, while internal noise (due to fluorescence from the instrument itself) is shown with the dotted blue line. The Si-K complex from Bennu, which is the prominent set of lines between 2 and 3 keV, is most strongly subject to the effects of noise. Internal background has been scaled from Chandra data (Chandra X-ray Center, Chandra Project Science MSFC, & Chandra IPI Teams 2014). This model does not include oxygen as it falls just below our model cut-off.

6.2.2 Instrument Response

The next step after simulating the observables is to estimate how these will convolve with the instrument response. Thus we simulate the data collection process by applying the instrument response for both the spectrometer and the SXM to our model spectrum. Inputs into the instrument response models include (along with the symbols that we use to denote each):

- Observation time, T_{obs}

- Coded aperture mask throughput, F (spectrometer only)

- Grasp, $G = A_E \Omega$

    - Effective detector area, A_E

    - Solid angle subtended by source with respect to detector, Ω

- Detector quantum efficiency, $Q(E)$ (a function of energy E)

- Detector histogram bin width, ΔE

- Gain drift, eV

- Detector spectral resolution, FWHM

In all cases where we are evaluating our results, we assume our measurements are well described within the realm of Poisson statistics. The origin of the values used for each of these inputs varies. In sections below, we detail how each of these inputs is derived for our simulations. After the asteroid and Solar spectra have been convolved with the detector response functions, the basic output for each will be a histogram of photon counts as a function of energy. In Table 6.2, we summarize some of the major observational inputs into our simulations, while others are

given in the text that follows.

Table 6.2: Observational inputs for spectrometer instrument response.

Parameter	Value
Open fraction	40.5%
Histogram binning ΔE [eV/bin]	~ 15
Gain drift [eV]	± 15
Total observation time T_{obs}	423 hours
CXB calibration period T_{CXB}	3 hours
Internal background calibration period T_{int}	112 hours
Solar state	Quiet Sun
Regolith composition	$\sim$CI chondritic

Observation Time

The observation time, T_{obs}, for the spectrometer is taken to be 423 hours. For the SXM, Solar spectra are recorded as histograms in 32 s intervals, roughly the time scale over which the Solar state can vary substantially. Time T_{CXB} and T_{int} is also allocated for CXB and internal calibration, respectively (see Sec. 6.2.3).

Coded Aperture Mask Throughput

The overall throughput of the instrument depends on the open fraction of the coded aperture mask (the fraction of open mask pixels to total mask pixels). For REXIS, nominally half of the coded aperture pixels are open. However, the presence of a structural grid network to support the closed pixels reduces the throughput further. Since the grid width is 10% of the nominal pixel spacing, the count rate of photons incident on the

237

REXIS detectors will be reduced by a factor $F = 0.5 \times (1 - 0.1)^2 = 0.405$ due to the presence of the coded aperture mask.

Grasp

The grasp G, which has units of $cm^2 Sr$, is the quantity that encodes the solid angle Ω subtended by the target with respect to the detector, and the averaged detector geometric area A_E that sees the target; detector efficiency is not accounted for in this term. The detector area does not comprise a single point, and since the field of view is not a simple cone, G must in general be calculated numerically. We calculate G for the CCDs using custom ray tracing routines in MATLAB and IDL. Since portions of the detector area can see the CXB that extends beyond the limb of the asteroid during observation, we keep track of this as well during our calculations.[‡] In Table 6.3, we summarize G for the CCDs, including individual contributions from Bennu and the CXB.

For the SXM, we assume that the solid angle subtended by the Sun is given by that for a distant source, $\pi \left(R_\odot / 1 \text{ AU} \right)^2$ where $R_\odot$ is the Sun's radius. Since the Sun is located at such a distance that its incident rays can be treated as parallel, and since there are no structural elements driving the SXM viewing geometry substantially, we calculate G for the SXM by simply multiplying the solid angle subtended by the Sun by the detector's 0.25 cm^2 area (Table 6.4).

[‡]For a given differential area element on the detector surface, we calculate the solid angle subtended by Bennu and the CXB and multiply each by the differential area; we then average these values over the area of the detector.

Table 6.3: Geometric inputs for spectrometer instrument response during primary observation period. These inputs assume a 280 m spherical asteroid radius, average 1 km asteroid centroid-to-spacecraft orbit, 9.84 cm mask coded area diameter, and a 20 cm focal length. During the calibration period, when REXIS observes only the sky, the entire 4.24 cm^2Sr grasp is devoted to the CXB.

Parameter	Value		
	Bennu	CXB	Total REXIS
Averaged geometric detector area A_E [cm^2]	15.16	2.09	—
Solid angle Ω [Sr]	0.254	0.185	—
Grasp G [cm^2Sr]	3.85	0.388	4.24

Table 6.4: Inputs for Solar X-ray Monitor (SXM) response.

Parameter	Value
Histogram binning ΔE [eV/bin]	~ 30
Single integration time	32 s
SDD area A_E [cm^2]	0.25
Solid angle Ω [Sr]	$\pi \left(R_\odot / 1 \text{ AU} \right)^2 = 6.79 \times 10^{-5}$
Grasp G [cm^2Sr]	1.70×10^{-6}

Detector Quantum Efficiency

The detector quantum efficiency, $Q(E)$, gives the overall reduction in counts registered by the detector due to absorption of incoming X-rays both by material overlying the CCDs and by the CCD material itself. In the case of the CCDs, we use the known material stackup (Koyama, Tsunemi, Dotani et al. 2007) and widely available photoabsorption cross-section data (Hubbell & Seltzer 1996) to determine the energy-dependent attenuation and hence quantum efficiency of the detector. We also include other possible sources of detection inefficiency, including built-up molecular contamination (Chandra X-ray Center, Chandra Project Science MSFC, & Chandra IPI Teams 2014) and the optical blocking filter (OBF), which is a thin aluminum film deposited on the CCDs in order to prevent saturation from optical light. The combined contribution of all these to $Q(E)$ is shown in Fig. 6-7.[§] For the SXM, SDD efficiency curves are taken from manufacturer data (AMPTEK 2014).

Detector Histogram Binning

The histograms that are generated by the spectrometer data are binned in intervals of width ΔE. Photons detected by the REXIS CCDs are assigned a 9 bit energy value, so that over an energy range of $0.5 - 7.5$ keV, $\Delta E = 7$ keV$/2^9 \sim 15$ eV (Table 6.2). For the SXM, there are 256

[§]When the paper on which this chapter is based was submitted, some of the quantum efficiency data shown in Fig. 6-7 were no longer up to date. However, the impact on our results is negligible, and future work will incorporate more accurate data. For more on the characterization of the REXIS CCDs, see Ryu, Burke, Clark et al. (2014)

energy bins, so that $\Delta E \sim 30$ eV (Table 6.4).

Gain Drift

Our ability to accurately define line features depends on our ability to accurately calibrate the gain of the detectors. In the case of the spectrometer, we employ on-board ^{55}Fe calibration sources in order to determine the line centers. The strength of the ^{55}Fe sources has been chosen to ensure that within a given time period, the sources' line centers can be determined with 3σ accuracy to within one bin width. In our work, we shift the gain at 5.9 keV randomly by $\pm$ 15 eV for each simulation we perform.

In the case of the SXM, we will use known Solar spectral features to accurately calibrate the gain over each integration period. Since the count rate for the SXM is so high, we can accurately determine line centers without counting statistics having too great an effect.

Detector Spectral Resolution

The detector energy resolution, which we denote by FWHM (full width half maximum), describes the width of the Gaussian distribution that a delta function-like spectral line would assume due to broadening. Natural broadening, which is typically on the order of a few eVs, is negligible in comparison to broadening from the detector itself. For the CCDs, the FWHM is a function of both photon energy and detector temperature (Koyama, Tsunemi, Dotani et al. 2007). The detector temperature drives

dark current, which subsequently increases the width of the Gaussian. REXIS's required detector operating temperature T is -60 °C or below, while the predicted temperature at the time of writing is ~ 20 °C colder than the requirement. Since the CCD temperature is the strongest driver of spectral resolution for a given line, and since the CCDs are passively cooled, in our results below, we calculate the performance over the range of detector temperatures between the requirement and prediction. CCD FWHM is determined using a combination of experimental data and analytical expressions, in a procedure outlined in Appendix E. Initial test results show that CCD performance is near or at Fano-limited. FWHM as a function of detector temperature for energies at the line centers of interest is shown in Fig. 6-7. For the SXM, the situation is somewhat simpler, since the SXM is cooled actively via a thermoelectric cooler. In this case, based on the manufacturer's test data, we have assumed that FWHM $(5.9 \text{ keV}) = 125 \text{ eV}$.
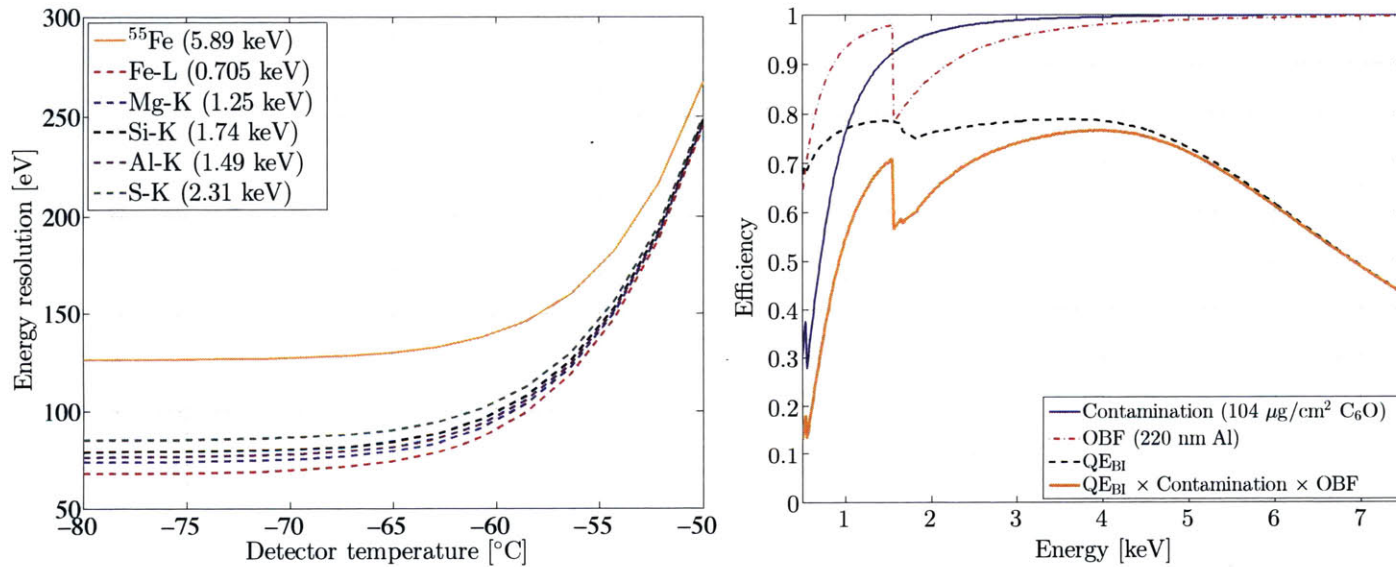
Figure 6-7: Drivers of spectral resolution and instrument response. In the left panel, we show the energy resolution of the CCDs as a function of detector temperature for various energies. The energies indicated are those associated with the line centers of spectral lines of interest. Initial tests have indicated that detector performance is at or near Fano-limited at the required detector operating temperature of $-60\,^{\circ}$C. Curves have been calculated using the method given in Appendix E. In the right panel, we show the quantum efficiency of the CCDs as a function of energy. Several sources of efficiency degradation are indicated. Molecular contamination is indicated by the solid blue line. The effect of the optical blocking filter, whose purpose is to attenuate optical light from Bennu that could cause saturation of the detectors, is indicated by the dash-dotted red line. The dashed black line indicates the radiation attenuation due to the composition stackup of the back-illuminated CCID-41. The total effect of all these are indicated by the thick orange line. Note the quantum efficiency estimate for our back illuminated CCD includes a conservative margin.

Calculating the Instrument Response Function

In this section, we summarize how all the above inputs combine to generate the instrument response and a spectrum histogram. We denote the baseline intensity as $I_0(E)$ [= $I_B(E)$, $I_\odot(E)$, or $I_{CXB}(E)$], and multiply $I_0(E)$ by the relevant geometrical and time integration factors. The number of counts $C_0(E)$ accumulated by the detector over a given integration period T_{obs} is thus

$$C_0(E) = I_0(E) \cdot G \cdot T_{obs} \cdot \Delta E. \tag{6.1}$$

During primary observation, G for the asteroid and the CXB are those given in Table 6.3. During the calibration period for the CXB, however, G is that for the whole spectrometer (i.e., 4.24 cm^{2}Sr) and instead of T_{obs}, we have T_{CXB}.

$C_0(E)$ will be reduced due to the quantum efficiency of the detector, and the resulting count distribution $C_1(E)$ is given by

$$C_1(E) = C_0(E) \cdot Q(E)$$
$$= I_0(E) \cdot Q(E) \cdot G \cdot T_{obs} \cdot \Delta E. \tag{6.2}$$

In Fig. 6-6, we show how Bennu's modeled spectrum compares to the CXB and internal background, plotting for each $C_1(E)/T_{obs}/\Delta E$. Consider a function Poisson$[C(E)]$ which takes as an input a number of counts for a given energy $C(E)$, and outputs a Poisson-distributed random num-

244

ber from a distribution whose mean is $C(E)$. Applying Poisson statistics to $C_1(E)$ then gives

$$C_2(E) = \text{Poisson}\left[C_1(E)\right]$$
$$= \text{Poisson}\left[I_0(E) \cdot Q(E) \cdot G \cdot T_{\text{obs}} \cdot \Delta E\right]. \qquad (6.3)$$

The effect of the detector state upon the spectrum is accounted for by imposing an effective broadening upon each count value in the spectrum, the broadening having the shape of a Gaussian with a given FWHM. For the CCDs, FWHM = FWHM(E, T), where T is the detector temperature and E is photon energy. This broadening will not have the shape of a precise Gaussian, however, and to simulate the stochastic nature of the broadening, we generate a random distribution sampled from a Gaussian of given energy and FWHM, with the total number of counts given by $C_2(E)$. Denote the generic Gaussian function that takes as an input the energy E, the counts at that energy $C(E)$, and the FWHM(E, T) as Gaussian $\left[E, C(E), \text{FWHM}(E, T)\right]$. Then the distribution of counts $C_3(E)$ is given by the convolution of Gaussian and $C_2(E)$:

$$C_3(E) = C_2(E) * \text{Gaussian}\left[E, C_2(E), \text{FWHM}(E, T)\right]. \qquad (6.4)$$

A histogram is then generated by binning $C_3(E)$ into the required number of bins. Consider a generic binning function that takes as inputs a counts profile $C(E)$ over what may be regarded as a continuous energy range $E \in [0.5 \text{ keV}, 7.5 \text{ keV}]$, and bins it into a new profile $C'(E)$ over an

energy range E', where $E' = \{0.5 \text{ keV}, 0.5 \text{ keV} + \Delta E, 0.5 \text{ keV} + 2\Delta E, ..., 7.5 \text{ keV}\}$. Denote this function $C'(E') = \text{Binning}\left[C(E), E; E'\right]$. Then the final histogram profile $C_3'(E')$ over the binned energies is given by

$$C_3'(E') = \text{Binning}\left[C_3(E), E; E'\right]. \tag{6.5}$$

In Fig. 6-8 we show a simulated histogram for a detector temperature $T = -60$ °C. For reference, the spectral features associated with our lines of interest are shown in thick colored lines. Noise subtraction (see Sec. 6.2.3 below) has been applied. In Fig. 6-10, we show the simulated histogram for the quiet Sun (solid magenta line), along with the idealized spectrum from which it is derived (dotted red line).

6.2.3 Data Processing

In this section, we detail our methods for processing the simulated spectrometer and SXM data (right hand and bottom side of Fig. 6-5). We first begin by detailing the process of noise background subtraction. Next, we discuss how we perform histogram counts for all our lines of interest. We then discuss the method for reconstructing the Solar spectrum, and finally how we generate calibration curves to map from histogram count ratios back to elemental abundance weight ratios.

246
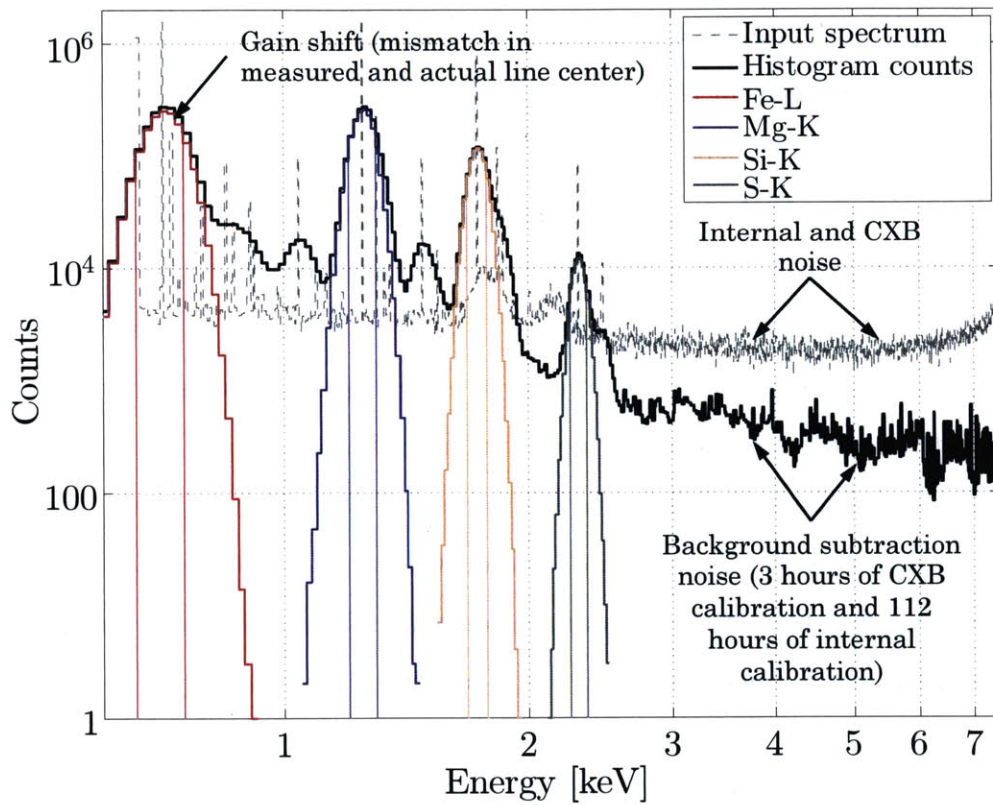
Figure 6-8: Example histogram of detector at temperature of $-60\ °C$, where these results follow from our preliminary model in Fig. 6-6.

Spectrometer Background Subtraction

As noted, asteroid spectral features (in particular, sulfur) are sensitive to noise from the CXB and from the instrument itself. As a result, REXIS devotes periods of time for both CXB and internal noise calibration. The data gathered during the calibration period are used to subtract out sources of noise from the data product. To simulate the noise subtraction procedure, we consider the total observed histogram counts $C_3'(E')$, which includes both the internal and CXB signal. We then simulate the calibration period data. Let internal counts as a function of energy be given by $C_{int}'(E')$ and CXB counts by $C_{CXB}'(E')$, where we have assumed that Poisson statistics and binning have been performed on each, and that the counts have been accumulated over the calibration periods, T_{int} and T_{CXB}, respectively, for each. Then the procedure for background subtraction is to scale each calibration count value up so that the integration time matches that of the primary observation. Thus, the spectrum that we consider after accounting for noise subtraction is given by

$$C_3'(E') - C_{int}'(E')\frac{T_{obs}}{T_{int}} - C_{CXB}'(E')\frac{T_{obs}}{T_{CXB}}. \tag{6.6}$$

In Fig. 6-8, we see the effect of noise subtraction. The thin, dashed line shows the input spectrum with CXB and internal noise especially prominent at higher energies. After subtracting out CXB and internal background, there remains some high frequency noise (right side of Fig. 6-8) since Poisson statistics are included for both the simulated observation

and calibration data. We note again that, since REXIS performs its CXB calibration with a sky observation in the absence of Bennu, $C'_{\mathrm{CXB}}(E')$ is calculated with a grasp given by that for the entire spectrometer.

Spectrometer Line Counting

The quantities of various elements present in Bennu's regolith are determined by the strength of the corresponding spectral features and hence counts in the spectrometer histogram. When counting, for a given line center, we consider all the counts within the FWHM of a Gaussian centered about that line center. As discussed above, the FWHM is a function of both line center energy E and detector temperature T. Onboard calibration data, which give us FWHM at 5.9 keV, and pre-flight test data, allow us in principle to estimate the detector FWHM for each CCD frame that is processed. For the purposes of our simulations here, we assume complete knowledge of FWHM. Furthermore, as we demonstrate below (Sec. 6.3.1), our results are relatively insensitive to FWHM. We do not assume that we know the actual line centers with complete certainty. By employing gain drift (see Sec. 6.2.2 above), we allow for the misidentification of the line centers. In Fig. 6-8, we indicate the counting zones for the lines of interest by vertical lines. *All* counts within a given FWHM counting zone are considered to come from that line of interest. Therefore, there will naturally be contamination within each zone from both the continuum background, noise sources, and other lines. During our simulations, we therefore keep track of the ideal, expected count number

from each line in addition to those we count directly from the histogram. The error between the two affects how well we are able to reconstruct weight ratios from these counts. More details on the counting scheme and the definition of "accuracy" are given in Appendix C.

Solar Spectrum Reconstruction

The method by which we use the SXM histogram to reconstruct the Solar spectrum is shown schematically in Fig. 6-9. First we take the quiet Sun spectrum (see Sec. 6.2.1) and convolve it with the detector response to generate a synthetic "observed" histogram (two blue boxes on the upper left of Fig. 6-9). Then we utilize a database of isothermal spectra which we generate beforehand, and convolve those with the known detector response to generate isothermal spectrum-derived histograms (boxes on the lower left). We use the observed histogram and that from the database to determine a best fit. The unconvolved isothermal spectrum whose convolved form provides the best fit is used in our later analysis. In Fig. 6-10, we summarize the results of the fitting procedure. In Fig. 6-10, we see that there are good fits over the energy ranges corresponding to our elements of interest.

We have found that using a linear combination of isothermal spectra when fitting against the observed Solar spectrum can provide better fits. This result is to be expected since the realistic quiet spectrum is indeed, via integration over the DEM, a linear combination of isothermal spectra. For simplicity here, however, we focus only on single-temperature

250

fits. The quality of the fit depends also on the characteristics of the isothermal spectrum database. These spectra are dependent on factors such as the coronal elemental abundance model employed. While we do not claim to have explored the full model space of elemental abundance models available, we have ensured that the abundance models used for both the DEM-convolved realistic spectrum and that from which the isothermal spectrum database is derived are distinct and randomly chosen.
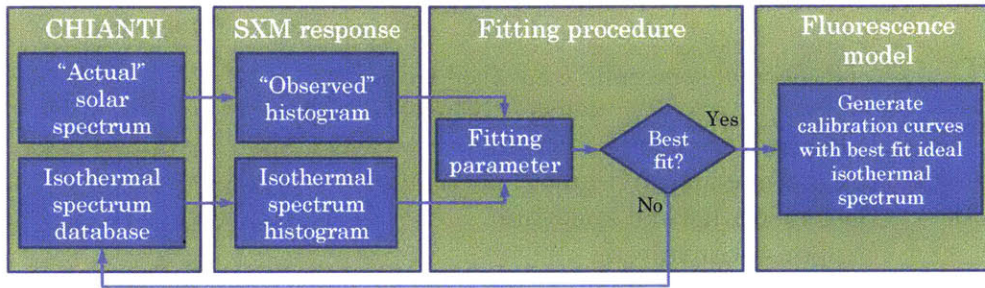
Figure 6-9: Solar X-ray Monitor (SXM) modeling flow. Here, we start with the "actual" Solar spectrum, which is multithermal (or equivalently, differential emission measure convolved) and which we model using the CHIANTI atomic database (Dere, Landi, Mason et al. 1997, Landi, Young, Dere et al. 2013) and the SolarSoftWare package (Freeland & Handy 1998). This spectrum is convolved with the SXM instrument response function, which subsequently creates a synthetic Solar X-ray histogram data product. We rely on the fact that, in the quiet state, the Solar X-ray spectrum can be approximated as isothermal. We hence generate a database of isothermal spectra, which we likewise convolve with the SXM response function, and then fit to the response-convolved quiet Sun histogram. The isothermal spectrum corresponding to the best fit isothermal histogram is then used to generate calibration curves.
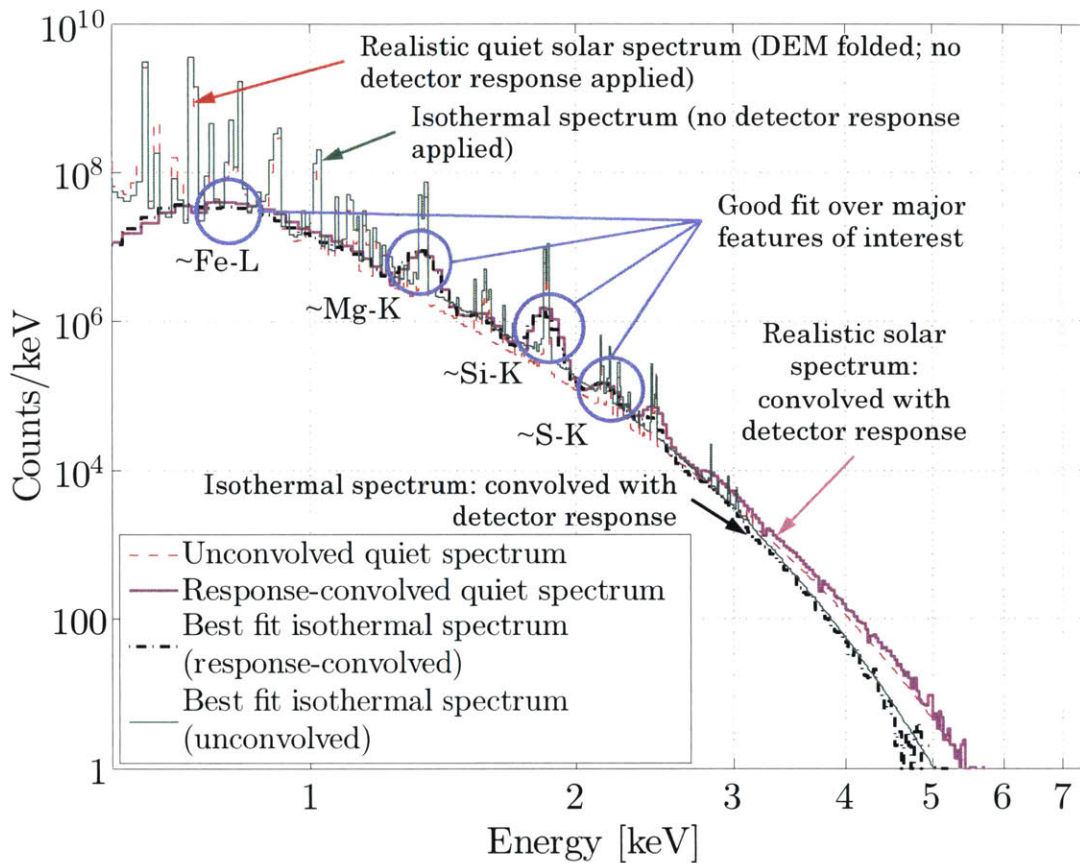
Figure 6-10: Fitting of quiet Solar spectra in the isothermal approximation. Example of fitting procedure for a simulated quiet Sun. First, a DEM-folded, quiet Sun Solar spectrum model is convolved with the SXM detector response function, which simulates ground data for the Solar state. Then, we search for a best fit isothermal spectrum and convolve that with the detector response function. The isothermal spectrum whose detector response has the best fit with that of the realistic spectrum is then used to generate the calibration curves (see Fig. 6-11). The histogram shown is over 32 s, while in this particular case, the isothermal best fit is a 3.1 MK spectrum.

252

Calibration Curve Generation: Mapping Count Ratios to Weight Ratios

For a given Solar state, in order to make the transition from histogram count ratios to elemental abundance ratios, we make use of so-called calibration curves (Lim & Nittler 2009). Calibration curves map, for a given Solar state, elemental abundance ratios to ideal count ratios. To generate calibration curves, we take the unconvolved Solar spectrum and simulate asteroid spectra corresponding to a wide range of meteorite compositions. The range of elemental abundances afforded by this range allows us to consider realistic weight ratios that may be expected from Bennu. The simulated spectra allow us to determine the expected count ratios for a given weight ratio, which then allows us to map histogram counts back to elemental abundances. In Table 6.5, we detail the various meteoritic compositions used to generate the calibrations curves for our elements of interest. In Fig. 6-11, we show calibration curves for our elemental ratios of interest. Various meteorite groups are indicated by the symbols. Solid lines indicate second-order fits to the realistic quiet Solar spectrum, while the dotted line indicates the fit based on the reconstructed Solar spectrum, which we discuss further below. For simplicity, we show the symbols only for the realistic quiet Sun fit and omit those for the reconstructed fit. The baseline weight ratios and expected count ratios for a CI chondrite-like regolith are indicated by the blue circles.

Table 6.5: Summary of inputs for calibration curves. Values taken from Lodders & Fegley (1998) unless marked with "*", in which case the data come from Nittler, McCoy, Clark et al. (2004). If "—" appears, weight percent values were not available in the reference for that element, and the remainder of the percent balance was allocated to that element for the sake of the ideal asteroid spectrum simulation. Since O is not one of our elements of interest, the development of the calibration curves is not incumbent upon accurate knowledge of O.

Class	Weight percent by element					
	O	Mg	Al	Si	S	Fe
CI	46.4	9.70	0.865	10.64	5.41	18.2
CM	43.2	11.5	0.130	12.70	2.70	21.3
CM*	38.92	8.99	1.334	11.916	2.122	25.466
CV	37.0	14.3	1.680	15.70	2.20	23.5
CO	37.0	14.5	1.400	15.80	2.20	25.0
CK	—	14.7	1.470	15.80	1.70	23.0
CR	—	13.7	1.150	15.00	1.90	23.8
CH	—	11.3	1.050	13.50	0.35	38.0
H	35.70	14.1	1.06	17.1	2.0	27.2
L	37.70	14.9	1.16	18.6	2.2	21.75
LL	40.00	15.3	1.18	18.9	2.1	19.8
EH	28.00	10.73	0.82	16.6	5.6	30.5
EL	31.00	13.75	1.00	18.8	3.1	24.8
R	—	12.9	1.06	18.0	4.07	24.4
K	—	15.4	1.30	16.9	5.5	24.7
Acap.*	—	15.6	1.20	17.7	2.7	23.5
Lod.*	25.858	16.299	0.0952	11.248	0.4257	43.92
Dio.*	20.42	16.528	1.00	24.28	0.204	12.729
IAB*	26.62	11.92	1.31	14.48	7.04	33.9
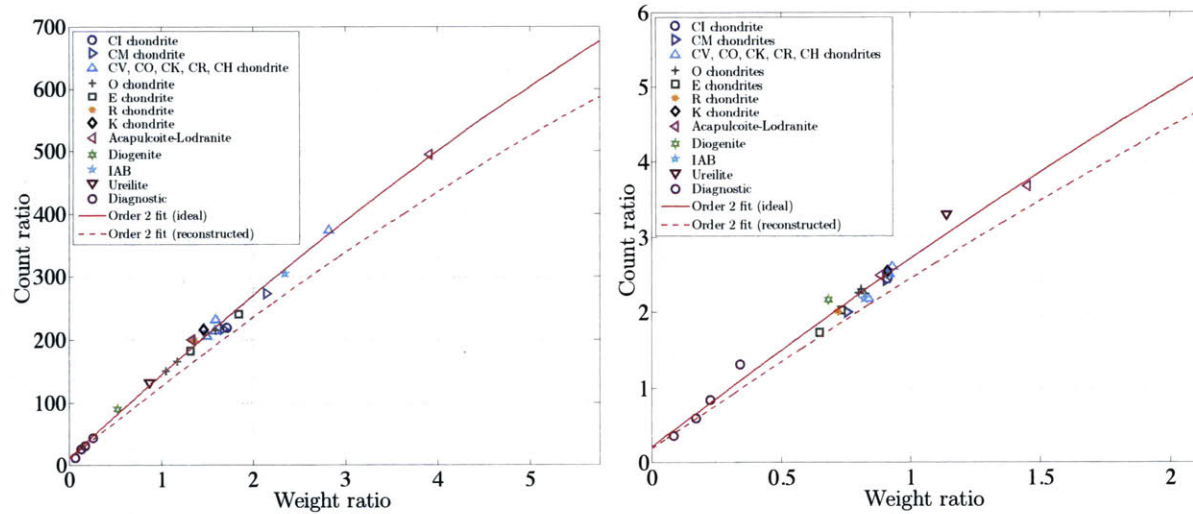
(a) Fe/Si weight ratio to count ratio calibration curve. (b) Mg/Si weight ratio to count ratio calibration curve.
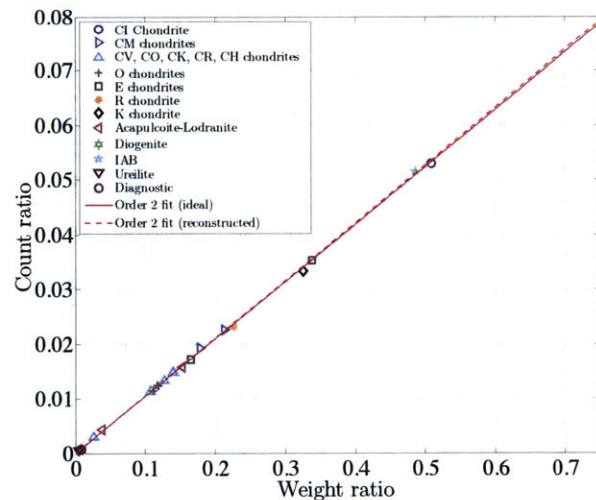
(c) S/Si weight ratio to count ratio calibration curve.

Figure 6-11: Weight ratio to count ratio calibration curves for the elemental abundance ratios of interest. Calibration curves are a function of weight ratio and Solar spectrum. Curves indicate second-order fits to simulated spectra of asteroids with the same composition as the major meteorite types indicated (see Table 6.5). Several artificial, "diagnostic" compositions have also been included to improve the fidelity of the fit. The solid curve indicates spectra generated using the DEM-folded quiet Solar spectrum. The dotted curve indicates spectra generated using the isothermal best fit (see Fig. 6-10). The baseline weight ratios and expected count ratios for a CI chondritic-like regolith are indicated by the blue circles.

6.3 Results

In this section, we detail our results based on the simulations and analysis presented in the previous sections. We first show how, assuming perfect knowledge of the Solar state, the count ratios we measure map back to elemental abundance weight ratio errors with respect to CI chondrite-like baseline composition. Next, based on these count ratio errors, we present the required observation times to achieve statistical significance on our measurements. Finally, we present a qualitative discussion of how the error generated during Solar spectrum reconstruction develops a permissible error space in which we interpret our results.

6.3.1 Weight Ratio Accuracy

Using the calibration curves given in Fig. 6-11, we can map the errors incurred by our histogram counting procedure into subsequent errors in weight ratio. In general, since the relationship between counts and regolith weight is roughly linear, the correspondence between count ratio error and weight ratio error is also roughly linear. In Table 6.6, we list the weight ratio errors for our required detector temperature ($T = -60$ °C) and our current best prediction for the detector temperature ($T \sim -80$ °C). In all cases, the predicted error is less than the requirement. The weight ratio errors over the range of temperatures between $T = -60$ °C and -80 °C are shown on the left panel of Fig. 6-12. The error bars in the figure represent the error spread over 20

simulations at each detector temperature, with each simulation incorporating the effect of factors such as Poisson statistics, gain drift, and noise subtraction. We note that, over most of the temperatures, there is not necessarily a degradation of performance with increasing detector temperature, as we might naively expect due to the decrease in spectral resolution. The relative insensitivity of our spectral performance on detector temperature (or equivalently FWHM) is primarily due to the fact that taking count ratios effectively cancels some of the effect of this systematic error present in each of the individual lines. In Fig. 6-4, we indicate with magenta ellipses the accuracy error due to these systematic effects at $T = -60\,^\circ\mathrm{C}$.
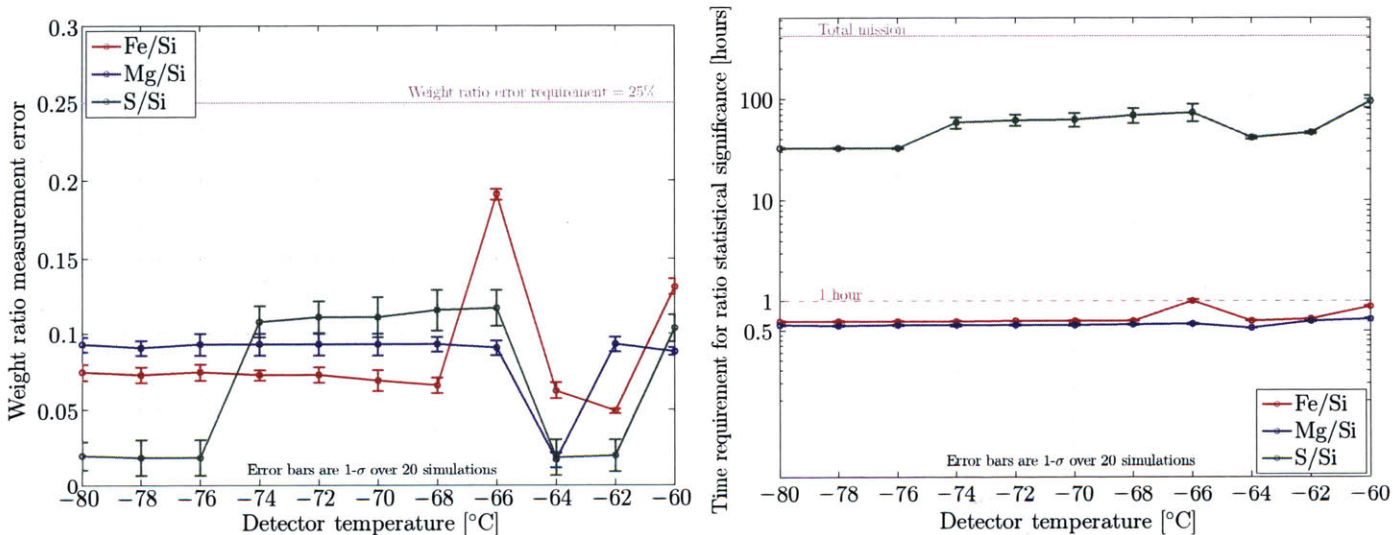
(a) Weight ratio error as a function of detector temperature.

(b) Observation time requirement to achieve statistical significance.

Figure 6-12: Spectral simulation results assuming perfect knowledge of the Solar state. In the left panel, we show the weight ratio error for our line ratios of interest as a function of detector temperature. The 25% requirement is indicated by the magenta line. For all temperatures, we are able to meet our requirement with margin. In the right panel, we show the observation time required to achieve statistical significance on our spectral measurements. The requirement that we achieve significance within the allotted mission time is shown by the magenta line. The observation time is derived from the count rates from both the lines of interest, and CXB and internal noise, as well as the difference between predicted accuracy error and required accuracy error. S/Si, which is most susceptible to noise from the internal background and CXB, requires the greatest observation time. Error bars are 1σ over 20 simulations.

Table 6.6: Summary of REX-3 systematic accuracy error performance. We show the accuracy error for each elemental abundance ratio of interest at two different detector temperatures. The detector temperature of -60 °C represents the required detector operating temperature, while -80 °C represents the current best prediction for the detector temperature at the time of writing.

		Ratio	Predicted accuracy error [%]	Requirement [%]	Margin (Requirement/Prediction)
Det. Temp. T	-60 °C	Fe/Si	13.7	$\leq 25\%$	1.8
		Mg/Si	9.1	$\leq 25\%$	2.7
		S /Si	10.3	$\leq 25\%$	2.4
	-80 °C	Fe/Si	7.9	$\leq 25\%$	3.2
		Mg/Si	9.7	$\leq 25\%$	2.6
		S /Si	3.0	$\leq 25\%$	8.3

6.3.2 Observation Time

The results in the above section represent systematic error. That is, they represent errors intrinsic to the behavior of the instrumentation itself. We must also consider statistical error to account for the stochastic nature of photon emission and place a statistical significance on our expected results. In order to account for statistical error, we consider the quadratic difference between our expected *count* ratio error and allowed count ratio error. Then, assuming Poisson statistics, based on the count rates within each energy range of interest from both the fluorescent lines and noise sources, we determine the required observation time to achieve an $N\sigma$ statistical significance level. Here, we choose $N = 3.5$, corresponding to $> 99\%$ confidence. A detailed calculation of the above, as well as expected count rates, is given in Appendix D and Table D.1,

259

respectively. Our required observation times for the two detector temperatures discussed above are given in Table 6.7, while those for the range of temperatures in between are given in the right panel of Fig. 6-12. Again, in all cases, we are able to achieve our required performance with margin. As noted above, the S/Si ratio, which is most subject to the effect of CXB and internal noise, requires the greatest amount of observation time to achieve statistical significance. The magenta error ellipses shown in Fig. 6-4 thus have a 3.5σ statistical confidence associated them.

Table 6.7: Summary of REX-6 observation time requirement expected performance for detector temperature of -60 °C and -80 °C. The observation time is based on obtaining sufficient photon statistics to achieve 3.5σ confidence in the accuracy results.

		Ratio	Observation time for 3.5σ confidence [hours]	Requirement [hours]	Margin (Requirement/Prediction)
Det. Temp. T	-60 °C	Fe/Si	0.9	≤ 420	467
		Mg/Si	0.7	≤ 420	600
		S /Si	108	≤ 420	3.9
	-80 °C	Fe/Si	0.6	≤ 420	700
		Mg/Si	0.6	≤ 420	700
		S /Si	33	≤ 420	12.7

6.3.3 Calibration Curves and Mapping Errors

In the results above, we have assumed perfect knowledge of the Solar state in mapping count ratio errors to weight ratio errors. Hence, we have used the solid red calibration curves shown in Fig. 6-11. In reality, we will have to use the reconstructed Solar spectrum-derived calibration

curves (dotted red lines in Fig. 6-11) in order to perform the mapping. From the calibration curves shown in Fig. 6-11, it is clear that the difference between the curves based on the actual and reconstructed Solar spectra are within the error of the fit itself over the weight ratio ranges of interest, so that we cannot claim a truly meaningful difference between the quality of each fit. While we have not accounted for this effect in the results presented above, we demonstrate graphically in Fig. 6-13 the error space that develops from reconstructing the Solar spectrum. Fig. 6-13 shows how, in the most extreme case (Fe/Si) the calibration curves diverge for the different input Solar spectra. The lines marked "baseline" map to a CI chondrite-type composition under the actual Solar spectrum. The 25% identification requirement then places a range of permissible errors about this baseline (the shaded grey area). So long as the combined misidentification of the Solar spectrum and the count ratio error do not exceed the bounds set by the shaded region, REXIS can still achieve its objectives. In our case, for the $T = -60$ °C case, the Fe/Si error is the most marginal (13.7% error), although it still falls within the required performance region. Since Fe/Si is not subject to the same statistical fluctuation as, e.g., S/Si, statistical significance on Fe/Si can still easily be achieved within the allotted observation time.
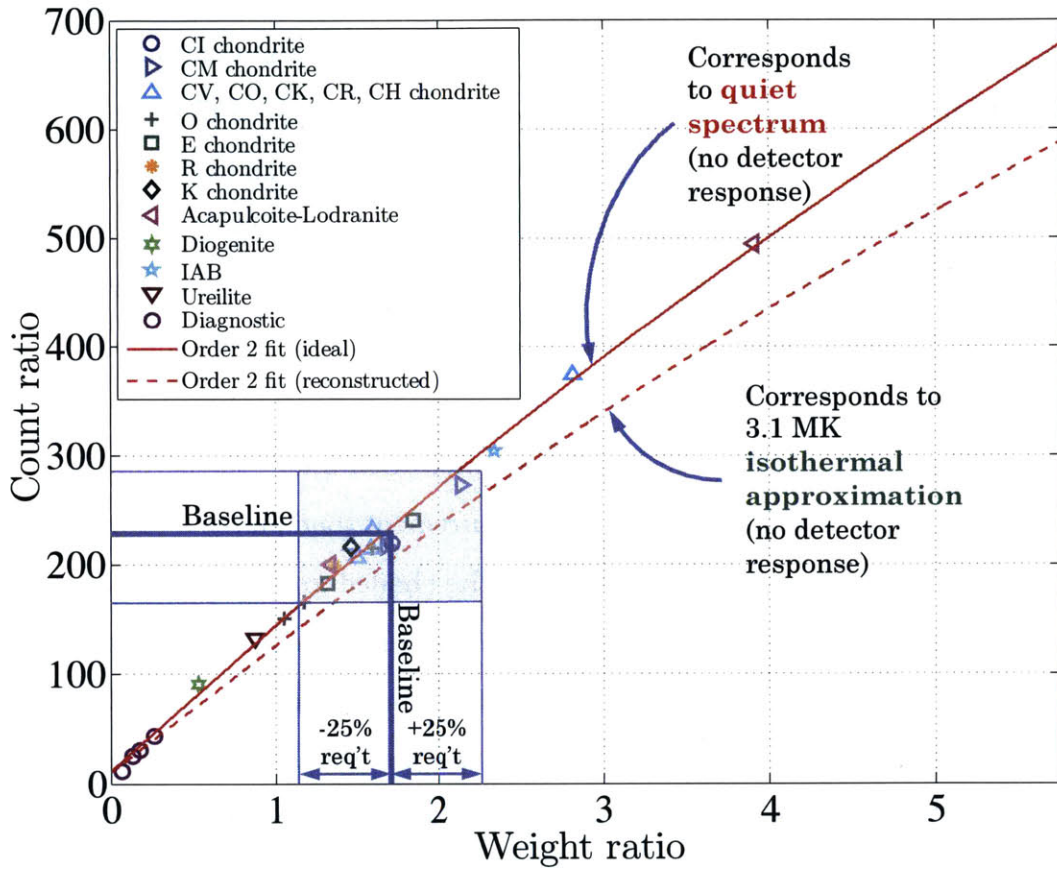
Figure 6-13: Calibration curve error space. The solid red line is the calibration curve generated by the quiet Sun (red line in Fig. 6-10). This is the calibration curve that would be generated if we had perfect knowledge of the Solar state. The dashed red line is the calibration curve generated by the best single temperature fit based on simulated SXM data (green line in Fig. 6-10).

6.4 Conclusions

In this chapter, we have presented the methodology and results of spectral performance modeling of REXIS. We have shown, by simulating Solar and asteroid X-ray spectra, the subsequent data product, and the data processing, how well REXIS can be expected to identify Bennu as a CI chondrite analogue. We have shown that our two primary requirements—that REXIS is capable of identifying a baseline CI chondrite meteorite analogue for Bennu to within 25%, and that it can accomplish this within the allotted mission observation time—are attainable with margin.

Conclusion

Chapter 7

Summary and Conclusion

The fields of astrophysics and planetary science are at a truly exciting time. In the months immediately preceding the writing of this thesis, *New Horizons* has arrived at Pluto and provided us with striking and compelling images of the dwarf planet and its companion Charon, and *Juno* has arrived at Jupiter to help solve (amongst others) the outstanding question of the nature of Jupiter's interior. In the months to come, OSIRIS-REx will begin its journey to the near-Earth asteroid Bennu, and the Transiting Exoplanet Survey Satellite (TESS) will be completed and launched, dramatically increasing the number of detected exoplanets in our observed sample. In the years since exoplanets were first discovered in 1995, our understanding of planets and their formation both in our own Solar System and in the galaxy has increased dramatically, and the same will be true in coming months and years. The purpose of this thesis has been to contribute to—and to help lay a framework to

continue—increasing that understanding.

7.1 Overview of the Thesis

The contributions of this thesis touch upon two distinct general topics. In the first, we addressed several of the outstanding questions in planet formation theory from the point of view of exoplanets. In particular, observations by the *Kepler* space telescope and other surveys have revealed to us that most observed planetary systems look distinctly different from our own. The preponderance of close-in super-Earths and mini-Neptunes and how they formed is one of the outstanding questions in astrophysics. We tackled this general problem in two ways.

7.1.1 Exoplanet Formation

First, we considered whether super-Earths and mini-Neptunes formed *in situ*, at their observed semimajor axes. In Chapter 4, using the fact that most observed close-in exoplanets appear to possess gaseous envelopes constituting a non-negligible fraction of the total planet mass, we constructed a self-consistent formation history that considered the accretion of gas envelopes in the presence of the full disc by isolation masses; the subsequent erosion of gas envelopes once the phase of giant impacts—which are required to assemble the final planet masses—sets in; and the accretion of gas envelopes by fully assembled planetary cores in the presence of the depleted gas disc. By following this formation his-

tory, we found that while it is possible that rocky, terrestrial planet analogues with envelope mass fractions $\lesssim 0.1\%$ formed *in situ*, it is likely that exoplanets whose gas envelopes comprise $\gtrsim 1\%$ their total planet mass likely formed further out and migrated inwards. This is due to the fact that isolation masses are too small to accrete significant gaseous envelopes in the inner disc, with accreted envelope masses being further eroded due to the giant impact assembly phase. This suggests that perhaps all close-in exoplanets formed further out, or intriguingly, that there are distinct formation pathways for rocky exoplanets and those with significant gas envelopes.

Next, we considered the fact that the observed population of super-Earths and mini-Neptunes possesses a considerable spread in its bulk structural composition, contrary to what might be predicted from gas accretion models. In Chapter 5, we proposed that late-stage giant impacts, which are likely responsible for bringing about dynamical stability to dynamically packed systems, are in part responsible for the observed spread in planet mean density. We constructed thermal evolution models to determine planetary radii at early times as well as giant impact models that showed one or two late-stage giant impacts could easily strip away at least half the envelope mass of an exoplanet and increase its mean density by up to an order of magnitude.

Since photoevaporation is often cited as a means by which planets can be devolatized (hence altering their mean densities), we constructed photoevaporation models to determine planet and envelope masses that

are unlikely to have been affected by photoevaporation. Using observed masses and radii for 76 sub-Neptune exoplanets, we inferred their envelope masses and, in concert with our photoevaporation models, identified exoplanetary systems that, based on orbital compactness and inferred envelope mass fraction variation as a function of mass, were probably shaped by late-stage giant impacts. From our results, more than half of the observed multiplanet systems for which masses and radii have been measured appear to have been sculpted by late-stage dynamical interactions and giant impacts.

7.1.2   Solar System Formation via REXIS

The second general topic we considered was planet formation in the context of our Solar System. In September 2016, the OSIRIS-REx spacecraft will begin its journey to the near-Earth asteroid Bennu, with the goal of characterizing it. REXIS is an X-ray imaging spectrometer traveling aboard OSIRIS-REx with the goal of using the fluorescent X-ray signature of Bennu to determine the asteroid's composition and place it within an analogue meteorite class. Meteorites are generally believed to be fragments of material from asteroids, and this, coupled with the fact that ground-based observations suggest Bennu is a carbonaceous chondrite, means that characterizing the chemistry of Bennu has the potential to shed light on the chemical and dynamical history of the early Solar System.

In Chapter 6, we presented analysis carried out to determine whether

REXIS is capable of inferring Bennu's elemental abundance ratios and placing it within an analogue meteorite class. To do so, we constructed asteroid and Solar X-ray emission models to determine the X-ray signatures of both the asteroid and the Sun. We then simulated the instrument response and data analysis and found that REXIS will be able to distinguish Bennu amongst various chondrite types when it arrives at Bennu in 2019. REXIS's determination of Bennu's bulk composition and chemical heterogeneity will, in concert with OSIRIS-REx's other instruments, help shed light on Bennu's history. Whether Bennu is in fact a transitional object between 24 Themis and 2 Pallas, and whether Bennu can be placed within an analogue meteorite class, will shed light on how small bodies in our Solar System coalesced and formed and were altered.

7.2   Directions for Future Work

As we have mentioned, planetary science and astronomy are both very dynamic fields, with advances made daily towards understanding the formation and evolution of planetary systems. As such, there are numerous opportunities for further work to be carried out that builds upon the work in this thesis. Below, we highlight a few of these.

7.2.1 Exoplanet Formation

Population Synthesis Models

In the first part of this thesis, we constructed planet formation and evolution models that for the first time account for the thermal evolution of exoplanets (in the presence of the gas disc and after the gas disc has dissipated) and the effect of dynamical interactions on their bulk structure and composition via giant impacts. In recent years, *population synthesis* models have been constructed to try to explain general features of exoplanet mass-period and radius-period diagrams (see, e.g., Ida & Lin 2004, Mordasini, Alibert, & Benz 2009). Typically, these models significantly parametrize the physics of gas accretion, thermal evolution, and photoevaporation for computational reasons. With increasing computational power, these models have begun including dynamical interactions between planets (allowing for scattering, for instance, between planetesimals), but thus far they have not accounted (as we have) for the effect of dynamical interactions on the bulk composition and structure of exoplanets. In Fig. 4-9, we have already done some work along these lines by performing Monte Carlo simulations to determine envelope mass loss for an ensemble of exoplanets through a phase of giant impacts. Future work would expand upon these results to account for further evolution after gas disc dissipation (including giant impacts and photoevaporation).

The Initial Conditions of Thermal Evolution Models

In our thermal evolution models (see, e.g., Section 5.2.1), we have assumed that planets start off with an initial, hot state with an arbitrarily high entropy or intrinsic luminosity. This is in contrast to "cold start" models, in which much of the planet's residual heat of formation has been radiated away or lost at disc dispersal. Whether this is indeed the case is still the matter of research and debate (e.g., Lagrange, Bonnefoy, Chauvin et al. 2010). It is important to understand how model assumptions such as initial conditions impact how we interpret the data.

In Fig. 7-1, we show some preliminary results indicating how varying initial conditions impact thermal evolution models and how we may interpret them. In the upper panel, we show the difference between hot start models and a model in which the initial planet radius is set such that the radiative convective boundary R_{RCB} is, at the time of disc dispersal, given by $2R_{\mathrm{core}}$. The latter result is motivated by models investigating the "sucking action" of the disc on an accreted gas envelope as the disc disperses (Ginzburg, Schlichting, & Sari 2015). We plot the fractional difference in planet radius after 1 Gyr of thermal evolution between planets that start off with a high intrinsic luminosity and an initial $R_{\mathrm{RCB}} = 2R_{\mathrm{core}}$ condition (solid lines) and between the $R_{\mathrm{RCB}} = 2R_{\mathrm{core}}$ case and a hot start in which the initial entropy is set at a high $10k_{\mathrm{B}}/\mathrm{baryon}$ (dash-dotted lines). The hot start models are those with large initial radius and denoted by $R_{\mathrm{large},0}$. The colder, $R_{\mathrm{RCB}} = 2R_{\mathrm{core}}$ case is denoted by $R_{\mathrm{small},0}$. We see that for massive envelopes

271

($M_{atm}/M_{core} \sim 0.1$) and for core masses $\sim 2M_\oplus$, the difference between a hot and cold start model can be over 100%. In the lower panel, we show an example of planet entropy versus planet radius for a core mass of $2M_\oplus$ and for various envelope mass fractions. We indicate with markers entropies and planetary radii corresponding to an initial condition of $R_{RCB} = 2R_{core}$, as well as the planet radius after 1 Gyr of thermal evolution for each of the three cases (initial planet state set by $R_{RCB} = 2R_{core}$, initial planet state set by high intrinsic luminosity L, and initial planet state set by $10k_B$/baryon). As noted in Chapter 5, the greatest impact is for planets with masses $\sim 2M_\oplus$. While these preliminary results show that planet radii can be misinterpreted by a significant amount, more work will help understand mass-radius relationships and envelope mass fraction inferences at these interesting masses.

Gas-to-Dust Ratios and Disc Thermochemistry

In Chapter 4, we assumed throughout that the amount of solid material available to form rocky cores was driven by the gas-to-dust ratio. The value adopted ($\sim 10^{-2}$) is motivated by cosmochemistry and widely used in the literature. However, recent work with new observational data has begun investigating variations in the gas-to-dust ratio in protoplanetary discs (Ansdell, Williams, van der Marel et al. 2016), which has the potential to shed light on how the availability of solid material may vary in different systems. While work in the past (e.g., Lewis 1972) considered the thermochemistry of the protoplanetary disc to better understand ice

272
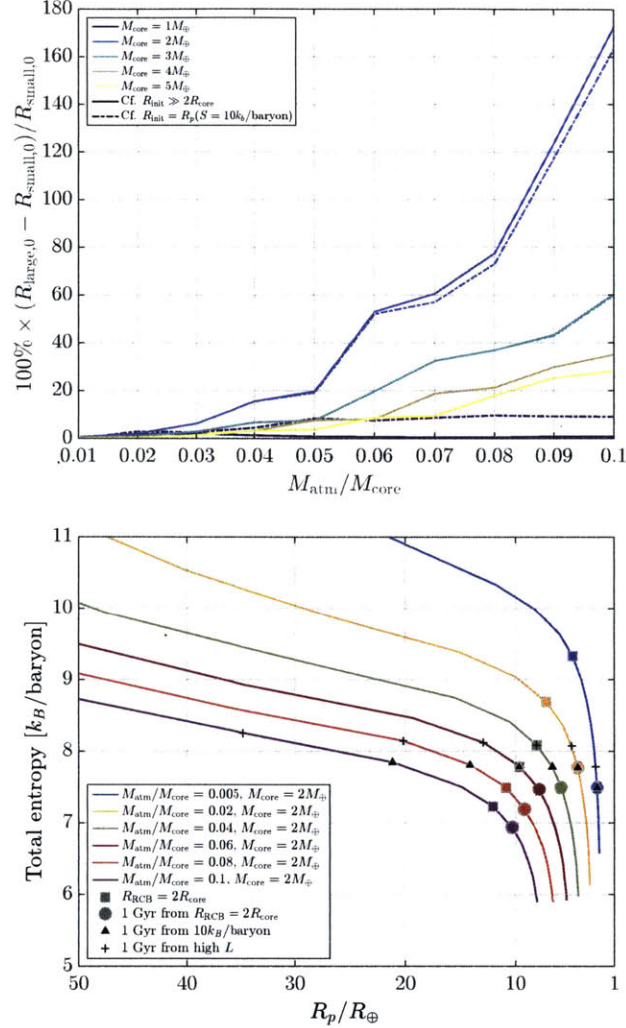
Figure 7-1: The effect of initial conditions on interpreting thermal evolution models. In the upper panel, we show the fractional error between hot start models (models with large initial radius $R_{\mathrm{large},0}$) and cold start models with smaller initial radii ($R_{\mathrm{large},0}$). The errors are shown for various core envelope mass fractions and core radii. Solid lines compare the cold start model described in the text and the high intrinsic luminosity hot start, while the dash-dotted lines compare the cold start model with a high, fixed envelope entropy model. Curves are shown for 1 Gyr of thermal evolution. In the lower panel, we show envelope entropy as a function of planet radius for various $M_{\mathrm{atm}}/M_{\mathrm{core}}$ values and fixed core mass $M_{\mathrm{core}} = 2M_{\oplus}$. We have indicated with a square the point where the radiative convective boundary $R_{\mathrm{RCB}} = 2R_{\mathrm{core}}$ at disc dispersal in the cold model, and have indicated planetary radii for the different models after 1 Gyr of thermal evolution.

lines, models of disc thermochemistry do not appear to have been used to directly place constraints on volatile and refractory material inventory for planet formation as a function of semimajor axis in a rigorous way using, e.g., chemical equilibrium models. Doing so will also shed light on the local availability of material for secondary atmospheres, and could be applied to models of Moon formation, in which chemical gradients in the protoplanetary disc (which are then mapped to interacting planetary embryos) are assumed to be linear functions of semimajor axis (Mastrobuono-Battisti, Perets, & Raymond 2015).

7.2.2 REXIS, and Asteroid and Solar System Characterization

The work presented in Chapter 6 represents the first step in understanding REXIS's science performance in Spectral Mode, and there are a number of opportunities to extend and refine this work. We summarize some of these below.

Other Baseline Regolith Compositions for Bennu

Throughout our work in Chapter 6, we assumed a baseline CI chondrite-like regolith composition for Bennu. However, ground measurements have suggested a possible CM chondrite-like composition. While we should not expect any substantial difference in expected performance if we assumed a CM-type composition, it is worthwhile to consider the

possibility that the baseline composition of the regolith is something radically different (for instance, achondritic).

Higher Order Observational Effects

We have assumed in our work that the orbit is perfectly circular and that Bennu is a perfect sphere, e.g., we do not take into account surface roughness. However, it is possible to use the Bennu shape model (Nolan, Magri, Howell et al. 2013) and OSIRIS-REx orbit for REXIS science operations in order to model the effect of both shape and orbit to a higher level of fidelity.

Active Sun Modeling and Reconstruction

We have assumed throughout this work that the Sun is in a quiet state, which it will be for the majority of REXIS's science operations. However, the Sun will occasionally flare, creating a higher flux and harder (i.e., greater intensity at higher energies) spectrum. This in turn will substantially affect Bennu's spectrum. Performing a similar analysis to that carried out in this thesis, but assuming a flaring Sun, should be carried out. The flaring Sun, however, cannot be approximated as isothermal, although a two-temperature model may suffice [see Appendix B and also Lim & Nittler (2009)].

Improved SXM Modeling

In our work here, we have assumed a relatively simple geometry and instrument response function for the SXM. Much as we have done for the spectrometer, it is possible to compute higher fidelity values for the SXM grasp and, with continued testing, better characterize the SXM response function in general.

Radiation Damage

Preliminary work has suggested that under OSIRIS-REx's expected radiation environment, degradation in CCD spectral resolution due to non-ionizing radiation damage still permits REXIS to meet its science objectives. This is in part due to the presence of the radiation cover (Inamdar 2013, Bralower 2013, Prigozhin, Kissel, Bautz et al. 2000) and the fact that REXIS is primarily concerned with the measurement of count ratios, which reduces the effect of spectral degradation on weight ratio reconstruction error (Sec. 6.3.1). However, continued characterization of the REXIS CCDs should allow for a more definite characterization of REXIS's radiation environment and the effect of radiation damage on spectral performance (especially as a function of time), which we have not accounted for here.

Internal Background

The internal background spectrum we have used here is scaled from Chandra data. In the future, a more accurate model making use of the

actual REXIS geometry should be employed to determine the fluorescent signature of the REXIS structure in response to X-rays from Bennu and the CXB. In this case, a simulation framework such as GEANT4 (Agostinelli, Allison, Amako et al. 2003) can be used to determine the intensity of X-ray emission from REXIS itself incident upon the CCDs.

Further Exploration of Error Space

While our discussion of the error space in Sec. 6.3.3 was somewhat qualitative, it is worthwhile to be more quantitative about our approach. Furthermore, we may continue to characterize how each of the model inputs discussed in Sec. 6.2 (such as molecular contamination, the OBF, and the spectral resolution) independently affect spectral performance.

Small Satellite Solar System Exploration

While the practical and modeling framework described in Chapter 6 was applied to a spacecraft-mounted instrument, it has the potential to be extended to independent, X-ray observing instruments (Hong & Romaine 2016). In particular, in the last decade, small satellites—especially so-called CubeSats (Heidt, Puig-Suari, Moore et al. 2000)—have proven very attractive in academic and commercial contexts due to their relatively low cost and the possibility of taking advantage of economies of scale to create high-throughput remote sensing fleets. Since the footprints of both REXIS and CubeSats are similar (with a typical dimension on the order of 10 cm) much of the engineering can in principle be

carried over. Asteroids and other small Solar System bodies are very attractive from a scientific perspective, providing (along with meteorites) many clues to the origin and history of the Solar System. Given their sheer number, characterizing as many of them as possible is a challenge, but it is conceivable that a fleet of X-ray spectrometers like REXIS could shed more light on the composition of asteroids.

7.3   Closing Remarks

Before 1995, our perceptions of how planets around other stars might look were limited in large part by what we knew about our Solar System and by our own imaginations. In 1977, the idea of a planet orbiting a binary star system was etched into the public consciousness with the memorable image in *Star Wars* of a double sunset on Tatooine. And the discovery of the first hot Jupiters was met with skepticism, since gas giants orbiting so close to their host stars was contrary to the picture of planet formation and evolution that had been developed for the Solar System. Since that time, we have discovered planets not only around binaries, but also around ternary star systems (Wagner, Apai, Kasper et al. 2016), and the characteristics of the hot Jupiter population—which is very real—have led to considerable insight into the formation and evolution of exoplanetary systems (e.g., Dawson & Murray-Clay 2013). It seems as if each day brings new and exciting results, many of which challenge what we thought we knew about planets, planetary systems,

and how they form.

Ultimately, however, the differences between imagination, possibility, and reality are mitigated by physics. Observations and theoretical studies of exoplanetary systems and our own Solar System—as well as the work presented in this thesis—have shown us that the possible outcomes of planet formation are relatively broad: the diversity that we see in planetary systems is somewhat expected. On the other hand, while the physical potential outcomes of planet formation may be broad, our own desire to answer existential questions requires us to understand which planet formation scenarios are *most likely*. The hope is that this thesis has, in some part, contributed towards that understanding. In this way, we will be able to understand how common systems like our Solar System and how common planets like the Earth (and the life it harbors) really are.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A

Modeling Bennu

A.1 Fluorescence

We model Bennu as a sphere of 280 m radius, with OSIRIS-REx viewing Bennu in a terminator orbit at 1 km from the asteroid barycenter. Denote a point on Bennu by P, the center of REXIS's detector plane by R, and the center of mass of the Sun by S. Let the angle between surface normal at P and $\overline{SP}$ be given by ψ_{in}, and that between the surface normal at P and $\overline{PR}$ by ψ_{out}. Then the intensity for the kth fluorescence line as measured by REXIS is given by (Jenkins 1995)

$$I_k(E_k) = \frac{\Omega_\odot}{\Omega_B} \int_{\text{Bennu}} \frac{d\Omega_B Q_k(E_k)}{4\pi \Delta E \sin \psi_{\text{in}}} \int_{E_k}^{\infty} \frac{I_\odot(E) dE}{\sum_j W_j \left[\mu_j(E) \csc \psi_{\text{in}} + \mu_j(E_k) \csc \psi_{\text{out}} \right]},$$

$$(A.1)$$

where $Q_k(E_k)$ is a factor that encompasses the probability and quantum yield associated with emission of the kth line. $\Omega_\odot$ is the solid angle sub-

tended by the Sun with respect to Bennu, ΔE is an arbitrarily chosen energy bin, and Ω_B is the solid angle subtended by Bennu with respect to REXIS. If we assume all incident Solar X-rays are parallel to one another (valid since the Sun can effectively be considered a point source with respect to Bennu), then ψ_{in}, ψ_{out}, and $d\Omega_B$ can be related to one another by straightforward geometry, and an integration over the entire surface area of Bennu amounts to an integration over all the ψ_{out} within the REXIS field of view.

If W_k is the weight fraction of the element associated with the kth line, r_k is the jump ratio, ω_k is the fluorescence yield, and f_k is the fraction of the series to which k belongs that is devoted to k, we have

$$Q_k(E_k) = W_k \frac{r_k - 1}{r_k} \omega_k f_k. \tag{A.2}$$

Since fluorescent intensity is monochromatic, the total contribution to the spectrum from fluorescence is the sum of all the I_k. An example of the calculation of the line/probability factor for the Fe-K series is given in Table A.1.

Table A.1: Series information for Fe-K and calculation of associated line probability/yield factor, $Q_k(E_k)$.

Edge/Series	r_k	ω_k	Line	E_k [eV]	f_k	Q_k/W_k
Fe-K	0.351	7.893	Kα_3	6,267.40	2.76096×10^{-4}	7.649×10^{-4}
			Kα_2	6,392.10	2.94023×10^{-1}	8.145×10^{-1}
			Kα_1	6,405.20	5.80277×10^{-1}	1.608
			Kβ_3	7,059.30	4.25566×10^{-2}	1.179×10^{-1}
			Kβ_1	7,059.30	8.21556×10^{-2}	2.276×10^{-1}
			Kβ_5	7,110.00	7.12115×10^{-4}	1.973×10^{-3}

A.2 Coherent scattering

The continuous spectrum due to coherent scattering is given by

$$I_{\text{scattering}}(E) = \frac{1}{\Omega_B} \int_\Omega \frac{d\sigma}{d\Omega} I_\odot(E) N_A d\Omega_B, \tag{A.3}$$

where N_A is Avogadro's number and the integration effected over the solid angle Ω. The differential scattering cross-section $d\sigma/d\Omega$ is given by

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{4\pi} \left(1 - \cos^2 \theta\right) |F(E,\theta)|^2, \tag{A.4}$$

where r_e is the classical electron radius, given by 2.82×10^{-15} m, and where $|F(E,\theta)|^2$ the modulus squared of the (complex) atomic form factor $F(E,\theta)$ for the element in question, dependent upon both the energy of the incident radiation and the scattering angle $\theta \equiv |\psi_{\text{in}} + \psi_{\text{out}}|$. Incoherent scattering, being at least an order of magnitude smaller than coherent scattering for all energies and elements of interest, is not considered. The total intensity of radiation emitted by the asteroid (in units of photons/Sr/s/eV/cm^2) is given by the sum of fluorescent and scattered radiation:

$$I(E) = \sum_k I_k(E_k) + I_{\text{scattering}}(E). \tag{A.5}$$

283

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix B

Modeling the Solar Spectrum

Here we briefly discuss how the Solar corona is modeled in order to determine its X-ray spectrum. Denote the power per unit volume emitted by a plasma undergoing an atomic transition from quantum states $j \rightarrow i$ by P_{ij}, and the wavelength (or equivalently, the energy) associated with this transition by λ_{ij}. Then the intensity $I_\odot(\lambda_{ij})$ of the radiation at the surface of the body of interest (say Bennu) for this transition is given by

$$I_\odot(\lambda_{ij}) = \frac{1}{4\pi R^2} \int_V P_{ij} dV, \qquad (B.1)$$

where R is the distance from the Sun to Bennu and V is the plasma volume. P_{ij} can be written (Golub & Pasachoff 1997)

$$P_{ij} = 0.8 A_X G(T, \lambda_{ij}) \frac{h_P c}{\lambda_{ij}} N_e^2, \qquad (B.2)$$

where N_e is the local electron density, A_X is the abundance of the Xth element with respect to hydrogen, and G is the so-called contribution function (not to be confused with the grasp), which is a function of both the plasma temperature T (not to be confused with the detector temperature) and the wavelength λ_{ij} associated with the transition. h_{P} is the Planck constant, and c the speed of light.

P_{ij} is dependent upon both temperature and electron density, and it is possible to decompose it into density and temperature dependent parts. We define the "differential emission measure", or DEM, as follows

$$\int_V N_e^2 dV = \int_T \mathrm{DEM}(T) dT, \qquad (\mathrm{B}.3)$$

so that

$$P_{ij} = 0.8 A_X G(T, \lambda_{ij}) \frac{h_{\mathrm{P}} c}{\lambda_{ij}} \mathrm{DEM}(T) \frac{dT}{dV}. \qquad (\mathrm{B}.4)$$

On the left-hand side of Eq. (B.3), the integration is effected over the plasma volume; on the right-hand side, over the possible coronal temperatures. Defining

$$\phi(T, \lambda) \equiv \sum_X \sum_{ij} 0.8 A_X G(T, \lambda_{ij}) \frac{h_{\mathrm{P}} c}{\lambda_{ij}} \qquad (\mathrm{B}.5)$$

(the first sum extending over all species and the second sum extending

over all transition pairs), we have, using Eq. (B.3),

$$I_\odot(\lambda) = \int_T \phi(T, \lambda) \text{DEM}(T) dT. \tag{B.6}$$

Instead of integrating over all possible coronal temperatures, in certain cases, it is possible to consider only those coronal temperatures at which the emission measure is greatest. For quiet Solar regions, it has been found that the emission measure peaks at one value, while the DEM for active regions tends to peak at two values. Thus we may write

$$I_\odot(\lambda) = \phi(T_1, \lambda) \text{EM}(T_1) \qquad \text{(quiet Sun)} \tag{B.7}$$

$$I_\odot(\lambda) = \phi(T_1, \lambda) \text{EM}(T_1) + \phi(T_2, \lambda) \text{EM}(T_2) \qquad \text{(active Sun)} \tag{B.8}$$

where $\text{EM}(T)$ is a single temperature emission measure, which for all practical purposes here amounts to a simple numerical prefactor.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix C

Definition of Accuracy

REXIS's spectral performance requirement states that the reconstructed weight ratios of the asteroid regolith be within a certain percent of that of the baseline composition. REXIS itself can only measure count ratios, and we use the calibration curves to make the correspondence between count ratios and weight ratios. For convenience, we denote this function CC : count ratio $\rightarrow$ weight ratio, with the inverse mapping CC^{-1} : weight ratio $\rightarrow$ count ratio.

Fig. C-1 demonstrates REXIS's counting procedure in forming count ratios. For simplicity, the figure shows only two spectral features I_1 and I_2, shown in blue and red, respectively. First, we define counting zones centered about each line center. The widths of these zones are given by the full-width half-maximum FWHM of the Gaussian centered at each energy. All the counts within each zone are considered to be from the respective line, although in reality, there will be some contamination

from other spectral features. Thus in our simplified case, the total number of counts in FWHM_1 will include contributions from I_1, denoted C_{11}, and those from I_2, denoted C_{12}. Likewise, the total number of counts in FWHM_2 will include contributions from I_2 (given by C_{22}) and I_1 (given by C_{21}). The total count ratio of the first feature to the second $\rho_{1/2}$ is then given by

$$\rho_{1/2} = \frac{C_{11} + C_{12}}{C_{22} + C_{21}}. \tag{C.1}$$

In the more general case applicable to REXIS, we have the total number of binned detector counts given by $C_3'(E')$, so that for the $\rho_{k/\text{Si}}$ line ratio,

$$\rho_{k/\text{Si}} = \frac{\sum_{E_k - \text{FHWM}_k/2}^{E_k + \text{FHWM}_k/2} C_3'(E')}{\sum_{E_\text{Si} - \text{FHWM}_\text{Si}/2}^{E_\text{Si} + \text{FHWM}_\text{Si}/2} C_3'(E')}. \tag{C.2}$$

We map $\rho_{k/\text{Si}}$ to the equivalent weight ratio $\varpi_{k/\text{Si}}$ by using the calibration curve: $\varpi_{k/\text{Si}} = \text{CC}\left(\rho_{k/\text{Si}}\right)$. The accuracy error is then calculated by comparing the measured weight ratio $\varpi_{k/\text{Si}}$ with the regolith input weight ratio $\varpi_{k/\text{Si},0}$ corresponding to a CI chondrite-like composition. If we denote the weight ratio error requirement by η, then we may write the requirement as

$$\left| 1 - \frac{\varpi_{k/\text{Si}}}{\varpi_{k/\text{Si},0}} \right| \leq \eta, \tag{C.3}$$
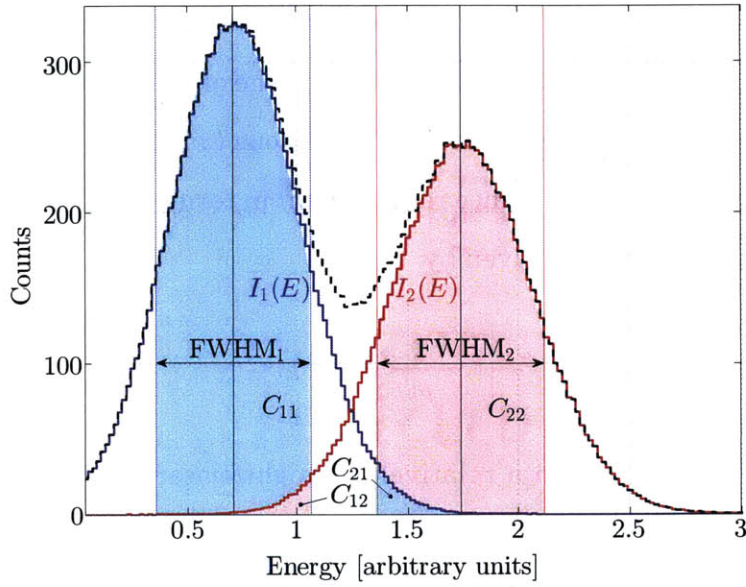
where $\eta = 0.25$.

Figure C-1: Example of line contamination in the simplified case of only two lines, $I_1(E)$ (blue) and $I_2(E)$ (red); their sum is in black. The counts $C_{11} + C_{12}$ account for all the counts under the sum in the FWHM zone of line 1 and the counts $C_{22} + C_{21}$ are all the counts in the FWHM zone of line 2. C_{11} are the counts that are due to line 1 in the line 1 FWHM zone, and C_{22} are the counts due to line 2 in the line 2 FWHM zone, while C_{12} is the contamination from line 2 into the FWHM zone of line 1, and C_{21} the contamination of line 1 into the FWHM zone of line 2.

In some cases (e.g., Appendix D), we may wish to consider only the errors in count ratios. In this case, we consider the expected count ratio from the lines of interest, ignoring contamination and other effects. If we denote this ratio by $\rho_{k/\mathrm{Si},0}$, we have

$$\rho_{k/\mathrm{Si},0} = \frac{C_1(E_k)}{C_1(E_{\mathrm{Si}})}, \tag{C.4}$$

where care has been taken to ensure that the effect of quantum effi-

ciency has been accounted for. Indeed, when the calibration curves are generated, $\rho_{k/\mathrm{Si},0}$ is calculated for a whole range of baseline compositions, and a second-order fit performed on the various $(\varpi_{k/\mathrm{Si},0}, \rho_{k/\mathrm{Si},0})$ pairs (Fig. 6-11). The REXIS performance requirement in terms of count ratio is denoted by η_C, and is then given by

$$\left| 1 - \frac{\rho_{k/\mathrm{Si}}}{\rho_{k/\mathrm{Si},0}} \right| \leq \eta_C. \tag{C.5}$$

η_C can be related to η in a relatively straightforward way that is most clearly demonstrated graphically by means of the shaded region shown in Fig. 6-13; it is given mathematically by

$$\left| 1 - \frac{\mathrm{CC}^{-1}\left[(1-\eta) \times \varpi_{k/\mathrm{Si},0}\right]}{\mathrm{CC}^{-1}\left(\varpi_{k/\mathrm{Si},0}\right)} \right| = \eta_C. \tag{C.6}$$

Appendix D

Statistical Error

Suppose for a given detector temperature that the total error in counts due to systematic error is Δ, that the requirement is given by η_C, and that statistical error is given by σ. We suppose that the errors can be summed quadratically:

$$\eta_C = \sqrt{\Delta^2 + \sigma^2},\tag{D.1}$$

so that

$$\sigma = \sqrt{\eta_C^2 - \Delta^2}.\tag{D.2}$$

For the kth line and an N confidence level,

$$\frac{\sigma^2}{N^2} = \frac{\dot{N}_k/T_{\text{tot}} + \dot{N}_{\text{CXB},k}/T_{\text{CXB}} + \dot{N}_{\text{int},k}/T_{\text{int}}}{\left(\dot{N}_k - \dot{N}_{\text{CXB},k} - \dot{N}_{\text{int},k}\right)^2} + \frac{\dot{N}_{\text{Si}}/T_{\text{tot}} + \dot{N}_{\text{CXB},\text{Si}}/T_{\text{CXB}} + \dot{N}_{\text{int},\text{Si}}/T_{\text{int}}}{\left(\dot{N}_{\text{Si}} - \dot{N}_{\text{CXB},\text{Si}} - \dot{N}_{\text{int},\text{Si}}\right)^2}.$$

$$\tag{D.3}$$

T_{CXB} and T_{int} are the CXB and internal calibration times (see Sec. 6.2.3). $\dot{N}_k$, $\dot{N}_{\mathrm{CXB},k}$, and $\dot{N}_{\mathrm{int},k}$ refer respectively to the total count rates, CXB count rates, and internal background count rates within the kth FWHM counting zone. More precisely, $\dot{N}_k$, $\dot{N}_{\mathrm{CXB},k}$, and $\dot{N}_{\mathrm{int},k}$ are given by C'_3/T_{obs}, $C'_{\mathrm{CXB}}/T_{\mathrm{obs}}$, and $C'_{\mathrm{int}}/T_{\mathrm{obs}}$ summed over each FWHM zone. Rearranging Eq. D.3, we get

$$\frac{\sigma^2}{N^2} = \underbrace{\left[\frac{\dot{N}_k}{\left(\dot{N}_k - \dot{N}_{\mathrm{CXB},k} - \dot{N}_{\mathrm{int},k} \right)^2} + \frac{\dot{N}_{\mathrm{Si}}}{\left(\dot{N}_{\mathrm{Si}} - \dot{N}_{\mathrm{CXB},\mathrm{Si}} - \dot{N}_{\mathrm{int},\mathrm{Si}} \right)^2} \right] \frac{1}{T_{\mathrm{tot}}}}_{\equiv L} +$$

$$\underbrace{\frac{\dot{N}_{\mathrm{CXB},k}/T_{\mathrm{CXB}} + \dot{N}_{\mathrm{int},k}/T_{\mathrm{int}}}{\left(\dot{N}_k - \dot{N}_{\mathrm{CXB},k} - \dot{N}_{\mathrm{int},k} \right)^2} + \frac{\dot{N}_{\mathrm{CXB},\mathrm{Si}}/T_{\mathrm{CXB}} + \dot{N}_{\mathrm{int},\mathrm{Si}}/T_{\mathrm{int}}}{\left(\dot{N}_{\mathrm{Si}} - \dot{N}_{\mathrm{CXB},\mathrm{Si}} - \dot{N}_{\mathrm{int},\mathrm{Si}} \right)^2}}_{\equiv R} . \qquad \text{(D.4)}$$

With L and R defined as above, the total observation time T_{obs} required for $N\sigma$ confidence is given by

$$T_{\mathrm{obs}} = \frac{L}{R - \sigma^2/N^2} . \qquad \text{(D.5)}$$

A summary of the expected count rates within each FWHM zone is given in Table D.1.

Table D.1: Summary of expected count rate with detector temperature $T = -60$ °C. E_- and E_+ are the lower and upper limits to the FWHM zone for each line, respectively. $\dot{N}_k$ is the total number of counts in each FWHM zone (i.e., from Bennu and background), while $\dot{N}_{\text{CXB},k}$ and $\dot{N}_{\text{int},k}$ are CXB and internal count rates, respectively, in each zone.

Line	E_- [keV]	E_+[keV]	$\dot{N}_k$	$\dot{N}_{\text{int},k}$	$\dot{N}_{\text{CXB},k}$
Fe-L	0.658	0.751	4.73	0.038	3.73
Mg-K	1.1830	1.3230	2.64	0.0291	1.69
Si-K	1.6620	1.8139	1.46	0.0745	0.948
S-K	2.2250	2.3910	0.967	0.0296	0.888

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix E

Calculating the Energy Resolution of the Detector

To determine the energy and detector temperature dependence of the detector resolution FWHM, we require two pieces of experimental data: FWHM as a function of energy E at a fixed temperature T_0, and FWHM as a function of temperature T at a fixed energy E_0. The two pieces of information can be combined then to determine the general dependence of FWHM on E and T:[*]

$$\text{FWHM}(E, T) = \sqrt{\text{FWHM}^2(E, T_0) + \text{FWHM}^2(E_0, T) - \text{FWHM}^2(E_0, T_0)}$$

(E.1)

In the case of REXIS, energy resolution for the CCID-41 has been experimentally determined as a function of energy at $T_0 = -90\ ^\circ\text{C}$,[†] and as a

---

[*]Personal communication, M. Bautz.
[†]Personal communication, M. Bautz.

function of temperature at $E_0 = 5.89$ keV.[‡] These two pieces of information together allow us to use Eq. E.1 to generate Fig. 6-7.

[‡]Personal communication, S. Kissel.

Bibliography

Adams, E. R., Seager, S., & Elkins-Tanton, L. 2008, Astrophys. J., 673, 1160. 0710.4941

Agnor, C. B., Canup, R. M., & Levison, H. F. 1999, Icarus, 142, 219

Agostinelli, S., Allison, J., Amako, K. a., et al. 2003, Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 506, 250

Akeson, R. L., Chen, X., Ciardi, D., et al. 2013, Pub. Ast. Soc. Aus., 125, 989. 1307.2944

Alfè, D., Gillan, M. J., & Price, G. D. 2002, Earth and Planetary Science Letters, 195, 91

Allen, B., Grindlay, J., Hong, J., et al. 2013, in SPIE Optical Engineering+ Applications, 88400M–88400M (International Society for Optics and Photonics)

ALMA Science Portal Accessed: 2016-05-01, The alma science portal, http://almascience.nrao.edu/. Accessed: 2016-05-01

AMPTEK 2014, Xr-100sdd silicon drift detector (sdd). http://www.amptek.com/products/xr-100sdd-silicon-drift-detector

Anders, E., & Grevesse, N. 1989, Geochimica Cosmochimica Acta, 53, 197

Ansdell, M., Williams, J. P., van der Marel, N., et al. 2016, ArXiv e-prints. 1604.05719

Armitage, P. J. 2013, Astrophysics of Planet Formation

Asphaug, E. 2010, Chemie der Erde / Geochemistry, 70, 199

Balbus, S. A., & Hawley, J. F. 1991, Astrophys. J., 376, 214

Barros, S. C. C., Almenara, J. M., Demangeon, O., et al. 2015, Mon. Not. R. Astron. Soc., 454, 4267. 1510.01047

Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, Astrophys. J. Supp., 204, 24. 1202.5852

Benz, W., & Asphaug, E. 1999, Icarus, 142, 5. astro-ph/9907117

Binzel, R. P., Lupishko, D., di Martino, M., et al. 2002, Physical Properties of Near-Earth Objects, 255–271

Blum, J., & Wurm, G. 2008, Ann. Rev. Astron. Astrophys., 46, 21

Bodenheimer, P., Laughlin, G. P., Rózyczka, M., et al. (eds.) 2007, Numerical Methods in Astrophysics: An Introduction (Taylor & Francis)

Bodenheimer, P., & Lissauer, J. J. 2014, Astrophys. J., 791, 103. 1407.1433

Bodenheimer, P., & Pollack, J. B. 1986, Icarus, 67, 391

Borucki, W. J., Koch, D. G., Basri, G., et al. 2011, Astrophys. J., 736, 19. 1102.0541

Bottke, W. F., Jr., Vokrouhlický, D., Rubincam, D. P., et al. 2006, Annual Review of Earth and Planetary Sciences, 34, 157

Bralower, H. 2013, Mechanical design, calibration, and environmental protection of the REXIS DAM, Master's thesis, Massachusetts Institute of Technology

Britt, D. T., Yeomans, D., Housen, K., et al. 2002, Asteroid Density, Porosity, and Structure, 485–500

Buckingham, E. 1914, Physical Review, 4, 345

Bus, S. J., Vilas, F., & Barucci, M. A. 2002, Visible-Wavelength Spectroscopy of Asteroids, 169–182

Butler, R. P., Marcy, G. W., Williams, E., et al. 1997, Astrophys. J., 474, L115

Campins, H., Morbidelli, A., Tsiganis, K., et al. 2010, The Astrophysical Journal Letters, 721, L53

Canup, R. M. 2012, Science, 338, 1052

Caroli, E., Stephen, J. B., Di Cocco, G., et al. 1987, Space Science Reviews, 45, 349

Carter, J. A., Agol, E., Chaplin, W. J., et al. 2012, Science, 337, 556. 1206.4718

Chambers, J. E. 2001, Icarus, 152, 205

Chambers, J. E., & Wetherill, G. W. 1998, Icarus, 136, 304

Chandra X-ray Center, Chandra Project Science MSFC, & Chandra IPI Teams 2014, Acis: Advanced ccd imaging spectrometer. http://cxc.harvard.edu/proposer/POG/html/chap6.html

Chandrasekhar, S. 1960, Principles of stellar dynamics

Chapman, C. R. 1996, Meteoritics and Planetary Science, 31, 699

Chassefière, E., & Leblanc, F. 2004, Planet. Space Sci., 52, 1039

Chiang, E., & Laughlin, G. 2013, Mon. Not. R. Astron. Soc., 431, 3444. 1211.1673

Clark, B. E., Binzel, R. P., Howell, E. S., et al. 2011, Icarus, 216, 462

Cohen, O., Drake, J. J., Glocer, A., et al. 2014, Astrophys. J., 790, 57. 1405.7707

Cossou, C., Raymond, S. N., Hersant, F., et al. 2014, Astron. Astrophys., 569, A56. 1407.6011

Courant, R., Friedrichs, K., & Lewy, H. 1928, Mathematische Annalen, 100, 32

Cowling, T. G. 1934, Mon. Not. R. Astron. Soc., 94, 768

Ćuk, M., & Stewart, S. T. 2012, Science, 338, 1047

D'Alessio, P., Calvet, N., & Hartmann, L. 2001, Astrophys. J., 553, 321. astro-ph/0101443

Dawson, R. I., Chiang, E., & Lee, E. J. 2015, Mon. Not. R. Astron. Soc., 453, 1471. 1506.06867

Dawson, R. I., & Murray-Clay, R. A. 2013, Astrophys. J., 767, L24. 1302.6244

de Pater, I., & Lissauer, J. J. 2015, Planetary Sciences

Deck, K. M., Payne, M., & Holman, M. J. 2013, Astrophys. J., 774, 129. 1307.8119

DeMeo, F. E., Binzel, R. P., Slivan, S. M., et al. 2009, Icarus, 202, 160

Dere, K., Landi, E., Mason, H., et al. 1997, Astronomy and Astrophysics Supplement Series, 125, 149

Dressing, C. D., Charbonneau, D., Dumusque, X., et al. 2015, Astrophys. J., 800, 135. 1412.8687

Dupree, A. K., Huher, M. C. E., Noyes, R. W., et al. 1973, Astrophysical Journal, 182, 321

Eddington, A. S. 1926, The Internal Constitution of the Stars

Elam, W., Ravel, B., & Sieber, J. 2002, Radiation Physics and Chemistry, 63, 121

Fermi, E. 1956, Thermodynamics (New York: Dover, 1956)

Freeland, S., & Handy, B. 1998, Solar Physics, 182, 497

Fressin, F., Torres, G., Charbonneau, D., et al. 2013, Astrophys. J., 766, 81. 1301.0842

Fressin, F., Torres, G., Rowe, J. F., et al. 2012, Nature, 482, 195. 1112.4550

Fu, R. R., Weiss, B. P., Shuster, D. L., et al. 2012, Science, 338, 238

Galli, D., & Shu, F. H. 1993, Astrophys. J., 417, 220

Gammie, C. F. 2001, Astrophys. J., 553, 174. astro-ph/0101501

Genda, H., & Abe, Y. 2003, Icarus, 164, 149

Gilliland, R. L., Marcy, G. W., Rowe, J. F., et al. 2013, Astrophys. J., 766, 40. 1302.2596

Ginzburg, S., Schlichting, H. E., & Sari, R. 2015, ArXiv e-prints. 1512.07925

Goldreich, P., Lithwick, Y., & Sari, R. 2004, Ann. Rev. Astron. Astrophys., 42, 549. astro-ph/0405215

Goldreich, P., & Schlichting, H. E. 2014, Astron. J., 147, 32. 1308.4688

Goldreich, P., & Ward, W. R. 1973, Astrophys. J., 183, 1051

Golub, L., & Pasachoff, J. 1997, The Solar Corona (Cambridge University Press)

Gu, P.-G., Lin, D. N. C., & Bodenheimer, P. H. 2003, Astrophys. J., 588, 509. astro-ph/0303362

Haisch, K. E., Jr., Lada, E. A., & Lada, C. J. 2001, Astrophys. J., 553, L153. astro-ph/0104347

Halliday, A. N. 2008, Philosophical Transactions of the Royal Society of London Series A, 366, 4163

Hansen, B. M. S., & Murray, N. 2012, Astrophys. J., 751, 158. 1105.2050

Hartigan, P., Edwards, S., & Ghandour, L. 1995, Astrophys. J., 452, 736

Hartmann, L. 2009, Accretion Processes in Star Formation: Second Edition (Cambridge University Press)

Hayashi, C. 1981, Progress of Theoretical Physics Supplement, 70, 35

Heidt, H., Puig-Suari, J., Moore, A., et al. 2000

Helled, R., Anderson, J. D., Podolak, M., et al. 2011, Astrophys. J., 726, 15. 1010.5546

Henyey, L., Vardya, M. S., & Bodenheimer, P. 1965, Astrophys. J., 142, 841

Hillenbrand, L. A. 2005, in A Decade of Discovery: Planets around other stars, STScI Symposium Series, vol. 19, ed. M. Livio

Hong, J., & Romaine, S. 2016, Earth, Planets, and Space, 68, 35. 1601. 01300

Housen, K. R., & Holsapple, K. A. 2003, Icarus, 163, 102

Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, Astrophys. J. Supp., 201, 15. 1103.2541

Howard, A. W., Marcy, G. W., Johnson, J. A., et al. 2010, Science, 330, 653. 1011.0143

Howe, A. R., & Burrows, A. 2015, Astrophys. J., 808, 150. 1505.02784

Howe, A. R., Burrows, A., & Verne, W. 2014, Astrophys. J., 787, 173. 1402.4818

Hubbell, J. H., & Seltzer, S. M. 1996, National Institute of Standards and Technology

Hunten, D. M. 1973, Journal of Atmospheric Sciences, 30, 1481

Ida, S., & Lin, D. N. C. 2004, Astrophys. J., 604, 388. astro-ph/ 0312144

Iglesias, C. A., & Rogers, F. J. 1996, Astrophys. J., 464, 943

Ikoma, M., & Hori, Y. 2012, Astrophys. J., 753, 66. 1204.5302

Inamdar, N. K. 2013, Radiation damage for rexis ccds. REXIS internal whitepaper

Inamdar, N. K., Binzel, R. P., Hong, J. S., et al. 2014, in Imaging Spectrometry XIX, Proc. SPIE, vol. 9222, 922207. 1410.3051

Inamdar, N. K., & Schlichting, H. E. 2015, Mon. Not. R. Astron. Soc., 448, 1751. 1412.4440

Inamdar, N. K., & Schlichting, H. E. 2016, Astrophys. J., 817, L13. 1510.02090

International Astronomical Union Accessed: 2016-05-23, Minor planet center, http://www.minorplanetcenter.net/data. Accessed: 2016-05-23

Jenkins, R. 1995, Quantitative X-ray Spectrometry (CRC Press)

Jones, M. P., Smith, M. J., & Masterson, R. A. 2014, in SPIE Optical Engineering+ Applications (International Society for Optics and Photonics)

Jontof-Hutter, D., Rowe, J. F., Lissauer, J. J., et al. 2015, Nature, 522, 321. 1506.07067

Kippenhahn, R., Weigert, A., & Weiss, A. 2012, Stellar Structure and Evolution

Kissel, L. 2000, Rtab: the rayleigh scattering database

Koyama, K., Tsunemi, H., Dotani, T., et al. 2007, Publications of the Astronomical Society of Japan, 59, S23

Koyama, K., Tsunemi, H., Dotani, T., et al. 2007, Publications of the Astronomical Society of Japan, 59, 23

Kretke, K. A., & Levison, H. F. 2014, ArXiv e-prints. 1409.4430

Lagrange, A.-M., Bonnefoy, M., Chauvin, G., et al. 2010, Science, 329, 57. 1006.3314

Lambrechts, M., & Johansen, A. 2012, Astron. Astrophys., 544, A32. 1205.3030

Lammer, H., Güdel, M., Kulikov, Y., et al. 2012, Earth, Planets, and Space, 64, 179

Landau, L. D., & Lifshitz, E. M. 1959, Fluid mechanics (Butterworth Heinemann)

Landi, E., & Drago, F. C. 2008, The Astrophysical Journal, 675, 1629

Landi, E., Young, P., Dere, K., et al. 2013, The Astrophysical Journal, 763, 86

Lauretta, D. S., & OSIRIS-REx Team 2012, in Lunar and Planetary Science Conference, Lunar and Planetary Science Conference, vol. 43, 2491

Lee, E. J., & Chiang, E. 2015, Astrophys. J., 811, 41. 1508.05096

Lee, E. J., Chiang, E., & Ormel, C. W. 2014, ArXiv e-prints. 1409.3578

Leinhardt, Z. M., & Stewart, S. T. 2012, Astrophys. J., 745, 79. 1106.6084

Lewis, J. S. 1972, Icarus, 16, 241

Lim, L. F., & Nittler, L. R. 2009, Icarus, 200, 129

Lissauer, J. J. 1993, Ann. Rev. Astron. Astrophys., 31, 129

Lissauer, J. J., Fabrycky, D. C., Ford, E. B., et al. 2011, Nature, 470, 53. 1102.0291

Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., et al. 2011, Astrophys. J. Supp., 197, 8. 1102.0543

Liu, S.-F., Hori, Y., Lin, D. N. C., et al. 2015, Astrophys. J., 812, 164. 1509.05772

Lodders, K., & Fegley, B. 1998, The Planetary Scientist's Companion

Lopez, E., & Fortney, J. J. 2013a, in American Astronomical Society Meeting Abstracts #221, American Astronomical Society Meeting Abstracts, vol. 221, #333.04

Lopez, E. D., & Fortney, J. J. 2013b, Astrophys. J., 776, 2. 1305.0269

Lopez, E. D., & Fortney, J. J. 2014, Astrophys. J., 792, 1. 1311.0329

Lopez, E. D., Fortney, J. J., & Miller, N. 2012, Astrophys. J., 761, 59. 1205.0010

Love, S. G., & Ahrens, T. J. 1996, Icarus, 124, 141

Lumb, D., Warwick, R., Page, M., et al. 2002, arXiv preprint astro-ph/0204147

Mamajek, E. E. 2009, in American Institute of Physics Conference Series, American Institute of Physics Conference Series, vol. 1158, eds. T. Usuda, M. Tamura, & M. Ishii, 3–10. 0906.5011

Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014, Astrophys. J. Supp., 210, 20. 1401.4195

Mastrobuono-Battisti, A., Perets, H. B., & Raymond, S. N. 2015, Nature, 520, 212. 1502.07346

Mayor, M., & Queloz, D. 1995, Nature, 378, 355

McCord, T. B., & Sotin, C. 2005, Journal of Geophysical Research (Planets), 110, E05009

Mewe, R., Gronenschild, E., & Van Den Oord, G. 1985, Astronomy and Astrophysics Supplement Series, 62, 197

Meyer, J.-P. 1985, The Astrophysical Journal Supplement Series, 57, 173

Milani, A., Chesley, S. R., Sansaturio, M. E., et al. 2009, Icarus, 203, 460

Mizuno, H. 1980, Progress of Theoretical Physics, 64, 544

Mohanty, S., & Basri, G. 2003, Astrophys. J., 583, 451. astro-ph/0201455

Mordasini, C., Alibert, Y., & Benz, W. 2009, Astron. Astrophys., 501, 1139. 0904.2524

Movshovitz, N., Bodenheimer, P., Podolak, M., et al. 2010, Icarus, 209, 616. 1005.3875

Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2015, Astrophys. J. Supp., 217, 31. 1502.02038

Nettelmann, N., Holst, B., Kietzmann, A., et al. 2008, Astrophys. J., 683, 1217-1228. 0712.1019

Nittler, L. R., McCoy, T. J., Clark, P. E., et al. 2004, Antarctic Meteorite Research, 17, 231

Nolan, M., Magri, C., Howell, E., et al. 2013, Asteroid (101955) bennu shape model v1.0. ear-a-i0037-5-bennushape-v1.0. NASA Planetary Data System

Pahlevan, K., & Stevenson, D. J. 2007, Earth and Planetary Science Letters, 262, 438. 1012.5323

Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, Proceedings of the National Academy of Science, 110, 19273. 1311.6806

Petit, J.-M., Chambers, J., Franklin, F., et al. 2002, Primordial Excitation and Depletion of the Main Belt, 711–723

307

Piso, A.-M. A., & Youdin, A. N. 2014, Astrophys. J., 786, 21. 1311.0011

Polishook, D. 2014, Icarus, 241, 79. 1406.3359

Pollack, J. B., Rages, K., Baines, K. H., et al. 1986, Icarus, 65, 442

Prigozhin, G. Y., Kissel, S. E., Bautz, M. W., et al. 2000, in X-Ray Optics, Instruments, and Missions III, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 4012, eds. J. E. Truemper, & B. Aschenbach, 720–730

Rafikov, R. R. 2006, Astrophys. J., 648, 666. astro-ph/0405507

Raymond, S. N., & Cossou, C. 2014, Mon. Not. R. Astron. Soc., 440, L11. 1401.3743

Raymond, S. N., Quinn, T., & Lunine, J. I. 2004, Icarus, 168, 1. astro-ph/0308159

Rogers, F. J., Swenson, F. J., & Iglesias, C. A. 1996, Astrophys. J., 456, 902

Rogers, L. A. 2015, Astrophys. J., 801, 41. 1407.4457

Rogers, L. A., Bodenheimer, P., Lissauer, J. J., et al. 2011, Astrophys. J., 738, 59. 1106.2807

Russell, C. T., Raymond, C. A., Coradini, A., et al. 2012, Science, 336, 684

Rybicki, G. B., & Lightman, A. P. 1986, Radiative Processes in Astrophysics

Ryu, K. K., Burke, B. E., Clark, H. R., et al. 2014, in SPIE Astronomical Telescopes+ Instrumentation, 91444O–91444O (International Society for Optics and Photonics)

Safronov, V. S. 1972, Evolution of the protoplanetary cloud and formation of the earth and planets. (Keter Publishing House.)

Saha, M. N. 1921, Proceedings of the Royal Society of London Series A, 99, 135

Sánchez, P., & Scheeres, D. J. 2014, Meteoritics and Planetary Science, 49, 788. 1306.1622

Sanz-Forcada, J., Micela, G., Ribas, I., et al. 2011, Astron. Astrophys., 532, A6. 1105.0550

Sari, R. 2006, Physics of Fluids, 18, 027106. astro-ph/0505174

Saumon, D., Chabrier, G., & van Horn, H. M. 1995, Astrophys. J. Supp., 99, 713

Schlichting, H. E. 2014, Astrophys. J., 795, L15. 1410.1060

Schlichting, H. E., Sari, R., & Yalinewich, A. 2015, Icarus, 247, 81. 1406.6435

Schmidt, S. J., Cruz, K. L., Bongiorno, B. J., et al. 2007, Astron. J., 133, 2258. astro-ph/0701055

Schneider, J., Dedieu, C., Le Sidaner, P., et al. 2011, Astron. Astrophys., 532, A79. 1106.0586

Seager, S., Kuchner, M., Hier-Majumder, C. A., et al. 2007, Astrophys. J., 669, 1279. 0707.2895

Sedov, L. I. 1959, Similarity and Dimensional Methods in Mechanics (New York: Academic Press, 1959)

Segura, A., Walkowicz, L. M., Meadows, V., et al. 2010, Astrobiology, 10, 751. 1006.0022

Sekiya, M. 1998, Icarus, 133, 298

Shu, F. H. 1977, Astrophys. J., 214, 488

Steffen, J. H., Fabrycky, D. C., Agol, E., et al. 2013, Mon. Not. R. Astron. Soc., 428, 1077. 1208.3499

Stevenson, D. J. 1982, Planet. Space Sci., 30, 755

Stevenson, D. J., & Salpeter, E. E. 1977, Astrophys. J. Supp., 35, 221

Stewart, S. T., Lock, S. J., & Mukhopadhyay, S. 2014a, in Lunar and Planetary Science Conference, Lunar and Planetary Inst. Technical Report, vol. 45, 2869

Stewart, S. T., Lock, S. J., & Mukhopadhyay, S. 2014b, AGU Fall Meeting Abstracts

Taylor, G. 1950, Proceedings of the Royal Society of London Series A, 201, 159

Thompson, A. C., & Vaughan, D. (eds.) 2001, X-ray Data Booklet (Lawrence Berkeley National Laboratory, University of California), second edn.

Vardya, M. S. 1960, Astrophys. J. Supp., 4, 281

Vokrouhlický, D., Bottke, W. F., Chesley, S. R., et al. 2015, The Yarkovsky and YORP Effects, 509–531

Wagner, K., Apai, D., Kasper, M., et al. 2016, ArXiv e-prints. 1607.02525

Walsh, K. J., & Morbidelli, A. 2011, Astron. Astrophys., 526, A126. 1101.3776

Wang, H., Weiss, B. P., Downey, B. G., et al. 2015, in Lunar and Planetary Science Conference, Lunar and Planetary Science Conference, vol. 46, 2516

Ward, W. R. 1997, Icarus, 126, 261

Watson, A. J., Donahue, T. M., & Walker, J. C. G. 1981, Icarus, 48, 150

Weidenschilling, S. J. 1977, Astrophys. Space Sci., 51, 153

Weisberg, M. K., McCoy, T. J., & Krot, A. N. 2006, Systematics and Evaluation of Meteorite Classification, 19–52

Weiss, B. P., Berdahl, J. S., Elkins-Tanton, L., et al. 2008, Science, 322, 713

Weiss, L. M., & Marcy, G. W. 2014, Astrophys. J., 783, L6. 1312.0936

Wetherill, G. W. 1989, in Asteroids II, eds. R. P. Binzel, T. Gehrels, & M. S. Matthews, 661–680

Winn, J. N. 2010, Exoplanet Transits and Occultations, 55–77

Wyatt, M. C., & Jackson, A. P. 2016, Space Sci. Rev.. 1603.04857

Youdin, A. N., & Kenyon, S. J. 2013, From Disks to Planets, 1

Youdin, A. N., & Shu, F. H. 2002, Astrophys. J., 580, 494. astro-ph/0207536

Zel'dovich, Y. B., & Raizer, Y. P. 1967, Physics of shock waves and high-temperature hydrodynamic phenomena (Dover)

Zhu, Z., Hartmann, L., & Gammie, C. 2009, Astrophys. J., 694, 1045. 0811.1762

Zombeck, M. 1990, Handbook of Space Astronomy and Astrophysics, vol. 104 (Cambridge University Press Cambridge)