Orbits and Interiors of Planets

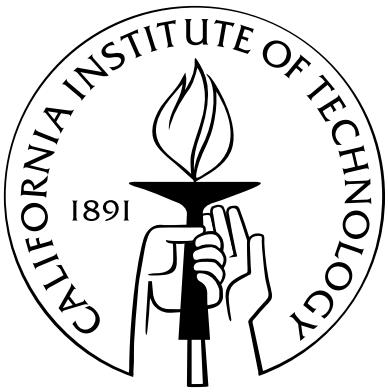
Thesis by

Konstantin Batygin

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

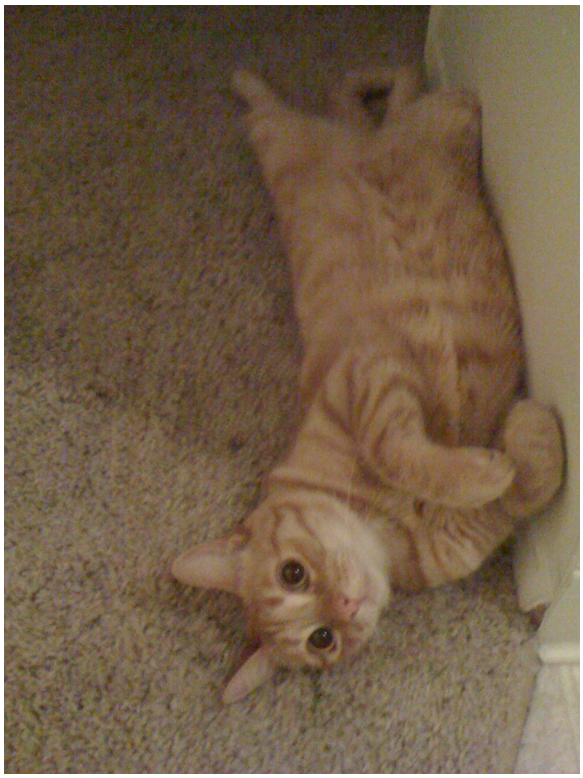
Pasadena, California

2012

(Defended April 4^{th}, 2011)

I dedicate this thesis to my cat, Orion. You are a purr-fect combination of smart and furry.

In other words, you are everything I look for in a cat.



Acknowledgements

I wish to begin by expressing my deep gratitude towards my advisors, Mike Brown and Dave Stevenson (listed in alphabetical order by last name). Their guidance has been nothing but outstanding, while their creativity, scientific approach and enthusiasm have been amazing sources of inspiration. I feel that I have learned a great deal as a graduate student and I owe the majority of my newly acquired knowledge to Mike and Dave's mentorship. In other words, learning from them has been awesome in every way. In addition to our professional relationship, I am proud to call Mike and Dave my good friends. Our intellectual debates have been consistently filled with humor and laughter and I will always remember them as good times.

I have immencely enjoyed closely collaborating with Alessandro "morby" Morbidelli. In particular, a combination of his humility and his near-infinite analytical capabilities has made learning celestial mechanics from him completely effortless for me. I additionally want to thank him for making me feel at home during my stay at Observatoire de la Cote d'Azur, in Nice, France. A special place in my heart will always be reserved for my undergraduate advisor and close friend, Greg Laughlin. I remain continuously influenced by his ingenuity and approach to science and life and appreciate our joint scientific efforts.

Much of the research I have been involved in as a graduate student would not have been

possible without the help of my collaborators. I am thankful to Peter Bodenheimer for providing me with the necessary tools to model structural evolution of planetary interiors and teaching me about the numerical approach to the problem. Over the last two years, I've had a great time working with Sabine Stanley and learning from her deep insight into dynamo theory. Additionally, I want to thank her for making me feel so welcome during our collaboration at Harvard. I wish to thank Kleomenis Tsiganis for sharing his understanding of orbital dynamics with me as well as his help on chaos theory and modeling self-gravitating systems. I am thankful to John Johnson, with whom I have been collaborating on various exo-planetary problems over the last year. Finally, my understanding of the Kuiper belt as well as my ability to use graphics processing units for gravitational dynamics problems has been aided by my friend, Wesley Fraser.

Casual interactions with many people have greatly enhanced my life as a graduate student at Caltech. Numerous conversations with Andy Ingersoll, Peter Goldreich, Oded Aharonson, Heather Knutson and Geoff Blake have all enriched my academic life. My office mates and friends, Ajay Limaye, Mike Line, Alex Lockwood, Nick Heavens, Adam Waszczak, and Masha Kleschova, as well as Aaron Wolf, Alejandro Soto, Miki Nakajima, and Peter "Pistol Pete" Gao have all been instrumental to creating a supportive atmosphere at work. I'm very thankful to you all. A special thanks goes out to my friends/gym partners Alex Hayes and Jeff Thompson - good times.

I'm thankful to Mike Black and Scott Dungan for providing me with the oft-needed computer support. I am also grateful to the administrative support staff, Irma Black, Margaret Carlos, Nora Oshima and Norma Teran for making my life so easy. A special thanks

goes out to our custodian, John Brown for being so awesome.

Throughout graduate school, music has been a very important part of my life. As such, I would like to thank my band members/partners in crime, Scott Shepherd (drums), Ajay Limaye (guitar/bass), my dad Yuri (bass), and Mike Ferrara (drums) for rocking with me. Our band, The Seventh Season has been an amazing creative outlet for me and these past few years would not have been nearly as great without your company. I've enjoyed every gig we've played tremendously.

Finally, I wish to express my infinite gratitude for the love and support that my family has provided over the last few years. My wife and best friend, Olga has consistently been an amazing partner and confidant. I love her more than anyone will ever know and cherish every moment I have with her. I am thankful to our baby girl (at least that what we've been told), Emma, for enriching my life beyond any expectation - I'm so excited about meeting you in person. My parents, Yuri and Galina, have provided me with nothing but constant love, comfort and wise counsel my entire life and I'm eternally grateful to them. I further want to thank them for being fun-loving, permissive people, making my childhood entertaining and cultivating a spirit of natural curiosity which eventually transformed into a passion for science - I'm inspired by both of you. Additionally, I am thankful to my sister Ekaterina and her husband, Paul, Olga's brother Andrew as well as Olga's parents Andrey & Galina for their encouragement through the years. Last but not least, I would like to thank my cat, Orion for numerous useful conversations and being very supportive throughout my graduate career.

Abstract

The focus of this thesis is a collection of problems of timely interest in orbital dynamics and interior structure of planetary bodies.

The first three chapters are dedicated to understanding the interior structure of close-in, gaseous extrasolar planets (hot Jupiters). In order to resolve a long-standing problem of anomalously large hot Jupiter radii, we proposed a novel magnetohydrodynamic mechanism responsible for inflation. The mechanism relies on the electro-magnetic interactions between fast atmospheric flows and the planetary magnetic field in a thermally ionized atmosphere, to induce electrical currents that flow throughout the planet. The resulting Ohmic dissipation acts to maintain the interior entropies, and by extension the radii of hot Jupiters at an enhanced level. Using self-consistent calculations of thermal evolution of hot Jupiters under Ohmic dissipation, we demonstrated a clear tendency towards inflated radii for effective temperatures that give rise to significant ionization of K and Na in the atmosphere, a trend fully consistent with the observational data. Furthermore, we found that in absence of massive cores, low-mass hot Jupiters can over-flow their Roche-lobes and evaporate on Gyr time-scales, possibly leaving behind small rocky cores. In systems where a transiting hot Jupiter is perturbed by a long-period companion, apsidal precession of the hot Jupiter that results from its tidal bulge plays an important, and often dominant

role in determining the nature of the dynamical state onto which the system settles. This precession is in turn a strong function of the planet's degree of central concentration and is characterized by the planetary Love number. Utilizing this connection, we have shown that in tidally relaxed systems, measurement of the hot Jupiter's eccentricity directly yields the planetary Love number, which can then be used to place meaningful constraints on the physical structure of the planet with the aid of thermal evolution calculations.

Chapters four through six focus on the improvement and implications of a model for orbital evolution of the solar system, driven by dynamical instability (termed the "Nice" model). Hydrodynamical studies of the orbital evolution of planets embedded in protoplanetary disks suggest that giant planets have a tendency to assemble into multi-resonant configurations. Following this argument, we used analytical methods as well as self-consistent numerical N-body simulations to identify fully-resonant primordial states of the outer solar system, whose dynamical evolutions give rise to orbital architectures that resemble the current solar system. We found a total of only eight such initial conditions, providing independent constraints for the solar system's birth environment. Next, we addressed a significant drawback of the original Nice model, namely its inability to create the physically unique, cold classical population of the Kuiper Belt. Specifically, we showed that a locally-formed cold belt can survive the transient instability, and its relatively calm dynamical structure can be reproduced. We developed a simple analytical model for dynamical excitation in the cold classical region and showed that comparatively fast apsidal precession and nodal recession of Neptune, during its eccentric phase, are essential for preservation of an unexcited state. Subsequently, we confirmed our findings with self-consistent N-body

simulations, suggesting that the cold classical Kuiper belt's unique physical characteristics are a result of its remote formation site. Finally, we showed that the solar system may have initially hosted an additional ice-giant planet, that was ejected from the system during the transient phase of instability. Namely, we demonstrated that a large array of 5-planet (2 gas giants + 3 ice giants) multi-resonant initial states can lead to an adequate formation of the outer solar system, deeming the construction of a unique model of solar system's early dynamical evolution impossible.

The last four chapters of this thesis address various aspects and consequences of dynamical relaxation of planetary orbits through dissipative effects as well as the formation of planets in binary stellar systems. Using octopole-order secular perturbation theory, we demonstrated that in multi-planet systems, tidal dissipation often drives orbits onto dynamical "fixed points," characterized by apsidal alignment and lack of periodic variations in eccentricities. We applied this formalism towards investigating the possibility that the large orbital eccentricity of the transiting Neptune-mass planet Gliese 436b is maintained in the face of tidal dissipation by a second planet in the system and computed a locus of possible orbits for the putative perturber. Following up along similar lines, we used various permutations of secular theory to show that when applied specifically to close-in low-mass planetary systems, various terms in the perturbation equations become separable, and the true masses of the planets can be solved for algebraically. In practice, this means that precise knowledge of the system's orbital state can resolve the $\sin(i)$ degeneracy inherent to non-transiting planets. Subsequently, we investigated the onset of chaotic motion in dissipative planetary systems. We worked in the context of classical secular perturbation the-

ory, and showed that planetary systems approach chaos via the so-called period-doubling route. Furthermore, we demonstrated that chaotic strange attractors can exist in mildly damped systems, such as photo-evaporating nebulae that host multiple planets. Finally, we considered planetary formation in highly inclined binary systems, where orbital excitation due to the Kozai resonance apparently implies destructive collisions among planetesimals. Through a proper account of gravitational interactions within the protoplanetary disk, we showed that fast apsidal recession induced by disk self-gravity tends to erase the Kozai effect, and ensure that the disk's unwarped, rigid structure is maintained, resolving the difficulty in planet-formation. We also showed that the Kozai effect can continue to be wiped out as a result of apsidal precession, arising from planet-planet interactions in a mature planetary system. However, if such a system undergoes a dynamical instability, its architecture may change in such a way that the Kozai effect becomes operative, giving rise to the near-unity eccentricities, observed in some extrasolar planetary systems.

Contents

List of Figures

List of Tables

Chapter 1

Inflating Hot Jupiters with Ohmic Dissipation

Originally published as:
Batygin, K., & Stevenson, D. J. 2010, The Astrophysical Journal Letters, 714, L238

1.1 Abstract

We present a new, magnetohydrodynamic mechanism for inflation of close-in giant extrasolar planets. The idea behind the mechanism is that current, which is induced through interaction of atmospheric winds and the planetary magnetic field, results in significant Ohmic dissipation of energy in the interior. We develop an analytical model for computation of interior Ohmic dissipation, with a simplified treatment of the atmosphere. We apply our model to HD209458b, Tres-4b and HD189733b. With conservative assumptions for wind speed and field strength, our model predicts a generated power that appears to be large enough to maintain the transit radii, opening an unexplored avenue towards solving a decade-old puzzle of extrasolar gas giant radius anomalies.

1.2 Introduction

The detection of the first transiting extrasolar planet HD209458b (Charbonneau et al., 2000; Henry et al., 2000) marked the first observation of a planet whose radius is anomalously large. With the current aggregate of transiting planets exceeding 60, over-inflated "hot Jupiters" are now known to be common (Fig.1), and understanding their radii has become recognized as an outstanding problem in planetary astrophysics (Baraffe et al., 2010). Most proposed explanations require an interior power source that would replace the radiated heat from gravitational contraction and cause a planet to reach thermal equilibrium with a larger-than-expected radius. In the context of such solutions, the generated heat must be deposited into the interior envelope, i.e. below the radiative/convective boundary, in order to maintain the core entropy (and therefore the radius) of the planet. Notably, eccentricity tides (Bodenheimer et al., 2001), obliquity tides of a Cassini state (Winn & Holman, 2005), and deposition of kinetic energy to adiabatic depths by dynamical and convective instabilities (Guillot & Showman, 2002) have been invoked to provide an extra power source in the interior of the planet. It has been shown that the required powers are rather modest (Burrows et al., 2007), but it is unlikely that any of the proposed solutions alone are able to account for all observed radii (Baraffe et al., 2010; Fortney & Nettelmann, 2009).

Here we show that the anomalous sizes of close-in exo-planets can be explained by a magnetohydrodynamic mechanism. The interactions of zonal winds with the expected planetary magnetic field in a thermally ionized atmosphere induce an emf that drives electrical currents into the interior. These currents dissipate Ohmically and thus maintain the interior entropy of the planet. The primary controlling factors in our model are the atmo-

spheric temperature, wind velocity and strength of the magnetic field, as they dictate how much current is allowed to penetrate the interior. Other variables, such as metalicity also contribute, but to a smaller degree. Our results predict that interior heating of this kind occurs in all close-in exoplanets with magnetic fields, but it is negligible if the atmospheric temperature is not high enough for sufficient thermal ionization to take place. Smaller, but hot exoplanets are attributed to heavy element enrichment in the interior. While the inflation mechanism we present here is general, the quantitative modeling in this work is specific to HD209458b, Tres-4b, and HD189733b which are arguably the better studied transiting exoplanets.

1.3 Structural Model & Electrical Conductivity

Unlike Jupiter and Saturn, close-in extra-solar gas giants are exposed to high irradiation due to their proximity to parent stars. This forces their atmospheric temperature-pressure profiles to be significantly shallower than their solar system counterparts (Fortney & Nettelmann, 2009). In particular, the lower atmospheres ($P \gtrsim 0.1$ Bars) of hot Jupiters are believed to be almost isothermal while the radiative/convective boundaries are thought to lie at $P \sim 100 - 1000$ Bars, depending on the planet (Showman et al., 2008).

The isothermal sections of extrasolar gas giant atmospheres often reach temperatures close to 2000 K (Spiegel et al., 2009) and in some cases, even higher (Borucki et al., 2009). These temperatures are not high enough to ionize H or He significantly, however, alkali metals such as Na and K will be partially ionized. As a result, electrical conductivity in the interior of a hot Jupiter is dominated by ionization of hydrogen, while in the outer region

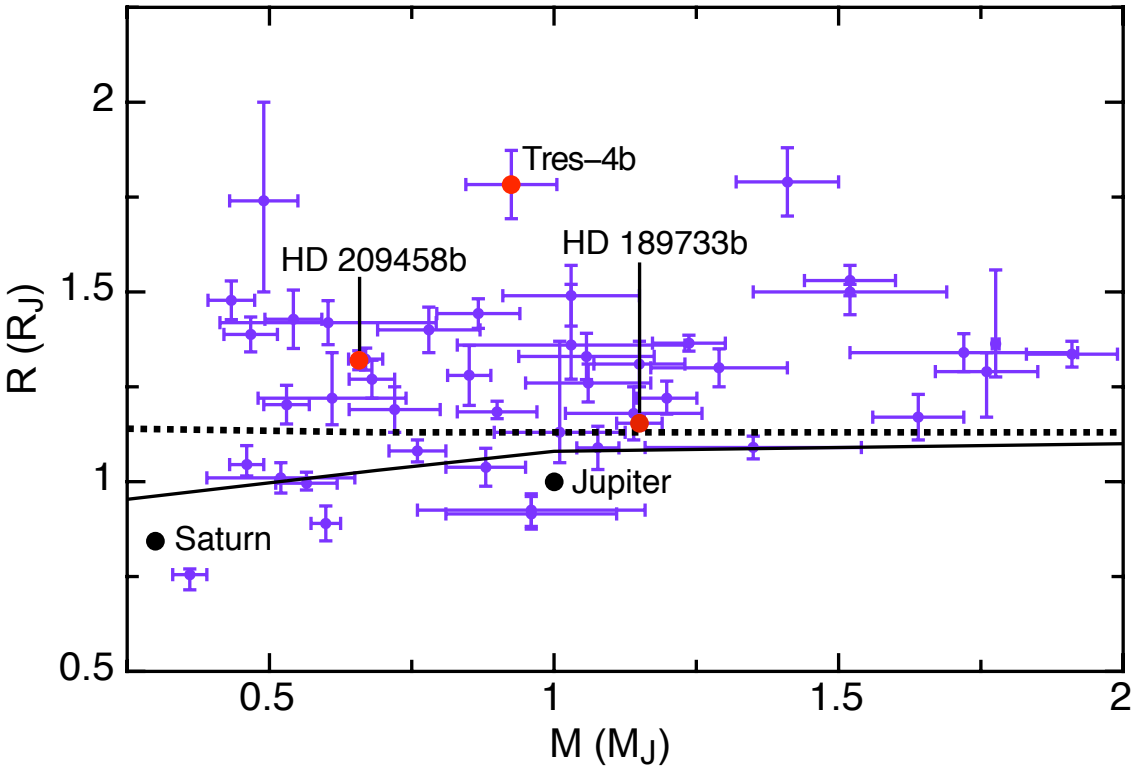


Figure 1.1: Scatter-plot of mass vs. radius of transiting Jovian exo-planets. The three planets considered in the text as well as Jupiter & Saturn are labeled. The two lines represent the theoretical mass-radius relationships for a core-less planet (dashed) and one with a $40M_{\oplus}$ core (solid) from Bodenheimer et al. (2003). Planets above the dashed line require an inflation mechanism to halt gravitational contraction.

of the planet, electrical conductivity is primarily due to the ionization of alkali metals, with the transition between the two inoization regimes taking place at $P \sim 300$ Bars.

Thermal ionization is governed by the Saha equation:

$$\frac{n_j^+ n_e}{n_j - n_j^+} = \left(\frac{m_e k_b T}{2\pi\hbar^2}\right)^{\frac{3}{2}} \exp\left(-I_j/k_b T\right), \tag{1.1}$$

where n_j and n_j^+ are the total and positively ionized number densities of constituent j respectively, $n_e = \sum n_j^+$ is the total electron number density, m_e is the electron mass, k_b is Boltzmann's constant, T is temperature, $\hbar$ is Plank's constant, and I_j is the ionization potential of constituent j. If the ionization is far from complete ($n_j^+ \ll f_j n$), the abundances of alkali metals, f_j are held constant, and the atmosphere is isothermal, it is easy to show that the electron number density takes on an exponential profile with an ionization scale-height that is twice as large as the density scale-heght:

$$n_e = n_0 \sqrt{\sum_{i=1}^{N} f_i \chi_i e^{\frac{r_0 - r}{2H}}}, \tag{1.2}$$

where χ's are the RHS's of equation (1), r_0 is the radial distance at some reference point ($P = 10$ Bars) and $H = k_b T/\mu g$ is the density scale-height. In our ionization calculations, we considered the following alkali metals: Na, K, Li, Rb, Fe, Cs and Ca. Their abundances and ionization potentials were inferred from Lodders (1999) and Cox et al. (2000) respectively.

The atmospheric temperatures above the isothermal layer differ significantly from planet to planet. In particular, thermal inversions have been detected in the atmospheres of HD209458b

(Burrows et al., 2007) and Tres-4b (Knutson et al., 2009) but not in HD189733b. In our models, we adopt atmospheric temperature profiles similar to that of Spiegel et al. (2009) for HD209458b and Tres-4b, and the 1D profile of Fortney et al. (2010) for HD189733b. The relatively cool temperatures attained above P $\lesssim 0.1$ Bars are of significant importance to our models because they provide insulating shells which are impenetrable to radial current. Consequently, current loops are necessarily setup through the interior, and any current flowing in the ionosphere is not relevant. We place the radiative/convective boundary at $P \sim 100$ Bars in all of our models.

We did not have to explicitly compute the ionization fractions of H and He, as they are published in the equation of state (Saumon et al., 1995), which we employed in our model. In particular, we used the "interpolated" version of the equation of state, where ionization occurs smoothly with pressure and temperature. Although the planetary structure was coreless, we mimicked the presence of a core by changing the Helium content from $Y = 0.24$ to $Y = 0.3$ (Burrows et al., 2003) in some of our models.

Having computed the electron number density, the electrical conductivity of a gas is given by (Tipler & Llewellyn, 2002)

$$\sigma = \frac{n_{e^-}}{n} \frac{e^2}{m_e A} \sqrt{\frac{\pi m_e}{8 k_b T}} \tag{1.3}$$

where n and A are the number density, and the number density weighted cross-section of everything other than electrons. Strictly speaking, the above equation is only valid for non-degenerate gas. However, by the point matter becomes degenerate in our models, the resistivity is completely negligible and the details of its profile have no noticeable effect on

the results.

Since we are only interested in the part of the planet, interior to the atmospheric temperature minimum, we define the model radius $r = R$ as the point of maximal conductivity in the atmosphere ($P = 75$ mbars), and we set the outer edge of our model at the conductivity minimum, $r = R + \gamma$ ($P = 30$ mbars). We place the bottom boundary of the "weather" layer of the atmosphere at a pressure of $P = 10$ Bars and denote it as $r = R - \delta$. Consequently, the "inert" layer of the atmosphere is between $100 \lesssim P \lesssim 10$ Bars. A computed electrical conductivity profile for HD209458b is presented in figure (2), along with a simplified conductivity profile resulting from equation (2). Because the functional profiles (dashed curve) are in good agreement with the numerically computed profile, we utilize them in all future calculations (see appendix).

1.4 Analytical Theory

Global circulation models (Showman et al., 2008, 2009; Langton & Laughlin, 2008; Menou & Rauscher, 2009) have shown that winds on hot Jupiters, specifically HD209458b and HD189733b, can attain velocities of order $v \sim$ 1km/s. It appears that two qualitative wind patterns are present. In the upper atmosphere ($P \lesssim 30$mbars), wind flows from the sub-stellar point to the anti-stellar point symmetrically across the terminator. Deeper down, a strong eastward zonal jet develops. Importantly, the development of zonal jets have been observed in virtually all simulations (see Showman et al. (2009) for a comprehensive review).

Prior to obtaining a formal solution to the problem, we can identify some of its features.

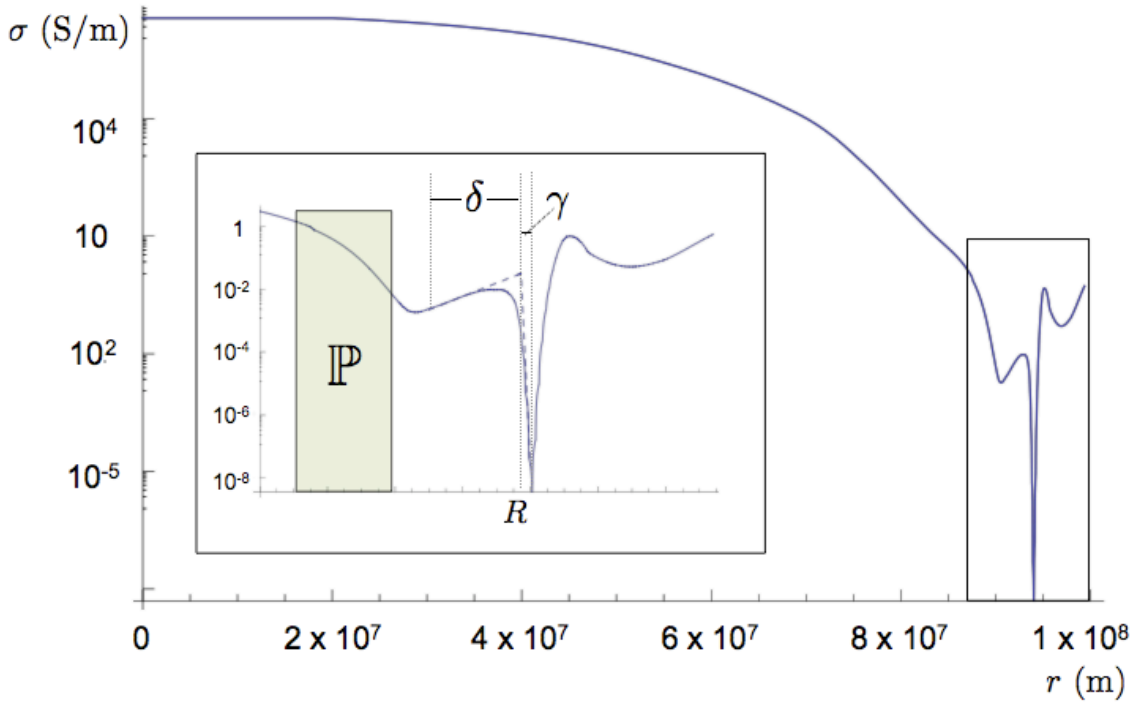


Figure 1.2: Electrical conductivity profile of the nominal HD209458b model with $T_{iso} = 1700$K, $Y = 0.24$ and $Z = 1\times$solar. The inset is a magnification of the profile in the outer part of planet. The model parameters R, δ and γ are labeled. The dashed lines are functional approximations to the conductivity profile where zonal flow is present. The highlighted region corresponds to the upper convective envelope $(100 - 3000)$Bars, where most of the interior dissipation takes place.

First, if the planet's dipole moment is aligned with the rotation axis and we consider only zonal flow, then there is azimuthal symmetry. Second, it is immediately apparent from the geometry of the zonal jet and the dipole field that the induced current will be meridional. In the atmosphere, we expect the current to flow from the poles to the equator where it penetrates the interior of the planet and completes the loop (Fig 3).

The general induction equation can be written as:

$$\frac{\partial \vec{B}}{\partial t} = -\vec{\nabla} \times \lambda \left(\vec{\nabla} \times \vec{B} \right) + \vec{\nabla} \times \left(\vec{v} \times \vec{B} \right), \tag{1.4}$$

where $\vec{B}$ is the magnetic field and $\lambda \equiv 1/\mu_0 \sigma$ is the magnetic diffusivity (Moffatt, 1978). We express the magnetic field as a dipole background component and an induced component: $\vec{B} = \vec{B}_{dip} + \vec{B}_{ind}$ with $\nabla \times \vec{B}_{dip} = 0$. This assumes no dynamo generation in the region. The induced magnetic field will tend to point in the same direction as the velocity field, so we can make the approximation $\vec{v} \times \vec{B} \approx \vec{v} \times \vec{B}_{dip}$. We assume that the prescribed velocity field and background magnetic field are not strongly modified by the induced field i.e. $Rm \equiv vL/\lambda \lesssim 1$, an assumption satisfied in our models with T$\leqslant$ 1700K. Finally, we seek a steady-state solution, so we require $\partial \vec{B}/\partial t = 0$. With these assumptions, the induction equation simplifies to:

$$\vec{\nabla} \times \lambda \left(\vec{\nabla} \times \vec{B}_{ind} \right) = \vec{\nabla} \times \left(\vec{v} \times \vec{B}_{dip} \right). \tag{1.5}$$

We can "uncurl" this equation and use Ampere's law $\vec{\nabla} \times \vec{B} = \mu_0 \vec{J}$ to recover Ohm's law:

$$\overrightarrow{J}_{ind} = \sigma \left(\overrightarrow{v} \times \overrightarrow{B}_{dip} - \vec{\nabla}\Phi \right). \tag{1.6}$$

By continuity, $\nabla \cdot \vec{J}$ must vanish. As a result,

$$\overrightarrow{\nabla} \cdot \sigma \overrightarrow{\nabla} \Phi = \overrightarrow{\nabla} \cdot \sigma \left(\overrightarrow{v} \times \overrightarrow{B}_{dip} \right). \tag{1.7}$$

If the conductivity takes on an exponential form, there exists an analytical solution for Φ and in our models, we confine the atmospheric flow to the region where conductivity is exponential. In the interior region, the electric potential is also governed by the above equation, with the right-hand side is set to zero. However, since the interior conductivity does not take on a simple analytical form, the above equation there must be solved numerically.

We take a nominal value for the "strength" of the field at the surface of the planets to be $||B||_R = 10^{-3}$T, approximately the value expected from scaling the field via the Elsasser number $\Lambda \equiv \sigma B^2 / 2\rho\Omega \sim 1$, where Ω is the planetary rotation rate (assumed tidally locked). The magnetic field scaling argument based on energy flux also suggests a similar value (Christensen et al., 2009). For comparison, Jupiter's surface field is $||B||_{R_{jup}} = 4.2 \times 10^{-4}$T (Stevenson, 2003). We approximate the zonal wind as $v \propto v_m \sin(\theta)\hat{\phi}$ where v_m is the maximum speed attained by the wind and set $v_m = 1$ km/s (see appendix for more details).

Once we have the solution for the current, we can compute the total Ohmic dissipation

rate below some radius r:

$$\mathbb{P} = \int\int\int \frac{\vec{J}^2}{\sigma(r)} dV. \tag{1.8}$$

In order to satisfy continuity, the magnitude of the current density must be constant along its path in the interior. As a result, it is apparent from the above equation that most of the dissipation takes place in the upper layers of the planet, where conductivity is not too great, and the solution is insensitive to the details of the conductivity profile in the deep interior, as long as it remains high. The Ohmic heat that is generated in the convective envelope of the planet replaces gravitational contraction, and is lost by radiative cooling at the radiative/convective boundary. Consequently, to ensure a null secular cooling rate, we need the Ohmic dissipation rate to at least compensate for the the radiative heat flux at the radiative/convective boundary (Clayton, 1968).

1.5 Model Results

It has been shown that extra-solar gas giants require between 10^{-6} and 10^{-2} of the irradiation they receive to be deposited into the adiabatic interior to maintain their radii (Bodenheimer et al., 2001; Burrows et al., 2007; Ibgui et al., 2009), although the exact number depends on the metallicity of the atmosphere and the mass of the heavy element core in the interior of the planet[1]. Under the assumption of solar metallicity and no core, HD209458b requires 3.9×10^{18}W, Tres-4b requires 8.06×10^{20}W and HD189733b requires no heating at all (Burrows et al., 2007; Ibgui et al., 2009). Within the context of our model, HD209458b

[1]If the dissipation is concentrated higher up in the atmosphere, 10-100$\times$ more heating is required (Guillot & Showman 2002).

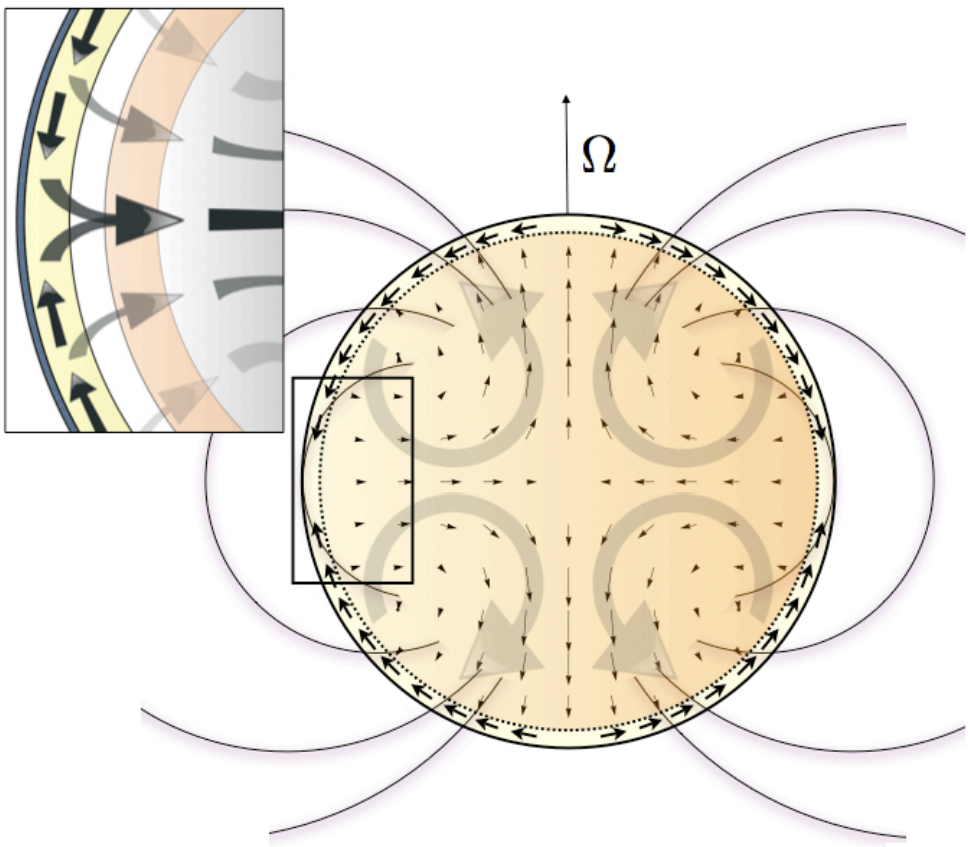


Figure 1.3: Side view cross-section of induced current due to zonal wind flow. The interior vector field, plotted with small arrows, is a quantitative result of the model. The large semi-transparent arrows are illustrations. The yellow shell in the inset represent the region to which we confine the zonal flow (10-0.03 Bars). The orange region denotes the region of interior heating.

and HD189733b are easily explained. To adequately explain Tres-4b however, we require an enhanced (10×solar) metallicity in the atmosphere to reduce the required heating down to 5.37×10^{19}W.

Table (1) presents a series of models with various temperatures, Helium contents, and metallicities of the planets under consideration. Upon inspection, it is apparent that the global heating rate scales exponentially with temperature, and as a square root of the metallicity. Both of these scalings can be easily understood by noting that scaling the conductivity profile by a multiplicative factor causes a corresponding change in dissipation while equations (1) and (2) relate temperature and metallicity (i.e. f) to the conductivity.

It is also noteworthy that the models with a simulated core produce approximately the same amount of heating as the coreless models. This is because most of the dissipation takes place in a region where ionization of alkali metals still dominates the electrical conductivity and the somewhat hotter interior isentrope makes little difference - recall that the value of the conductivity is unimportant in the deep interior.

There are a number of other scalings present within our model. For instance, the total Ohmic dissipation rate is proportional to the squares of the wind speed and the strength of the magnetic field, $\mathbb{P} \propto (B/10^3\text{T})^2(v_m/1\text{km s}^{-1})^2$. Additionally, to leading order, the dissipation in the atmosphere scales linearly with the thickness of the atmosphere, while the interior dissipation approximately scales quadratically. Consequently, along with the conductivity effect, hotter atmospheres also lead to more dissipation by virtue of a physically larger atmosphere.

It is important to understand that Ohmic heating does not only affect the interior. Be-

cause the induced current's "return path" lies in the atmosphere (Fig. 3), the atmosphere also gets heated. This heating, along with magnetic drag on the flow are the limiting factors of our theory.

Consider the nominal case of HD209458b with $T_{iso} = 1700$K and $Z = 1\times$solar. In this model, the heating required to inflate the planet is attained at a depth of ~ 90 Bars, essentially *at* the radiative/convective boundary. Note that if we shift the radiative/convective boundary to a higher pressure, the resulting change in the conductivity profile causes the heating to also shift to a somewhat greater depth. Ohmic heating in the atmosphere is small (only $\sim 4\%$) in comparison with the irradiation and $R_m \sim 0.3$. In other words, the assumptions implicit in our calculation are satisfied and the mechanism seems to explain the transit radius adequately. However, if we go to the model with $T_{iso} = 2000$K, $R_m \sim 3$, the Ohmic dissipation in the atmosphere is comparable with the insolation, and the assumptions of our model may no longer apply.

The nominal model of Tres-4b with $T_{iso} = 2250$K also runs into the same problem. Here, $R_m \sim 15$, and the Ohmic dissipation in the atmosphere is again comparable with the insolation. However, if we imagine that magnetic drag reduces the wind velocity by a factor of ~ 3, our results fall in the right ballpark to explain Tres-4b's radius, in the scenario where its atmospheric opacity is super-solar. Finally, consider the model of HD189733b. For this configuration of parameters, our mechanism does not predict a significant amount of Ohmic dissipation at adiabatic depths, consistent with an un-inflated radius. A similar scenario is observed for the model of HD209458b with $T_{iso} = 1400$K.

Overall, it appears that within the current setup of the model, the cumulative heating

below the weather layer of the atmosphere i.e. $r < R - \delta$ is of order a few$\times 10^{-2}$ of the heating that takes place in the atmosphere, which is in turn $\sim 10^{-2}$ of the insolation. Provided that this ratio of magnitudes holds up in a more dynamical treatment of the problem, it can provide an upper limit to the maximum inflation that can be explained with Ohmic dissipation. While in this paper we have only considered the effects of dissipation in the convective envelope, the intense heating present in the inert layers of the atmosphere may also play an important role in inflation (Bodenheimer, private communication). A quantitative description of this effect in the context of our model will be addressed in a future study.

1.6 Discussion

In this letter, we have presented a new, magnetohydrodynamic mechanism for inflation of extrasolar gas giants. Our calculations show that the heating, necessary to maintain the seemingly anomalous radii of transiting exo-planets, naturally emerges from considerations of interactions between partially ionized winds and the planetary magnetic field. Interestingly, there seems to be a set limit to the extent that Ohmic dissipation can heat the interior, making this theory testable, although it is noteworthy that a similar argument can be applied to other theories as well. Currently, there is significant uncertainty with respect to the calculation of the required interior heating, because core masses are unknown. However, dynamical determinations of interior structure (Batygin et al., 2009; Ragozzine & Wolf, 2009) may allow us to resolve the degeneracy for a fraction of observed planets, and provide a solid test-bed for the mechanism we've presented here.

There is a number of interesting additional questions that our model inevitably brings up. First, recall that our treatment of the induction equation is kinematic. In reality, flow modification by the Lorenz force may play an important role in determining the actual wind patterns. While this effect may be small for HD209458b and HD189733b, weather on hotter planets, such as Tres-4b or Wasp-12b may be more intimately linked with their magnetic fields, calling for a magentohydrodynamic treatment of the atmospheric circulation. Generally, when zonal winds interact with a background dipole field, they give rise to poloidal current which in turn, gives rise to a predominantly toroidal, unobservable field. However, the dayside-to-nightside flows that are present at higher levels in the atmosphere may modify the flow in an interesting way that may eventually be astronomically observable.

Second, we are neglecting the stellar magnetic field. The star's magnetic field is likely to be considerably smaller than the planetary field at the planetary orbital radius, but induction by stellar field as well as coupling of the stellar and planetary magnetic field lines is certainly plausible. This too, may produce an astronomically observable signature. Finally, we are neglecting the effects the induced current in the interior on the planetary dynamo. Considerations of this sort may influence the background magnetic field of the planet. All of these aspects call for a self-consistent treatment of the full problem. Such calculations would no-doubt provide further insight into the physical structure of extra-solar gas giants.

1.7 Appendix

We approximate the electrical conductivity profile in the atmosphere with exponential functions:

$$\sigma = \begin{cases} \sigma_\delta e^{\frac{r-(R-\delta)}{H_\delta}} & R-\delta < r \leqslant R \\ \sigma_\gamma e^{\frac{r-R}{H_\gamma}} & R < r \leqslant R+\gamma \end{cases} \tag{1.9}$$

where σ_δ and σ_γ are the conductivities at $r = R-\delta$ and $r = R$ respectively, while H_δ and H_γ are the conductivity scale-heights in the corresponding regions. We prescribe a parabolic radial dependence to the zonal flow over the thickness of the atmosphere, δ, and maintain the velocity constant over the outermost thin shell, γ:

$$\vec{v} = \begin{cases} 0 & 0 < r \leqslant R-\delta \\ \beta v_m \sin\theta \hat{\phi} & R-\delta < r \leqslant R+\gamma \end{cases} \tag{1.10}$$

where

$$\beta = \begin{cases} \left(\frac{r-(\tilde{R}-\delta)}{\delta}\right)^2 & R-\delta < r \leqslant R \\ 1 & R < r \leqslant R+\gamma \end{cases} \tag{1.11}$$

Assuming alignment of the dipole moment and the rotation axis, the background dipole magnetic field can be expressed as follows:

$$\vec{B}_{dip} = \vec{\nabla} \times k \left(\frac{\sin\theta}{r^2}\right) \hat{\phi}.09o \tag{1.12}$$

With these expressions, we can decompose the angular part of $\vec{v}\times\vec{B}$ into spherical harmonics. Upon inspection, one finds that the only harmonic of interest has $\ell = 2$ and $m = 0$. Consequently, we write the potential as $\Phi = g(r)Y_2^0(\theta,\phi)$ and equation (7) becomes a scalar equation.

Because the outer edge of our models is set at an insulating shell, we require the radial current at $r = R+\gamma$ to be zero:

$$g'_\gamma(R+\gamma) = \sqrt{\frac{\pi}{5}}\frac{4kv_m}{3(R+\gamma)^3}. \tag{1.13}$$

This boundary condition is appropriate when the electrical resistance that the current will encounter radially, greatly exceeds that of a path confined to a surface i.e. $\int_R^{R+2\gamma}\sigma^{-1}dr \gg R\int_0^{\frac{\pi}{2}}\sigma^{-1}d\theta$. This criterion is satisfied in our models.

With this boundary condition, the radial part of the solution to equation (7) in the outermost shell ($R < r \leqslant R+\gamma$) reads:

$$\begin{aligned}
g_\gamma(r) &= \frac{e^{-\frac{R+r+\gamma}{H_\gamma}}}{90H_\gamma r^3\left(1-4H_\gamma+6H_\gamma^2\right)\left(12H_\gamma^2-6H_\gamma(R+\gamma)+(R+\gamma)^2\right)} \\
&\times (12\sqrt{5\pi}kv_m e^{\frac{R+r+\gamma}{H_\gamma}}H_\gamma\left(1-4H_\gamma+6H_\gamma^2\right)(6H_\gamma^2(2R-5r+2\gamma) \\
&- 2H_\gamma(R^2-3Rr-3r^2+2R\gamma-3r\gamma+\gamma^2)-r(r^2+(R+\gamma)^2)) \\
&- 30e^{\frac{R+\gamma}{H_\gamma}}H_\gamma^2(4H_\gamma+r)(12H_\gamma^2-6H_\gamma(R+\gamma)+(R+\gamma)^2)A_1 \\
&+ 5e^{\frac{r}{H_\gamma}}\left(24H_\gamma^3-18H_\gamma^2r+6H_\gamma r^2-r^3\right)(12H_\gamma^2+6H_\gamma(R+\gamma) \\
&+ (R+\gamma)^2)A_1),
\end{aligned} \tag{1.14}$$

where A_1 is an undetermined constant of integration. In a similar fashion, we can write

down the solution to the radial part of equation (7) in the region ($R - \delta < r \leqslant R$):

$$\begin{aligned} g_\delta(r) &= \frac{1}{15r^3}\Big(\frac{5A_2\left(-24H_\delta^3 + 18H_\delta^2 r - 6H_\delta r^2 + r^3\right)}{6H_\delta^2 - 4H_\delta + 1} - \frac{5A_3 H_\delta e^{-\frac{r}{H_\delta}}(4H_\delta + r)}{6H_\delta^2 - 4H_\delta + 1} \\ &- \frac{2\sqrt{5\pi}kv_m e^{-\frac{r}{H_\delta}}}{\delta^2}\Big(12H_\delta^2(4H_\delta + r)\mathrm{Ei}\left(\frac{r}{H_\delta}\right) + e^{\frac{r}{H_\delta}} \\ &\times \left(-192H_\delta^3 + 96H_\delta^2 r + 2(-24H_\delta^3 + 18H_\delta^2 r - 6H_\delta r^2 + r^3)\log(r)\right. \\ &- \left. 2H_\delta(12r^2 + R^2 - \delta^2) + r(4rR - R^2 + \delta^2)\right)\Big), \end{aligned} \tag{1.15}$$

where A_2 and A_3 are again undetermined constants, and Ei is an exponential integral: $\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t} dt$. Although equation (7) must be solved numerically in the interior, towards the center of the planet, where conductivity can be taken to be constant, it reduces to Laplace's equation. As a result we can use the polynomial eigenfunction $g_{int}(r) = A_4 r^2$ in the vicinity of the origin, and A_4 is the last undetermined constant. The four constants of integration are determined by continuity.

Acknowledgments We thank G. Laughlin, P. Bodenheimer, D. Charbonneau, A. Showman, J. Liu, H. Knutson, A. Wolf and M. Line for useful discussions and the anonymous referee for insightful comments.

Table 1.1: Ohmic dissipation attained at different pressures in various models of HD209458b, Tres-4b, and HD189733b

Planet	Y	T_{iso} (K)	Z ($\times$solar)	$\mathbb{P}$ [$P < 10$Bars] (W)	$\mathbb{P}$ [$P > 10$Bars] (W)	$\mathbb{P}$ [$P > 100$Bars] (W)
HD209458b	0.24	1400	1	2.30×10^{19}	2.23×10^{17}	1.09×10^{16}
HD209458b	0.24	1400	10	7.28×10^{19}	7.06×10^{17}	3.43×10^{16}
HD209458b	0.24	1700	1	1.14×10^{21}	1.01×10^{19}	5.60×10^{17}
HD209458b	0.24	1700	10	3.61×10^{21}	3.19×10^{19}	1.77×10^{18}
HD209458b	0.24	2000	1	1.22×10^{22}	3.24×10^{20}	7.09×10^{19}
HD209458b	0.24	2000	10	3.89×10^{22}	1.05×10^{21}	2.29×10^{20}
HD209458b	0.3	1400	1	2.22×10^{19}	1.30×10^{17}	9.18×10^{14}
HD209458b	0.3	1400	10	7.01×10^{19}	4.10×10^{17}	2.89×10^{15}
HD209458b	0.3	1700	1	6.97×10^{20}	7.67×10^{18}	8.02×10^{17}
HD209458b	0.3	1700	10	2.21×10^{21}	2.43×10^{19}	1.90×10^{18}
HD209458b	0.3	2000	1	1.38×10^{22}	3.13×10^{20}	4.05×10^{19}
HD209458b	0.3	2000	10	4.52×10^{22}	1.05×10^{21}	9.42×10^{19}
Tres-4b	0.24	2000	1	6.87×10^{22}	2.57×10^{21}	1.42×10^{20}
Tres-4b	0.24	2250	1	1.44×10^{23}	3.33×10^{21}	3.68×10^{20}
Tres-4b	0.24	2500	1	4.62×10^{23}	7.87×10^{21}	1.54×10^{21}
Tres-4b	0.3	2000	1	4.80×10^{22}	9.56×10^{20}	3.16×10^{19}
Tres-4b	0.3	2250	1	1.98×10^{23}	5.92×10^{21}	6.16×10^{20}
Tres-4b	0.3	2500	1	5.13×10^{23}	8.75×10^{21}	1.55×10^{21}
HD189733b	0.3	1500	1	9.94×10^{18}	2.65×10^{16}	1.00×10^{16}

Bibliography

Batygin, K., Bodenheimer, P., & Laughlin, G. 2009, The Astrophysical Journal Letters, 704, L49

Baraffe, I., Chabrier, G., & Barman, T. 2010, Reports on Progress in Physics, 73, 016901

Bodenheimer, P., Lin, D. N. C., & Mardling, R. A. 2001, The Astrophysical Journal, 548, 466

Bodenheimer, P., Laughlin, G., & Lin, D. N. C. 2003, The Astrophysical Journal, 592, 555

Bodenheimer, P., 2010, personal communication

Borucki, W. J., et al. 2009, Science, 325, 709

Burrows, A., Hubeny, I., Budaj, J., & Hubbard, W. B. 2007, The Astrophysical Journal, 661, 502

Burrows, A., Hubeny, I., Budaj, J., Knutson, H. A., & Charbonneau, D. 2007, The Astrophysical Journal Letters, 668, L171

Burrows, A., Sudarsky, D., & Hubbard, W. B. 2003, The Astrophysical Journal, 594, 545

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, The Astrophysical Journal Letters, 529, L45

Christensen, U. R., Holzwarth, V., & Reiners, A. 2009, Nature, 457, 167

Cox, A. N., Becker, S. A., & Pesnell, W. D. 2000, Allen's Astrophysical Quantities, 499

Clayton, D. D. 1968, New York: McGraw-Hill, 1968

Fortney, J. J., & Nettelmann, N. 2009, Space Science Reviews, 108

Fortney, J. J., Shabram, M., Showman, A. P., Lian, Y., Freedman, R. S., Marley, M. S., & Lewis, N. K. 2010, The Astrophysical Journal, 709, 1396

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

Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, The Astrophysical Journal Letters, 529, L41

Ibgui, L., Spiegel, D. S., & Burrows, A. 2009, arXiv:0910.5928

Knutson, H. A., Charbonneau, D., Burrows, A., O'Donovan, F. T., & Mandushev, G. 2009, The Astrophysical Journal, 691, 866

Langton, J., & Laughlin, G. 2008, The Astrophysical Journal, 674, 1106

Lodders, K. 1999, The Astrophysical Journal, 519, 793

Menou, K., & Rauscher, E. 2009, The Astrophysical Journal, 700, 887

Moffatt, H. K. 1978, Cambridge, England, Cambridge University Press, 1978. 353 p.

Ragozzine, D., & Wolf, A. S. 2009, The Astrophysical Journal, 698, 1778

Saumon, D., Chabrier, G., & van Horn, H. M. 1995, The Astrophysical Journal Supplement, 99, 713

Showman, A. P., Y-K. Cho, J., & Menou, K. 2009, arXiv:0911.3170

Showman, A. P., Cooper, C. S., Fortney, J. J., & Marley, M. S. 2008, The Astrophysical Journal, 682, 559

Showman, A. P., Fortney, J. J., Lian, Y., Marley, M. S., Freedman, R. S., Knutson, H. A., & Charbonneau, D. 2009, The Astrophysical Journal, 699, 564

Spiegel, D. S., Silverio, K., & Burrows, A. 2009, The Astrophysical Journal, 699, 1487

Stevenson, D. J. 2003, Earth and Planetary Science Letters, 208, 1

Tipler, P. A., & Llewellyn, R. 2002, Modern Physics, 4th edition, by Paul A. Tipler and Ralph Llewellyn. W. H. Freeman Publishers, ISBN 0-71674345-0, 700pp, 2002.

Winn, J. N., & Holman, M. J. 2005, The Astrophysical Journal Letters, 628, L159

Chapter 2

Evolution of Ohmically Heated Hot Jupiters

Originally published as:
Batygin, K., Stevenson, D. J., & Bodenheimer, P. H. 2011, The Astrophysical Journal, 738, 1

2.1 Abstract

We present calculations of thermal evolution of Hot Jupiters with various masses and effective temperatures under Ohmic dissipation. The resulting evolutionary sequences show a clear tendency towards inflated radii for effective temperatures that give rise to significant ionization of alkali metals in the atmosphere, compatible with the trend of the data. The degree of inflation shows that Ohmic dissipation, along with the likely variability in heavy element content can account for all of the currently detected radius anomalies. Furthermore, we find that in absence of a massive core, low-mass hot Jupiters can over-flow their Roche-lobes and evaporate on Gyr time-scales, possibly leaving behind small rocky cores.

2.2 Introduction

Over the last decade, novel discoveries of transiting extra-solar planets have often been fraught by unexplained radius anomalies. In particular, many gas giant planets, residing on orbits in extreme proximity to their host stars, have been found to have much larger radii than what was previously thought possible in the context of "standard" gas giant theory (Stevenson, 1982; Guillot, 1999). A number of possible explanations to this problem have been proposed, most notably tidal heating (Bodenheimer et al., 2001, 2003), kinetic heating due to breaking gravity waves (Guillot & Showman, 2002), enhanced opacity (Burrows et al., 2007), semi-convection (Chabrier & Baraffe, 2007) and turbulent burial of heat by a mechanical greenhouse effect (Youdin & Mitchell, 2010). However, it appears unlikely that any of the above solutions can simultaneously explain all of the observed anomalies (Fortney & Nettelmann, 2010). Recently, a new Ohmic heating mechanism, that relies on the electro-magnetic interactions between atmospheric flows and the planetary magnetic field, has been suggested as a promising explanation to the inflation problem (Batygin & Stevenson, 2010).

A characteristic that is unique to hot Jupiters is the enormous amount of incident energy that they receive from their host stars. The extreme irradiation naturally results in high atmospheric temperatures (sometimes in excess of 2000K) (Spiegel et al., 2009), and the latitudinal dependence of this heating leads to fast, super-rotating equatorial jets with characteristic wind-speeds of order ~ 1 km/s (Showman et al., 2009). Consequently, the atmospheres of hot Jupiters are hot enough to thermally ionize alkali metals, that are present in trace abundances, and give rise to a finite electrical conductivity of order $\sim 10^{-3} - 1$ S/m,

depending primarily on the effective temperature. Rapid advection of the ions by the jets, while immersed in the planetary magnetic field, leads to an induction of an electro-motive force that sets up electrical current loops through the deep interior and the atmosphere of the planet. These electrical currents give rise to Ohmic heating as they flow throughout the planet.

It has been shown, in the context of a static structural model, that the level of resulting dissipation in the interior is approximately that required to maintain the radii of hot Jupiters (Batygin & Stevenson, 2010). However, time-evolution of Ohmically heated gas-giants remains an open question, since the degree of dissipation exhibits strong dependence on the structure of the planet, and particularly on the planet's T_{eff}, mass, and the location of the radiative/convective boundary. More explicitly, the thermal structure can be affected by the heating and its radial variation as well as the elapsed time. At the same time, the conductivity arises from thermal ionization, causing the interior structure to feedback on the heating. Thus, the structural profile of the planet dictates the Ohmic heating distribution. As a result, self-consistent evolutionary calculations are required to assess whether Ohmic dissipation can indeed resolve the inflation problem. To perform such calculations is the primary purpose of this study.

Here, we construct an approximate model with specific assumptions, that couples calculations of Ohmic heating and structural evolution, and show how this heating together with the likely variability of heavy element abundances can provide an explanation for all of the currently observed radius anomalies. We compute the maximum radii attainable by evolved Hot Jupiters as functions of planetary mass and effective temperature and find a tendency

for larger radii for effective temperatures that correspond to the onset of significant conductivity arising from alkali metals in the atmosphere. This behavior is compatible with the trend of the data. Furthermore, our calculations suggest that in absence of a relatively massive, high-metallicity core, low-mass hot Jupiters can expand beyond their Roche-lobes and spill their envelopes onto their parent stars on billion year time-scales, possibly leaving behind small rocky cores.

The paper is structured as follows. In section 2, we consider the energetics of the Ohmic mechanism, and show that in steady state, the degree of dissipation is limited by the thermodynamic efficiency of the atmosphere. In section 3, we present a simple model for the damping of the global circulation by the Lorentz force, and show that characteristic efficiency of the Ohmic mechanism is of order a few percent. In section 4, we describe the coupled Ohmic heating/themal evolution model that we employ in our calculations. In section 5, we present the theoretical curves that delineate mass-radius-T_{eff} space and compare our results with observational data. We conclude and discuss avenues for future improvement of the Ohmic inflation model in section 6.

2.3 Work - Ohmic Dissipation Theorem

Let us begin by considering the energetics of the Ohmic dissipation mechanism. We start by writing out the invicid Navier-Stokes equation, where turbulent stresses have been neglected and the last term is the Lorentz force:

$$\frac{D\vec{v}}{Dt} = -2\vec{\Omega} \times \vec{v} - \frac{\vec{\nabla} P}{\rho} + \vec{g} + \frac{\vec{J} \times \vec{B}}{\rho}. \tag{2.1}$$

In the above equation, $\vec{v}$ is the velocity, $\vec{\Omega}$ is the rotational velocity, P is pressure, ρ is density, g is the gravitational acceleration, $\vec{J}$ is the current density and $\vec{B}$ is the background planetary magnetic field. There exists a vast literature on the subject of energy budgets of the various terms in the Navier-Stokes equation (see for example Holton (1992) and Peixoto & Oort (1992)). Although a discussion of the global picture is important for understanding the details of atmospheric dynamics, the aim of our study is limited to the Ohmic mechanism, so we shall focus our discussion on the Lorentz term. The power, or rate of change of kinetic energy energy of the fluid per unit volume, provided solely by the Lorentz force is

$$\left(\frac{\rho}{2}\frac{Dv^2}{Dt}\right)_L = \vec{v}\cdot\vec{J}\times\vec{B} \tag{2.2}$$

If the quantity on the right is positive, the Lorentz force adds kinetic energy to the flow, while if it is negative, kinetic energy is drained from the flow. Recall that MHD Ohm's law reads

$$\vec{J} = \sigma(\vec{v}\times\vec{B} - \nabla\Phi) \tag{2.3}$$

where σ is the electrical conductivity and Φ is the electric scalar potential. Using a standard vector-calculus identity, we can write $\vec{v}\cdot\vec{J}\times\vec{B} = -\vec{J}\cdot\vec{v}\times\vec{B}$. Consequently, the energy equation can be re-written as

$$\vec{v}\cdot\vec{J}\times\vec{B} = -\frac{J^2}{\sigma} - \vec{J}\cdot\nabla\Phi \tag{2.4}$$

The first term on the RHS, is the Ohmic dissipation. Upon integration, by Gauss's theorem, the last term vanishes, since we require no radial current at the outer boundary:

$$\int\int\int \vec{J}\cdot\nabla\Phi dV = \int\int\int \nabla\cdot(\vec{J}\Phi)dV = \oint(\vec{J}\Phi)\cdot d\vec{a} = 0 \tag{2.5}$$

As a result, we discover that in steady-state, Ohmic dissipation is work done by the flow:

$$\int\int\int \left(\frac{\rho}{2}\frac{Dv^2}{Dt}\right) dV = -\int\int\int \frac{J^2}{\sigma} dV \tag{2.6}$$

It is important to note, however, that the work done by the flow is limited by the efficiency factor i.e. the fraction of insolation that is available to do useful work. In practice, this means that the total Ohmic dissipation rate should be rather insensitive to the magnetic field strength, once the field is larger than some critical "saturation" value. Particularly, in the saturated case, the flow velocity should scale inversely with the magnetic field. Thus, in such a regime, changing the field will not change dissipation or the planet radius. Rather the efficiency factor plays the governing role. Although numerical simulations are required to quantitatively understand the saturation field strength accurately, based on dimensionless number analysis, it is likely that the critical field strength is not overwhelmingly high (~ 1 Gauss or so).

2.4 Magnetic Damping of the Global Circulation and the Efficiency of the Ohmic Mechanism

In the limit where the Lorentz force is dynamically negligible, the wind speeds may be interpreted from global circulation models (GCMs). In this case, the computation of Ohmic dissipation is straight-forward, although it is noteworthy that GCM's carry an intrinsic range of numerical uncertainty (Cho et al., 2008; Showman et al., 2010; Heng et al., 2011). For sufficiently high atmospheric ionization levels and/or planets with strong magnetic fields, however, the onset of flux-freezing can act to slow down the winds, effectively capping the degree of Ohmic dissipation. The total Ohmic dissipation can be written as some fraction of the absorbed sunlight; we refer to this as the efficiency, ϵ. This efficiency must obviously go to zero as one goes to very low conductivities or low background field strength and is limited from above by thermodynamic (Carnot efficiency) considerations. It is further noteworthy that only a part of the total Ohmic dissipation ($\sim 5\%$ for $T_{eff} \gtrsim 1400$K; $\sim 10^{-4}\%$ for $T_{eff} = 1000$K) is deposited in the region of relevance to our inflation mechanism (the convective interior); the rest is dissipated at higher levels[1] (see Figure 4 and the associated discussion). The primary aim of this section is to derive a quantitative estimate for the Ohmic efficiency.

The limitation to efficiency can arise in two ways: through the reduction of the wind speeds as a direct consequence of the Lorentz force or through the reduction of the equator to pole temperature difference because of meridional circulation. We have not performed

[1]This is because for lower T_{eff} the atmosphere is characterized by higher resistance, and thus dissipates fractionally more energy.

full MHD global circulation models. Instead, following Schneider & Lindzen (1977) and Held & Hou (1980), we develop a simple, analytical model for the mid-latitude circulation in hot Jupiters, crudely accounting for magneto-hydrodynamic interactions.

Let us begin by simplifying the Lorentz force. Since the electric field is of the same order of magnitude as $\vec{v} \times B_{dip}$ as defined by the boundary condition, we write $\vec{J} \sim \sigma(\vec{v} \times B_{dip})$. Collecting the conductivity, magnetic field and fluid density into a damping timescale and retaining only the component that opposes the flow, we have

$$\frac{\vec{J} \times \vec{B}}{\rho} \sim -\frac{\sigma \vec{v} B^2}{\rho} \sim -\frac{\vec{v}}{\tau_L} \tag{2.7}$$

In other words, the effect of the Lorentz force is to introduce a term into the equation of motion that can be approximated as a Rayleigh drag with characteristic timescale

$$\tau_L = \rho/\sigma B^2 \sim 10^6 (\frac{\rho}{0.1\text{kg/m}^3})(\frac{0.1\text{S/m}}{\sigma})(\frac{10^{-3}\text{T}}{B})^2\text{sec.} \tag{2.8}$$

A damping time-scale of the same functional form was used in GCM simulations of Perna et al. (2010) to mimic the the Lorentz force, in order to obtain an estimate of Ohmic dissipation in the atmosphere. As will be revealed below, this term introduces both direct drag as well as meridional circulation into an otherwise geostrophic solution.

Next we assume steady state and azimuthal symmetry, so all time and zonal derivatives in the Navier-Stokes equation vanish. Finally, under an assumption of a constant Coriolis parameter (f-plane approximation), the horizontal Boussinesq equations of motion reduce

to an Ekman balance:

$$fv_y = \frac{v_x}{\tau_L} \tag{2.9}$$

$$fv_x = -\frac{1}{\rho_0}\frac{\partial P'}{\partial y} - \frac{v_y}{\tau_L} \tag{2.10}$$

where we have expressed all quantities in a local cartesian coordinate system i.e. $v_\phi \to v_x$, $v_\theta \to v_y$, and P' is the component of pressure that is not compensated by gravity. Under the same assumptions, the vertical momentum equation, neglecting Coriolis effects simplifies to the hydrostatic equation

$$\frac{1}{\rho_0}\frac{\partial P'}{\partial z} = \alpha g \Theta' \tag{2.11}$$

where $\alpha = 1/\langle\Theta\rangle$ is the coefficient of thermal expansion and Θ' is the potential temperature departure from the background state $\bar{\Theta}(z)$. Stable stratification is implicit in the problem and we write the heat equation in accord with Newtonian cooling:

$$v_z\left(\frac{d\bar{\Theta}}{dz}\right) = -\frac{\Theta' - \Theta^{rad}}{\tau_N} \tag{2.12}$$

where Θ^{rad} is the radiative deviation from the background state, and $\tau_N = c_p P/4g\sigma_{SB}T_{eff}^3$ is the Newtonian cooling timescale. The physical meaning of Θ^{rad} is that if no meridional flow is present to transport heat from the equator to the pole, the potential temperature profile takes the form $\Theta = \bar{\Theta} + \Theta' = \bar{\Theta} + \Theta^{rad}$.

To obtain a solution for the meridional circulation, let us introduce a stream-function $\Psi(y, z)$ such that $v_y = \partial\Psi/\partial z$ and $v_z = -\partial\Psi/\partial y$. Upon differentiating the hydrostatic equation with respect to z, differentiating the y-momentum equation with respect to y and

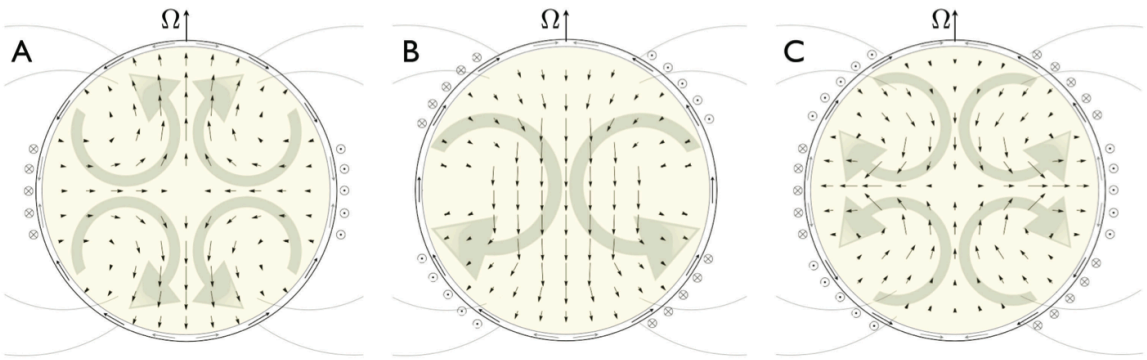


Figure 2.1: A comparison between interior current geometries, induced by (A) a single jet (B) two counter-rotating jets, each in one hemisphere and (C) a triple jet with retrograde equatorial flow. The small arrows inside the planet show the induced current (actual model output). Additionally, dipole magnetic field lines as well as graphical representations of the currents are presented. Atmospheric flows are represented as into-the-board and out-of-the-board arrows on the sides of the planets.

subtracting the first from the latter, we obtain a modified thermal wind equation.

$$f\frac{\partial v_x}{\partial z} = -\alpha g\frac{\partial \Theta'}{\partial y} - \frac{1}{\tau_L}\frac{\partial v_y}{\partial z} \tag{2.13}$$

Multiplying through by $d\bar{\Theta}/dz$, expressing the velocities in terms of the stream-function, and applying the heat equation, we arrive at a differential equation for the potential temperature perturbation:

$$\frac{\partial^2 \Theta'}{\partial z^2} + \frac{\tau_N}{\tau_L}\left(\frac{N^2}{f^2}\right)\frac{1}{1+(f\tau_L)^{-2}}\frac{\partial^2 \Theta'}{\partial y^2} = \frac{\partial^2 \Theta^{rad}}{\partial z^2} \tag{2.14}$$

where $N^2 = g\left\langle\Theta\right\rangle^{-1} d\Theta/dz$ is the square of the Brunt-Vaisala frequency. Note that the Lorentz force is not necessarily anti-parallel to the zonal velocity so the corespodence we use here is designed only to find the effect of the drag. Next we take the solutions to be of

the functional form:

$$\begin{aligned} \Psi &= \Psi_0 \sin(\pi y/R)\sin(\pi z/H) \\ v_x &= -v_{x0}\sin(\pi y/R)\cos(\pi z/H) \\ \Theta^{rad} &= \Theta_0^{rad}\cos(\pi y/R)\sin(\pi z/H) \\ \Theta' &= \Theta_0'\cos(\pi y/R)\sin(\pi z/H) \end{aligned} \tag{2.15}$$

where H is the atmospheric scale-height, R is the planetary radius, and all terms with the subscript "0" are undetermined constants. Upon substitution of these solutions into the above equation, we obtain an algebraic expression for the actual potential temperature deviation as a function of the radiative equilibrium deviation and other system parameters. This expression can be substituted back into the thermal wind equation to yield an order of magnitude equation for zonal wind velocity reduction due to the drag:

$$v_x \sim \frac{gH}{fR}\frac{\Delta\Theta^{rad}}{\langle\Theta\rangle}\left(1+(f\tau_L)^{-2}+\left(\frac{N}{f}\frac{H}{R}\right)^2\left(\frac{\tau_N}{\tau_L}\right)\right)^{-1} \tag{2.16}$$

where $\Delta\Theta^{rad}$ is the equator-to-pole potential temperature difference that the planet would have if no meridional circulation was present. Consider the limiting cases of this equation. If we take the drag to be negligible ($\tau_L \to \infty$), we have $\Theta' = \Theta^{rad}$, so the global potential temperature becomes $\Theta = \bar{\Theta} + \Theta^{rad}$, as $v_y \to 0$. Thus the wind velocity equation reduces to the standard thermal wind equation. If we consider the drag to be the dominant effect ($\tau_L \to 0$), then the potential temperature perturbation again becomes equivalent to the radiative potential temperature perturbation, since all circulation effectively stops. This is a

regime where the above treatment of the Lorentz force is highly inaccurate. The transition between the two regimes takes place when $f\tau_L \sim 1$.

The functional form for the stream-function that we introduced above satisfies the no-normal flow boundary conditions at the edges of the modeled "cell". The circulation it implies, however, is somewhat unrealistic and is not globally observed on either terrestrial or solar system gas giant atmospheres. This is largely because in order to obtain the analytical solution presented above, we made a series of simplifying approximations and implicitly omitted any discussion of turbulent Reynolds stresses, $\nabla \cdot (v'_x v'_y)$. In a detailed model for atmospheric circulation, the resulting eddy momentum fluxes are essential to closing the angular momentum cycle on gas giant planets (Liu & Schneider, 2010), although formally the above solution also closes the angular momentum cycle (Schneider & Lindzen, 1977). Consequently, we stress that the above picture is not intended to be a quantitatively good representation of the atmospheric dynamics on hot Jupiters. However, we do suggest that the model yields the correct behavior of the wind velocity reduction, to an order of magnitude.

Direct application of the above formalism to our models is not possible because of the large variation in conductivity and density. However, we can still use the resulting quantitative estimate of Ohmic efficiency as a guide, since any reasonable choices of parameters (e.g. the scale factors in equation 8), yield a Lorentz damping timescale, τ_L, that is far longer than the rotational timescale, implying a limited effect on the zonal winds. In other words, for most hot Jupiter regimes, it is reasonable to approximate the effect of the Lorentz force as a Rayleigh drag. The efficiency of conversion of zonal wind energy into heating is

then given by the ratio of the kinetic energy dissipated by the drag term in the equations of motion to the incident stellar flux:

$$\begin{aligned} \epsilon &\sim \rho v_\phi^2 H/(\tau_L \sigma_{SB} T_{eff}^4) \\ &\sim 0.01(\frac{\rho}{0.1\text{kg/m}^3})(\frac{v_\phi}{1\text{km/s}})^2(\frac{H}{1000\text{km}})(\frac{1500\text{K}}{T_{eff}})^4 \end{aligned} \tag{2.17}$$

A value of this order is in accord with the results of simplified GCM simulations of hot Jupiters where magnetic effects are modeled as drag (Perna et al., 2010). In order to qualitatively explore the effect of variable efficiency, we compute three sets of models, characterized by 1%, 3%, and 5% Ohmic efficiencies. Note that at low ($T_{eff} \sim 1000K$) temperatures, such efficiencies are likely to be overestimates, because of very low conductivity. However, the mechanism at very low conductivity is inactive anyway, so we retain the same efficiencies for consistency. From a computational point of view, a constant efficiency means that although the dissipation is a function of the radial coordinate, it is scaled such that the integrated Ohmic heating amounts to 1%, 3%, or 5% of the insolation. In other words, we only adjust the magnitude of the heating, not its radial distibution.

2.5 Coupled Ohmic Heating/Structural Evolution Model

In this study, the calculation of the Ohmic heating was performed in effectively the same way as in (Batygin & Stevenson, 2010), but with the atmospheric temperature structure computed rather than assumed. In other words, as the planet evolves along its evolutionary sequence, the conductivity profile everywhere in the planet is recalculated at every time-

step. Naturally, the heating profile is updated at every time-step as well. A number of studies in the past have concentrated all of the (tidal) heating in the deep interior of the planet (Bodenheimer et al., 2001, 2003; Ibgui et al., 2010). It is not known whether the tidal heating is deep-seated; it may be in the atmosphere and thus unavailable for inflation. By contrast the Ohmic model explicitly computes the relative amounts in the interior and atmosphere, so the actual calculated heating distribution is implemented in the structural calculations. Note that in our model, heating of the adiabatic interior as well as heating of the deep atmosphere (as in Guillot & Showman (2002)) contribute to the radial evolution.

We solve the simplified steady-state induction equation (see Batygin & Stevenson (2010) for derivation)

$$\nabla \cdot \sigma \nabla \Phi = \nabla \cdot \sigma \left(\vec{v} \times \vec{B}_{dip} \right) \tag{2.18}$$

All of these planets can be expected to have internal dynamos with field strengths similar to Jupiter ($|\vec{B}| \sim$ few Gauss) or perhaps larger because of the lower density in the dynamo generating region. It is implicitly assumed the magnetic field remains roughly dipolar at the planetary surface, in spite of the induced current. This assumption may be generally valid in most hot Jupiter atmospheres, as the magnetic Reynold's number, $Re_m = VL/\eta$, where η is the magnetic diffusivity, is typically of order unity at most, rendering dynamo action unlikely in the atmosphere. However, it is unclear if the induced current from the atmosphere will have an appreciable effect on the interior dynamo, and its role should be assessed with a detailed, numerical model. As in (Batygin & Stevenson, 2010), we used a simple analytical prescription for the velocity field and the flow is confined to the radiative atmosphere so that that the RHS of the above equation vanishes in the (assumed) rigidly

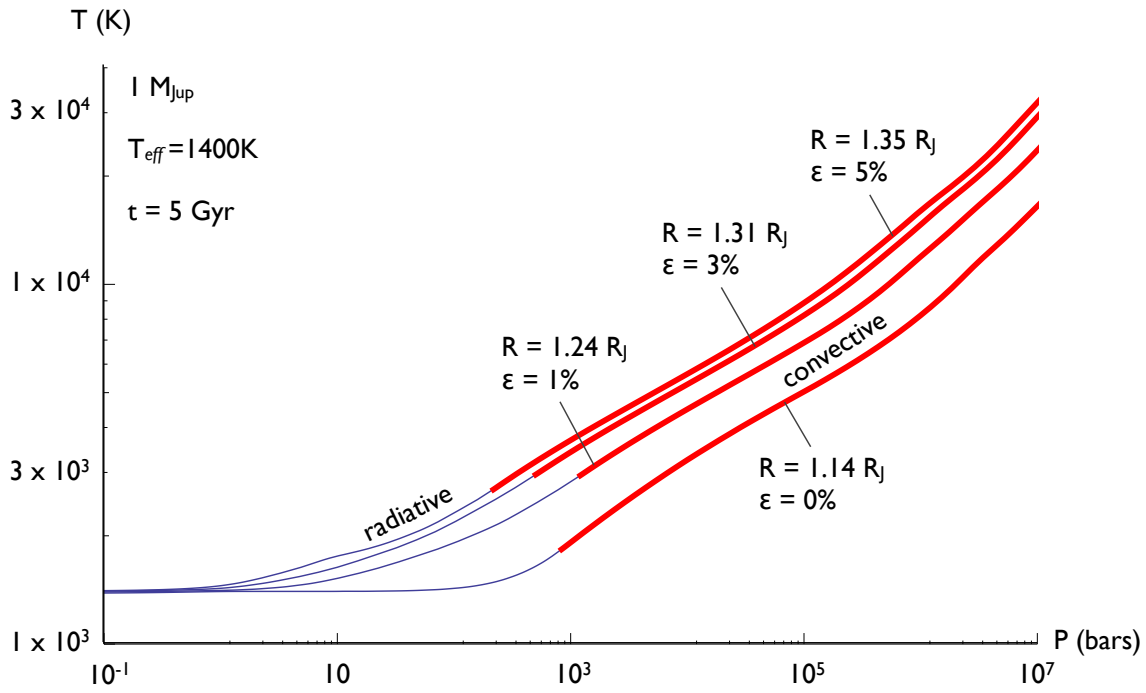


Figure 2.2: A series of representative pressure-temperature profiles of an evolved ($t = 4.5$Gyr) $1M_{Jup}$ planet, with $T_{eff} = 1400K$. The convective parts of the planet are plotted as thick red lines, while the radiative parts is plotted as thin purple lines. The four profiles correspond to $\epsilon = 0\%, 1\%, 3\%$ and 5% solutions.

rotating interior. Nearly solid body rotation of the deep interior is predicted by the same argument that was used to reach this conclusion for Jupiter (Liu et al., 2008). We assume that this is also the tidally synchronized state (Goldreich, 1963; Hut, 1981). Note that the assumption of tidal synchronization is present both, explicitly in our calculation of $\vec{v}\times\vec{B}_{dip}$, and implicitly in interpretation of GCMs of tidally locked planets. Once the electric scalar potential is known, the current, J, is determined from Ohm's law (including the induction emf i.e. equation 3). Thus, the Ohmic dissipation per unit mass, $\mathbb{P} = J^2/\rho\sigma$ everywhere in the planet can also be computed.

As already discussed above, we can interpret the qualitative characteristics of the velocity field from GCM results. Although the details of the flows that GCMs produce vary

somewhat, depending on the assumptions implicit to the solver (Heng et al., 2011), most solutions produce a broad single jet in the atmosphere. The physical mechanism for the jet's formation has been identified as standing planetary-scale Rossby and Kelvin waves (Showman and Polvani, 2011). However, there exists a notable exception in the literature, where retrograde jets are also produced (Thrastarson & Cho, 2010). This may be of importance for the Ohmic mechanism, since the atmospheric flows govern the induction of the interior current. To explore the effect of retrograde motions, we have computed the interior currents in a static model (as in Batygin & Stevenson (2010)) with three different prescriptions for the jets. Namely, we considered, a single broad jet ($v_\phi \propto \sin\theta$), two counter-rotating jets, one in each hemisphere ($v_\phi \propto \sin 2\theta$), as well as three jets, with retrograde equatorial flow ($v_\phi \propto \sin 3\theta$).

The results of these calculations are presented in side-view cross-sections of planets in Figure 1. In each of the three panels, the small arrows inside the planet show the induced current (actual model output). Additionally, the magnetic field lines as well as graphical representations of the currents and atmospheric flows are presented. Figure 1A shows the current induced by a single jet. As in the results of Batygin & Stevenson (2010), loops are set up such that the radial current, which is induced at the equator flows towards the poles in the interior, and turns around in the atmosphere. If the current is induced by a double jet (Figure 1B), its geometry changes considerably. Since the flows produce counter-acting electro-motive forces, the interior current now flows from one hemisphere to the other. However, the hemispheric symmetry is restored with three jets (Figure 1C), and due to a particular choice of signs (namely, prograde flow at the equator in Figure 1A and retrograde

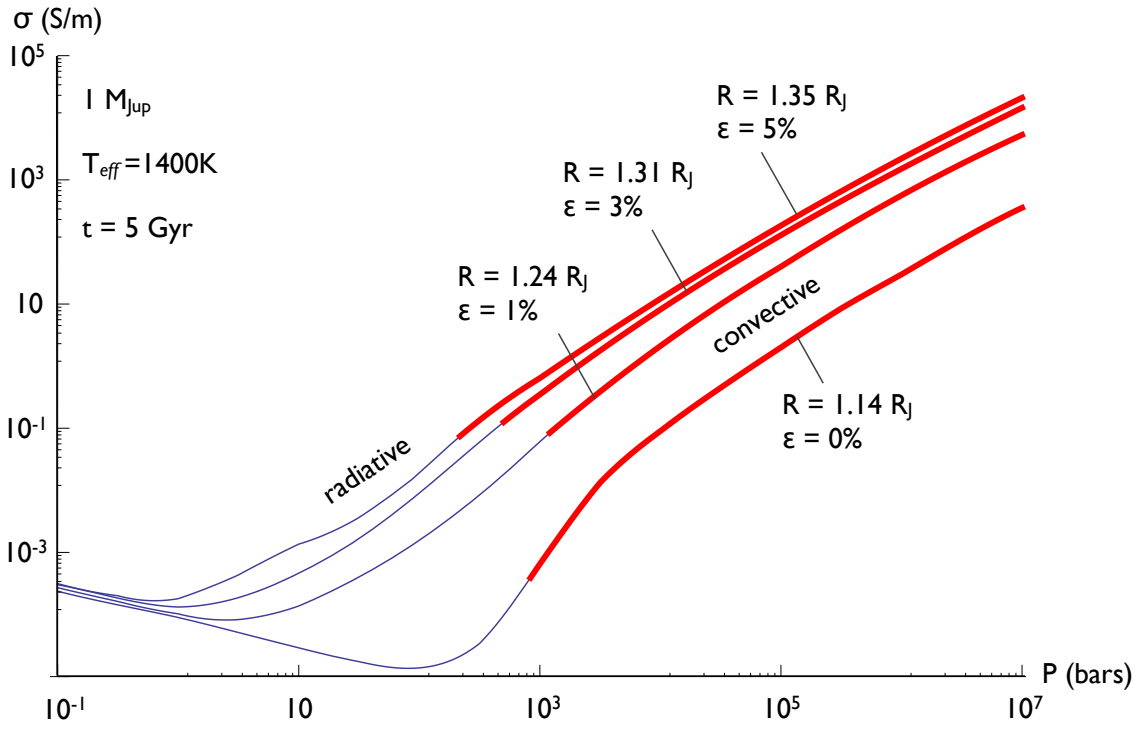


Figure 2.3: A series of representative electrical conductivity profiles of an evolved ($t = 4.5$Gyr) $1M_{Jup}$ planet, with $T_{eff} = 1400K$. The ionization was derived from the the temperature-pressure profiles, presented in Figure 2. The convective parts of the planet are plotted as thick red lines, while the radiative parts is plotted as thin purple lines. The four profiles correspond to $\epsilon = 0\%, 1\%, 3\%$ and 5% solutions.

flow at the equator in Figure 1C), the direction of the interior current is reversed.

Although, the interior current geometry changes considerably depending on the assumed flow, the resulting heating rate does not. Indeed, a comparison of the Ohmic dissipation profiles in the three static models described above reveals that the ratio of the cumulative dissipation below the radiative-convective boundary to the total is the same to within $\sim 10\%$. Consequently, in accord with most GCM results, we approximate the atmospheric flow as a single zonal jet in all following calculations.

Our calculations of planetary interior structure and evolution employed a descendant of the Berkeley stellar evolution code (Henyey et al., 1964). Over more than 40 years

of its lifetime, the model has been subject to a considerable amount of modification and improvement in its input physics and has been used extensively in the study of both extra-solar and solar system giant planets (Pollack et al., 1996; Bodenheimer et al., 2003; Batygin et al., 2009). A given evolutionary sequence starts at a radius of roughly $2R_{Jup}$ and ends at an age of 5 Gyr.

The program we used to compute the structural evolution of hot Jupiters assumes that the standard Lagrangian equations of stellar structure apply. Energy transport is accomplished either by convection or radiation, as dictated by the Schwarzschild criterion. Energy sources within the planet include gravitational contraction, cooling of the interior, and Ohmic dissipation. For the gaseous envelope, the interpolated Saumon et al. (1995) equation of state is used. In the models with a core, the solid/liquid core has constant density. The imperfections of the equation of state are expected to have a much smaller effect on the computed radii than the uncertainties in our model for Ohmic inflation.

The atmosphere is taken to be gray. Pure molecular opacities are used in the radiative outer layers of the planet (Freedman et al., 2008) while the high temperature and pressure, opacities of Alexander & Ferguson (1994) are used in the interior. In the atmosphere, it is assumed that dust grains contribute negligibly to the opacity. This may not be true but is again a smaller uncertainty than the other uncertainties in the modeling. We refer the reader to Hubickyj et al. (2005) and Dodson-Robinson et al. (2008) for further reading. Given the approximate treatment of magnetic induction in the model, it is sensible to utilize a simple gray atmosphere. However, it is noteworthy that a different atmospheric model could modify our results on a quantitative level, since the atmospheric boundary condition

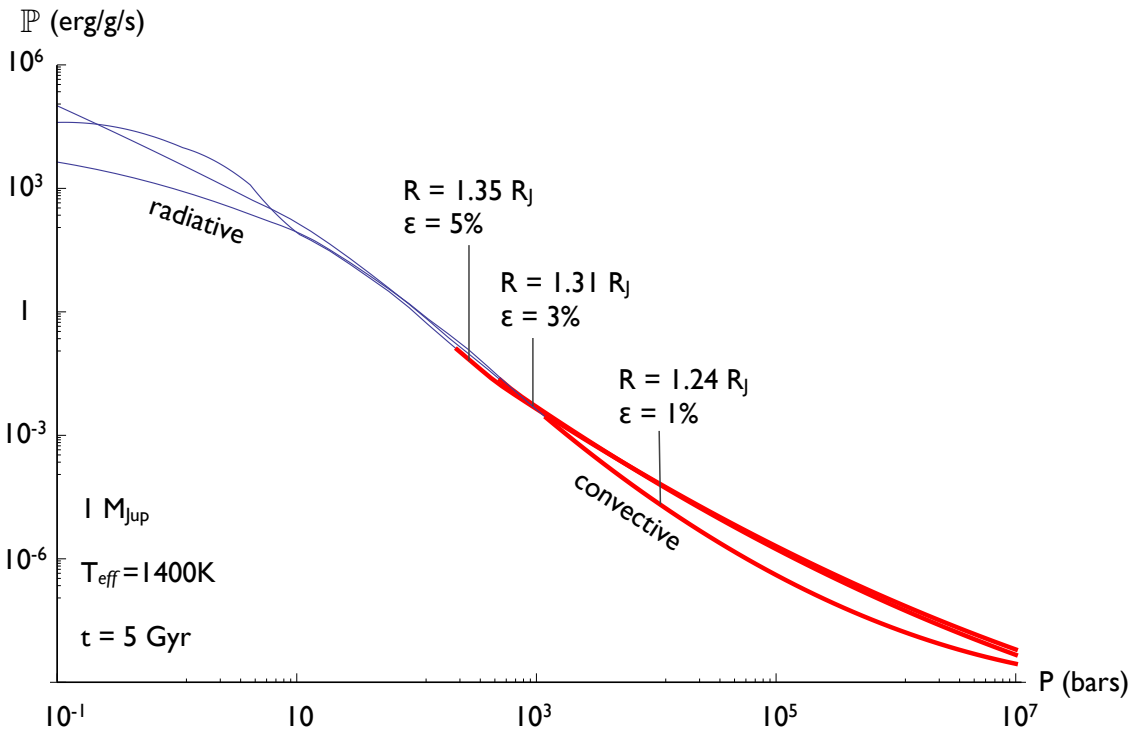


Figure 2.4: A series of representative Ohmic heating profiles (energy dissipation rate per unit mass) of an evolved ($t = 4.5$Gyr) $1M_{Jup}$ planet, with $T_{eff} = 1400K$, corresponding to the $\epsilon \neq 0$ temperature-pressure profiles shown in Figure 2. The thick red lines denote the heating of the adiabatic interior, while the thin purple lines represent the heating of the radiative atmospheres.

directly governs the cooling rate of the interior (Guillot & Havel, 2011).

A series of representative temperature-pressure profiles of an evolved $1M_{Jup}$ planet are plotted in Figure 2. The four presented evolutionary sequences were started from the same initial condition, but were evolved with different amount of Ohmic heating, namely $\epsilon = 0$, $\epsilon = 1\%$, $\epsilon = 3\%$ and $\epsilon = 5\%$, as labeled. It is noteworthy that in planets with higher T_{eff}, Ohmic dissipation in the atmosphere can lead to a formation of a small convective region, in the otherwise radiative atmosphere. Electrical conductivity profiles, corresponding to the presented temperature-pressure profiles are shown in Figure 3.

At the planetary surface, defined as the Rosseland mean photosphere, the luminosity

is composed of two parts: the internal luminosity generated by the planet, and the energy absorbed from the stellar radiation flux and re-radiated (insolation). In our simulations, the atmospheric circulation spans the region between 10 bars and the photosphere, in accord with GCM results (Showman et al., 2009). Additionally, in the heating calculation, a boundary condition of zero radial electrical current is employed at the surface, as a result of significant decrease in conductivity with height in the upper atmosphere (Batygin & Stevenson, 2010). This decrease in electrical conductivity is not explicitly present in our models, since the temperature at the photosphere is T_{eff} by definition. However, significant drops in temperatures shortly above the photosphere can be observed in more sophisticated atmospheric models (e.g. Spiegel et al. (2009)). Furthermore, if a given atmosphere exhibits a thermal inversion, the characteristic decrease in temperature at $P \sim 10^{-2}$ bars provides a thin, electrically insulating layer. In other words, the null radial current boundary condition at the photosphere is likely to hold true for most atmospheres, whether or not an inversion exists.

A series of representative Ohmic heating profiles (corresponding to the $\epsilon \neq 0$ temperature-pressure profiles presented in Figure 2) are shown in Figure 4. Note that although the heating is maximum at the very surface, the layers that contribute to the expansion lie much deeper (i.e. below ~ 10 bars at an early epoch and below $\sim 10^2 - 10^3$ bars for evolved planets). The location of maximal heating corresponds to the atmospheric height where the electrical current turns around. Thus the higher up, electrically insulating layers of the atmosphere discussed in the previous paragraph, would experience very little Ohmic heating.

2.6 Results: Radial Evolution

We computed the structural evolution of Ohmically heated hot Jupiters with the mass range spanning an order of magnitude between $0.23M_{Jup}$ and $3M_{Jup}$ and effective temperature range between $1000K$ and $1800K$. All but one of our models are core-less, and thus should be viewed as giving an upper bound on the radius that a planet of a given mass and T_{eff} may achieve. In our suite of simulations, we observed two distinct families of solutions that are characterized by either a monotonically decreasing or a monotonically increasing radius in time. A few representative evolution sequences are presented in Figure 5, with their masses and effective temperatures labeled. In the solutions where the radius is monotonically decreasing with time, the final answer (radius at $t = 5$Gyr) is largely independent of the initial condition: the solution always asymptotically approaches the same "equilibrium radius." If the radius monotonically increases with time, however, the quantity of importance is the total integrated heating rate, which is intrinsically a function of the planetary radius. Consequently, these "unstable" solutions are unavoidably dependent on the initial conditions.

Hot Jupiters can change their radii significantly primarily because of a change in the entropy of the deep interior, since the outermost several (radiative) scale heights constitute only a small fraction of the planet radius. In the idealized limit where the heat escaping from the deep interior is small compared to Ohmic heating, $\mathcal{P}$, the equation governing evolution for the deep interior reads

$$\frac{d}{dt}(E_{\text{grav}} + E_{\text{int}}) = \mathcal{P}. \tag{2.19}$$

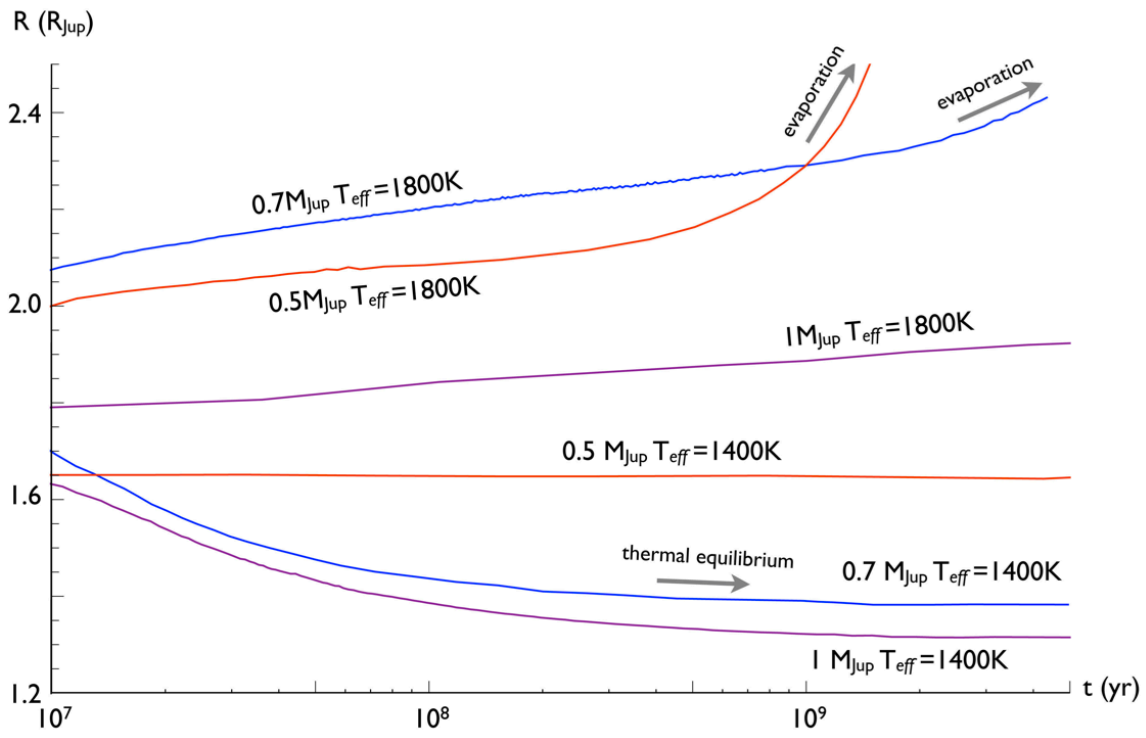


Figure 2.5: A sample of evolution sequences of $0.5, 0.7$ and $1 M_{Jup}$ coreless planets at $T_{eff} = 1400$ and $1800K$ with $\epsilon = 3\%$. The 0.7 and $1 M_{Jup}$ planets at $1400K$ planets asymptotically approach radii that are characterized by thermal equilibrium. The $0.5, 0.7$ and $1 M_{Jup}$ planets at $1800K$ are all on unstable evolutionary paths which eventually lead to Roche-Lobe overflow. However, the $1 M_{Jup}$ planet, however, is considerably more stable, owing to its more degenerate state. The evolution of the $0.5 M_{Jup}$ planet at $1400K$ lies between the two regimes - its starting radius is the equilibrium radius.

If the planet is a polytrope and the internal energy is entirely thermal then the sum of internal and gravitational energies can be written as $-kGM^2/R$ where k is a constant typically somewhat smaller than unity (Chandrasekhar, 1957). In the case of highly degenerate bodies (e.g. Jupiter itself) the zero temperature part of the internal energy changes in such a way as to exactly cancel the change in gravitational energy as radius changes (Hubbard, 1984) and the LHS above becomes only the time derivative of thermal energy to a good approximation. (This invalidates the claim, often made but erroneous, that the luminosity of degenerate planets is derived from contraction). The bodies of interest to us have non-ideal

thermodynamics and are not in the degenerate limit so no simple result holds. However, direct calculation shows that it is still approximately true to replace the LHS above with something like $d/dt(-kGM^2/R)$ with k of order unity, even though it is not correct to think of the energy as being derived solely from changes in gravitational energy. All of this discussion ignores possible effects arising from redistribution of the heavy elements.

The transition from stable to unstable solutions can only be understood through detailed models, because it depends on the details of opacity and conductivity structure, but the essential physics lies in the comparison of Ohmic heating at depth with the total radiative heat loss from the convective interior evaluated at the radiative/convective boundary. This heat loss scales as the adiabatic temparature gradient, which depends in turn on g, the gravitational acceleration. This is approxcimately linear in mass because radius does not vary much, so Ohmic heating overwhelms heat loss from the interior once the mass is sufficiently small. In the limit where the Ohmic heating dominates the evolution of the deep interior, we expect that the characteristic timescale of inflation $\tau_{infl} \equiv R/(dR/dt)$ is (to order of magnitude) the ratio of GM^2/R to Ohmic heating, or

$$\tau_{infl} \sim (\frac{0.01}{\epsilon})(\frac{M}{M_J})^2(\frac{R_J}{R})^3(\frac{1500\mathrm{K}}{T_{eff}})^4\mathrm{Gyr}. \tag{2.20}$$

As can be seen in Figure 5, after a sufficient amount of time, unstable solutions enter a phase of runaway growth, with the instability driven by the fact that the cumulative heating rate is proportional to the surface area of the planet. This inevitably leads to Roche-lobe overflow and evaporation of the planet (Laine et al., 2008). The timescale over which the planet can remain intact while on an unstable path is a function of the planetary mass

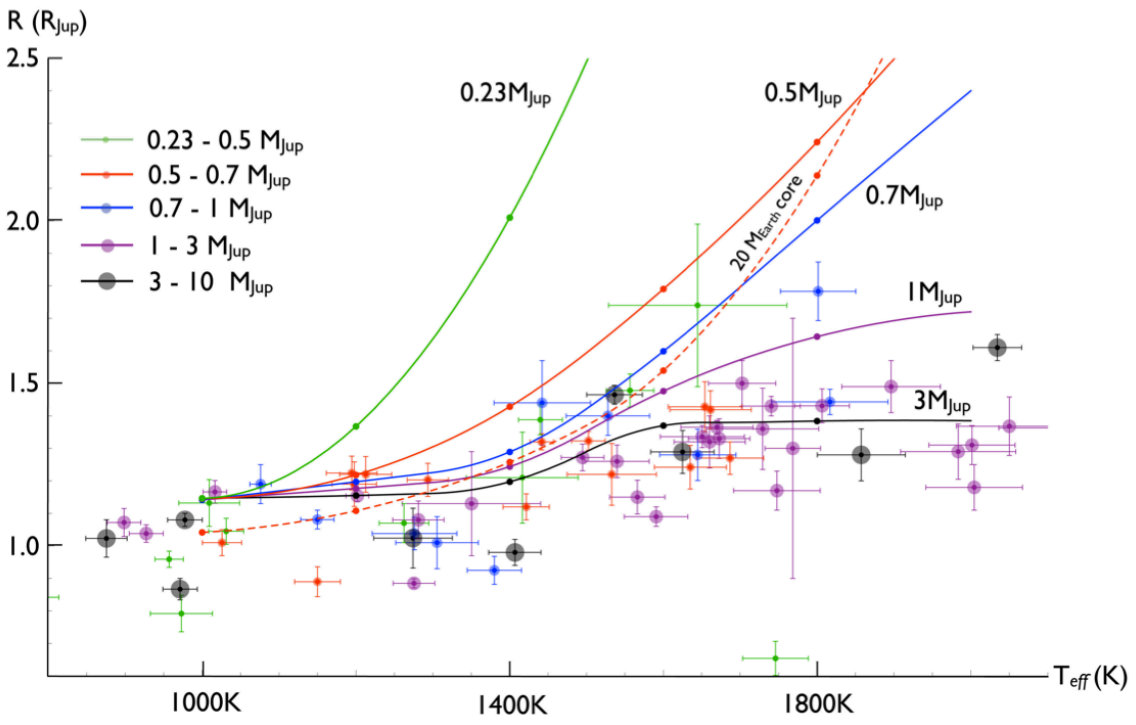


Figure 2.6: An $R - T_{eff}$ diagram of evolved models of various masses ($\epsilon = 1\%$) as well as the current aggregate of detected transiting hot Jupiters. The solid lines represent theoretical curves, corresponding to core-less models, while the dashed line depicts $M = 0.5 M_{Jup}$ models with a $20 M_{\oplus}$ solid core. The observational data points are presented, along with their corresponding 1σ error bars. The masses of the data are binned to fall between the theoretical curves and are labeled by color and size as shown on the figure.

and high-mass planets can, in principle, remain intact and grow slowly over many Gyr, while low-mass giant planets evaporate on a $\sim$ 1Gyr timescale. It is noteworthy that the timescale will also likely be affected by the $1/R$ slowdown of winds (see equation 16) and thus a possible modification of the efficiency (which we kept constant) with a growing radius. Still, it is very likely that evaporation is unavoidable in certain circumstances.

Low-mass planets can be stabilized against evaporation, if they possess sufficiently massive high-metallicity cores (Bodenheimer et al., 2003; Ibgui et al., 2010). However the critical core mass is far from being trivially small. In fact, a simulation of an $0.5 M_{Jup}$ planet with a ρ = 3g/cc, $20 M_{\oplus}$, core at $T_{eff} = 1800K$ reveals that it is only able to

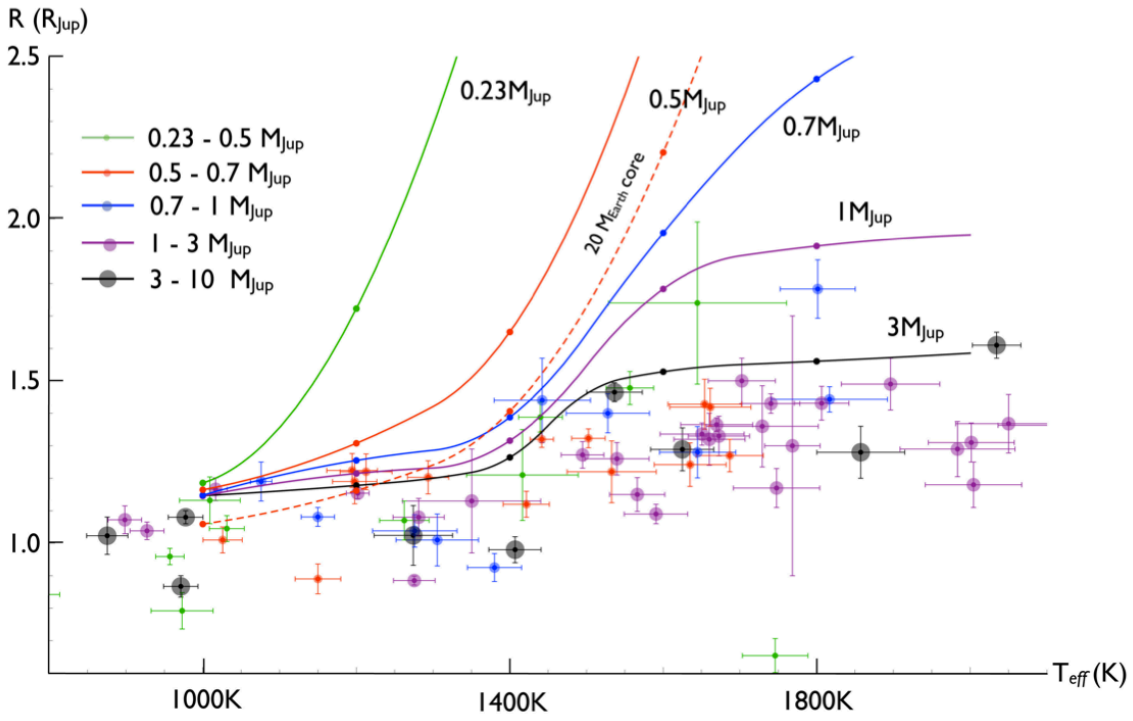


Figure 2.7: Same as Figure 2, but with $\epsilon = 3\%$

retain its envelope for $\sim$ 1 Gyr before overflowing the Roche-lobe. Given that hot low-mass planets tend to reside in compact multiple systems (Terquem & Papaloizou, 2007), while hot Jupiters usually lack close-by companions (Lo Curto et al., 2010), the discussion of evaporation is suggestive of the possibility that a significant fraction of hot, sub-giant planets, without companions, may have been born as giant planets and have since lost their gaseous envelopes[2].

Within the context of our mechanism, the planetary radius is most clearly visualized as function of effective temperature rather than mass. Consequently, we choose to plot equal-efficiency, equal-mass lines on an $R - T_{eff}$ diagram rather than a conventional $R - M$ diagram. Our results, along with the current aggregate of well-characterized transiting hot

[2]A similar idea has already been proposed in the context of thermal atmospheric escape (Baraffe et al., 2005).

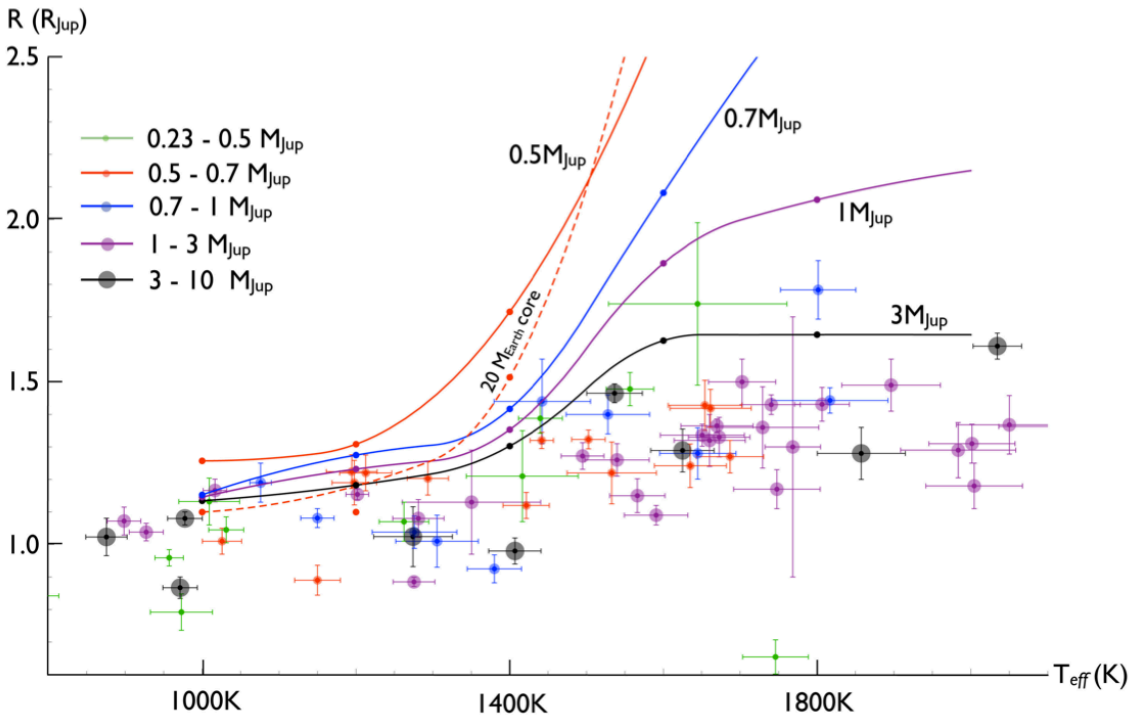


Figure 2.8: Same as Figure 2, but with $\epsilon = 5\%$

Jupiters, are illustrated in Figures 6-8. The evolved final model radii are plotted as points with spline interpolation between them and are labeled according to their masses. The observational data[3] are plotted as points with their corresponding 1σ error bars. The data are binned in mass, to fall between our model masses, and are labeled by color. As a result, if the Ohmic mechanism is indeed correct, and unmodeled effects such as variable opacity and age make a small contribution (Burrows et al., 2007), all observed data of a given color should fall below the model curve of the same color.

[3]The data was acquired from the Extrasolar Encyclopedia (http://www.exoplanet.eu)

2.7 Discussion

As is clearly demonstrated, if the approximations inherent to our model are valid, Ohmic dissipation, along with the likely presence of high-metalicity cores, is able to account for all of the currently observed radius anomalies. Furthermore, the self-similar family of model curves reproduces the characteristic increase and overturn of the radii in the effective temperature region between $1200K$ and $1800K$. This feature is unique to the Ohmic dissipation mechanism and is a consequence of ionization. Recall that the primary source of electrons in the atmosphere is alkali metals, with K playing a dominant role. Under densities that are characteristic of hot Jupiter atmospheres, K has an ionization temperature around $\sim 1800K$. This explains the rapid increase in Ohmic heating with rising T_{eff} around ~ 1400K. This increase comes about through a Boltzmann factor in the Saha equation, and will therefore have a roughly exponential dependence. We stress that this is uniquely characteristic of the Ohmic mechanism, since other proposed mechanisms will exhibit an algebraic dependence on T_{eff} (see Laughlin et al. (2011) for a statistical comparison of various inflation mechanisms with data).

As the temperature nears ~ 1800K, the ionization levels begin to saturate and the radiative/convective boundary moves deeper, so the radii don't grow as quickly with T_{eff}. However ionization of Na begins to contribute strongly above $2000K$. This renders all but the most massive of our models unstable. This is suggestive, given that out of a sample of 87 transiting Hot Jupiters, only four have been detected with $T_{eff} > 2100K$, although the mechanism by which Hot Jupiters halt their migration at small radii is still not fully understood.

It is noteworthy that our low-mass models significantly overestimate the data, to the extent where small ($\sim 20 M_{\oplus}$) cores are insufficient to explain the observed radii. This is not particularly surprising, given that the inflation is somewhat enhanced by a shallower adiabatic temperature gradient, g/c_p, allowing for a larger fraction of the generated heat to be deposited into the convective interior, and a diminished gravitational energy, that must be overcome by Ohmic heating. The discrepancy between the model and the data is either indicative of massive cores, or more likely curbed efficiency of the Ohmic mechanism at smaller mass. Both are certainly plausible, and the latter can be a consequence of diminished strength of the magnetic field, which is dictated by the intrinsic heat-flux (Christensen et al., 2009).

As the parameter space available to hot Jupiters is, to first order, delineated, the models presented in this work make two predictions. First and foremost, the radii of all hot Jupiters should fall below the theoretical curves presented in Figures 6-8. Second, a fraction of the hot sub-giant planet population may have originated from the giant planet population. Since this process can take place many Gyr after the birth of the star, Ohmically evaporating planets should be detected, given a sufficiantly large data set. These predictions will become testable, as the known aggregate of transiting exoplanets continues to grow. Accordingly, additional data will allow us to further understand the diversity of planetary bodies and to better define the Solar System's place in our galactic neighborhood.

We would like to conclude with a discussion of future prospects for improvement of the model. Let us begin by recalling the simplifications inherent to the approach presented here. In the structural calculation, we used a gray, one-dimensional atmosphere. Obvi-

ously this is not intended to be a quantitatively good representation of real atmospheres and improved treatment of the temperature (and thus conductivity) profile, accounting for asymmetry in insolation between the day and the night side as well as heat transport by global circulation is desired in the future. Indeed the considerable difference in interior cooling rates shown here, and those presented by Guillot & Havel (2011) suggests that the effect of changing the atmospheric boundary condition on structural evolution may be considerable. That said, the inability of current state-of-the-art atmospheric models to match observations in a satisfactory manner suggests that much additional effort must be applied towards improvement of our understanding of extra-solar planet atmospheres for implementation of a complex atmosphere into our model to be meaningful. In the heating calculation, our analytical parameterization of the circulation should be replaced by a self-consistent solution of the global circulation that fully accounts for magneto-hydrodynamic effects. The induced current in the interior would be a natural result of such a calculation. At the same time, the effect of the induced current on the interior dynamo should be assessed. Although in this study we assumed that the interior dynamo is largely unaffected by the currents induced in the atmosphere, in reality this may not be the case, since the induced current can in principle be comparable to the characteristic current of the interior dynamo. However, given the enormous computational cost of such an endeavor, it is not surprising that a calculation of this sort has not yet been done.

In general, future efforts in improvement of the Ohmic heating model should rely on both, modeling and observational constraints. After all, observations of primary and secondary eclipses can often yield information about the atmospheric temperature profile

(Knutson et al., 2008) as well as elemental abundances (Line et al., 2010). At the same time, ground-based high-dispersion spectroscopy can yield an estimate of the wind velocity (Snellen et al., 2010). In certain cases (e.g. WASP-12b), the planet's proximity to its host star results in considerable interactions between the planetary magnetosphere and the stellar wind. This can lead to formation of a bow-shock, which ultimately produces an observable signature, that manifests itself as an early ingress of the transit in near-UV frequency range. Quantitative modeling of this process may lead to meaningful constraints on the planetary magnetic field (Vidotto et al., 2010). As a result, an observationally guided improvement of the Ohmic inflation model will surely prove to be a useful venture, as quantitative enhancement of the first-order results presented in this paper will allow for a rigorous statistical comparison between the model and an expanded data set.

Acknowledgments We are indebted to A. Ingersoll, G. Laughlin, T. Guillot and T. Schneider for useful conversations, as well as the referee, Jonathan Mitchell for providing helpful suggestions.

Bibliography

Alexander, D. R., & Ferguson, J. W. 1994, The Astrophysical Journal, 437, 879

Batygin, K., & Stevenson, D. J. 2010, The Astrophysical Journal Letters, 714, L238

Batygin, K., Bodenheimer, P., & Laughlin, G. 2009, The Astrophysical Journal Letters, 704, L49

Baraffe, I., Chabrier, G., Barman, T. S., Selsis, F., Allard, F., & Hauschildt, P. H. 2005, Astronomy & Astrophysics, 436, L47

Bodenheimer, P., Lin, D. N. C., & Mardling, R. A. 2001, The Astrophysical Journal, 548, 466

Bodenheimer, P., Laughlin, G., & Lin, D. N. C. 2003, The Astrophysical Journal, 592, 555

Burrows, A., Hubeny, I., Budaj, J., & Hubbard, W. B. 2007, The Astrophysical Journal, 661, 502

Chabrier, G., & Baraffe, I. 2007, The Astrophysical Journal Letters, 661, L81

Chandrasekhar, S. 1957, [New York] Dover Publications [1957],

Cho, J. Y.-K., Menou, K., Hansen, B. M. S., & Seager, S. 2008, The Astrophysical Journal, 675, 817

Christensen, U. R., Holzwarth, V., & Reiners, A. 2009, Nature, 457, 167

Dodson-Robinson, S. E., Bodenheimer, P., Laughlin, G., Willacy, K., Turner, N. J., & Beichman, C. A. 2008, The Astrophysical Journal Letters, 688, L99

Fortney, J. J., & Nettelmann, N. 2010, Space Science Reviews, 152, 423

Freedman, R. S., Marley, M. S., & Lodders, K. 2008, The Astrophysical Journal Supplement, 174, 504

Guillot, T. 1999, Science, 286, 72

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

Guillot, T., & Havel, M. 2011, Astronomy & Astrophysics, 527, A20

Goldreich, P. 1963, Monthly Notices of Royal Astronomical Society, 126, 257

Heng, K., Menou, K., & Phillipps, P. J. 2011, Monthly Notices of Royal Astronomical Society, 370

Henyey, L. G., Forbes, J. E., & Gould, N. L. 1964, The Astrophysical Journal, 139, 306

Held, I. M., & Hou, A. Y. 1980, Journal of Atmospheric Sciences, 37, 515

Holton, J. R. 1992, International geophysics series, San Diego, New York: Academic Press, —c1992, 3rd ed.,

Hubbard, W. B. 1984, New York, Van Nostrand Reinhold Co., 1984, 343 p.

Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2005, Icarus, 179, 415

Hut, P. 1981, Astronomy & Astrophysics, 99, 126

Ibgui, L., Burrows, A., & Spiegel, D. S. 2010, The Astrophysical Journal, 713, 751

Knutson, H. A., Charbonneau, D., Allen, L. E., Burrows, A., & Megeath, S. T. 2008, The Astrophysical Journal, 673, 526

Laughlin, G., Crismani, M., & Adams, F. The Astrophysical Journal Letters, in press

Laine, R. O., Lin, D. N. C., & Dong, S. 2008, The Astrophysical Journal, 685, 521

Liu, J., Goldreich, P. M., & Stevenson, D. J. 2008, Icarus, 196, 653

Liu, J., & Schneider, T. 2010, Journal of Atmospheric Sciences, 67, 3652

Lo Curto, G., et al. 2010, Astronomy & Astrophysics, 512, A48

Line, M. R., Liang, M. C., & Yung, Y. L. 2010, The Astrophysical Journal, 717, 496

Perna, R., Menou, K., & Rauscher, E. 2010, The Astrophysical Journal, 724, 313

Peixoto, J. P., & Oort, A. H. 1992, New York: American Institute of Physics (AIP), 1992,

Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., & Greenzweig, Y., Icarus 1996, 124, 62

Saumon, D., Chabrier, G., & van Horn, H. M. 1995, The Astrophysical Journal Supplement, 99, 713

Schneider, E. K., & Lindzen, R. S. 1977, Journal of Atmospheric Sciences, 34, 263

Showman, A. P., Fortney, J. J., Lian, Y., Marley, M. S., Freedman, R. S., Knutson, H. A., & Charbonneau, D. 2009, The Astrophysical Journal, 699, 564

Showman, A. P., Cho, J. Y.-K., & Menou, K. 2010, Exoplanets, 471

Showman, A. P., Polvani, L. M. 2011. Equatorial superrotation on tidally locked exoplanets. ArXiv e-prints arXiv:1103.3101.

Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S. 2010, Nature, 465, 1049

Spiegel, D. S., Silverio, K., & Burrows, A. 2009, The Astrophysical Journal, 699, 1487

Stevenson, D. J. 1982, Annual Review of Earth and Planetary Sciences, 10, 257

Terquem, C., & Papaloizou, J. C. B. 2007, The Astrophysical Journal, 654, 1110

Thrastarson, H. T., & Cho, J. Y. 2010, The Astrophysical Journal, 716, 144

Vidotto, A. A., Jardine, M., & Helling, C. 2010, The Astrophysical Journal Letters, 722, L168

Youdin, A. N., & Mitchell, J. L. 2010, The Astrophysical Journal, 721, 1113

Chapter 3

Determination of the Interior Structure of Transiting Planets in Multiple-Planet Systems

Originally published as:

Batygin, K., Bodenheimer, P., & Laughlin, G. 2009, The Astrophysical Journal Letters, 704, L49

3.1 Abstract

Tidal dissipation within a short-period transiting extrasolar planet perturbed by a companion object can drive orbital evolution of the system to a so-called tidal fixed point, in which the apsidal lines of the transiting planet and its perturber are aligned, and for which variations in the orbital eccentricities of both planet and perturber are damped out. Maintenance of the fixed-point apsidal alignment requires the orbits of both planet and perturber to precess at the same rate. Significant contributions to the apsidal precession rate are made by the secular planet-planet interaction, by general relativity, and by the gravitational quadropole fields created by the transiting planet's tidal and rotational distortions. The precession arising from the planetary quadrupole can be the dominant term, and is strongly dependent on the planet's internal density distribution, which is in turn controlled by the

fractional mass of the planet incorporated into a heavy-element core. The fixed-point orbital eccentricity of the inner planet is therefore a strong function of the planet's interior structure. We illustrate these ideas in the specific context of the recently discovered HAT-P-13 exo-planetary system, and show that one can already glean important insights into the physical properties of the inner $P = 2.91\,\mathrm{d}$, $M = 0.85\,M_{\rm Jup}$, $R = 1.28\,R_{\rm Jup}$ transiting planet. We present structural models of the planet, which indicate that its observed radius can be maintained for a one-parameter sequence of models that properly vary core mass and tidal energy dissipation in the interior. We use an octopole-order secular theory of the orbital dynamics to derive the dependence of the inner planet's eccentricity, e_b, on its tidal Love number, k_{2b}. We find that the currently measured eccentricity, $e_b = 0.021 \pm 0.009$ implies $0.116 < k_{2b} < 0.425$, $0 M_\oplus < M_{core} < 120 M_\oplus$, and $10,000 < Q_b < 300,000$. Improved measurement of the eccentricity via transit and secondary eclipse timing, along with continued radial velocity monitoring, will soon allow for far tighter limits to be placed on all three of these quantities, and will provide an unprecedented probe into the interior structure of an extrasolar planet.

3.2 Introduction

The mounting detection rate for extrasolar planets inevitably produces a series of "firsts". Notable examples include 51 Peg b, the first Jovian-mass planet orbiting a main-sequence star (Mayor & Queloz 1995), Upsilon Andromedae b, c, and d, the first multiple extrasolar planet system (Butler et al. 1999), HD 209458 b, the first transiting planet (Charbonneau et al. 2000, Henry et al. 2000), and Gliese 581 e, the first truly terrestrial-mass planet (Mayor

et al. 2009).

The HAT-P-13 system (Bakos et al. 2009) presents an exoplanetary first that at first glance seems to lie perhaps one notch down on the novelty scale. This system contains the first transiting planet (HAT-P-13-b) that is accompanied by a well-characterized longer-period companion planet (HAT-P-13-c).

Transiting planets in multiple-planet systems have, however, been eagerly anticipated by the astronomical community. In configurations of this type, the planet-planet perturbations can lead to a host of observational effects (largely involving precise timing of the transits) that will potentially enable remarkable dynamical characterization of both the orbital and the planetary properties (see, for example, the review of Fabrycky 2009). On this count, the HAT-P-13 system does not disappoint. We show here that a combination of a tidal-secular orbital evolution model, coupled with interior evolution models of the inner planet, can be used to probe the planet's interior structure and to measure its current tidal quality factor, Q. Using information currently available, we outline this approach, and show that HAT-P-13-b has $Q < 300,000$.

The plan of this paper is as follows. In section 2, we describe the dynamics of a system at a tidal fixed point, and we outline the resulting connection between the interior structure of the planet and its orbital eccentricity. In section 3, we give an overview of the system and describe our interior evolution calculations. We show that, under the assumption of a tidal origin for the planet's inflated size, the observed planetary radius can be explained by a one-parameter sequence of models within a two-parameter space delineated by planetary core-mass and planetary tidal luminosity. We then summarize the prospects for improved

measurement of the orbital eccentricity, and proselytize the overall ramifications of our study.

3.3 A System at an Eccentricity Fixed Point

In a planetary system that resides far from a significant mean motion resonance, the non-Keplerian portion of the orbital motion can be well described by secular terms in the planetary disturbing function. In this paper, we focus on the particular circumstance in which a short-period transiting planet on a nearly circular orbit receives perturbations from a relatively distant companion planet lying on a significantly eccentric orbit. Due to the high eccentricity of the outer body, classical Laplace-Lagrange theory (e.g. Murray & Dermott 1999) cannot be used. Systems of this type are, however, well-described by the octopole-order secular theory presented in Mardling (2007). This theory requires that (1) the eccentricity of the inner planet be much less than that of the outer planet, and (2) that the mass of the inner planet be much less than that of the star. The theory makes no strong demands, however, on either the mass of the outer planet or on its eccentricity. If we assume coplanar orbits, the secular evolution generated by the planet-planet interaction is given by:

$$\dot{e_b} = -\frac{15}{16} n_b e_c \left(\frac{m_c}{M_\star}\right) \left(\frac{a_b}{a_c}\right)^4 \frac{\sin(\varpi_b - \varpi_c)}{(1-e_c^2)^{5/2}} - e_b \frac{21\pi}{P_b} \frac{k_{2b}}{Q_b} \frac{M_\star}{m_b} \left(\frac{R_b}{a_b}\right)^5 , \tag{3.1}$$

$$\dot{e_c} = \frac{15}{16} n_c e_b \left(\frac{m_b}{M_\star}\right) \left(\frac{a_b}{a_c}\right)^3 \frac{\sin(\varpi_b - \varpi_c)}{(1-e_c^2)^2} , \tag{3.2}$$

$$\dot{\varpi}_{bsecular} = \frac{3}{4} n_b \left(\frac{m_c}{M_\star}\right) \left(\frac{a_b}{a_c}\right)^3 \frac{1}{(1-e_c^2)^{3/2}} \left[1 - \frac{5}{4}\left(\frac{a_b}{a_c}\right)\left(\frac{e_c}{e_b}\right)\frac{\cos(\varpi_b - \varpi_c)}{1-e_c^2}\right] , \quad (3.3)$$

and

$$\dot{\varpi}_{csecular} = \frac{3}{4} n_c \left(\frac{m_b}{M_\star}\right) \left(\frac{a_b}{a_c}\right)^3 \frac{1}{(1-e_c^2)^{2}} \left[1 - \frac{5}{4}\left(\frac{a_b}{a_c}\right)\left(\frac{e_b}{e_c}\right)\frac{1+4e_c^2}{1-e_c^2}\cos(\varpi_b - \varpi_c)\right] , \quad (3.4)$$

where a corresponds to semi-major axis, e is eccentricity, ϖ is longitude of perihelion, n is mean motion, P is period, R is radius, m is mass, $M_\star$ is stellar mass, k_{2b} is the tidal Love number, and Q_b is the inner planet's effective tidal dissipation parameter.

The inner planet experiences additional contributions to its precession from the quadrupole potential that arises from the tidal and rotational bulges of the planet, and from the leading-order effects of general relativity. As discussed in Ragozzine & Wolf (2009), precession driven by the tidal and rotational bulges of the star is unimportant, unless the rotational period of the star is short (e.g. $P_{rot} \lesssim 10$ days). Derivations of the planet-induced tidal and rotational precessions are given in Sterne (1939), and are discussed in the planetary context by Wu & Goldreich (2002) and also, extensively, by Ragozzine & Wolf (2009). The relativistic advance has been discussed by many authors, for an up-to-date discussion in the extrasolar planet context see, for example Jordan & Bakos (2008). To linear order, we can treat the total precession of the inner planet as the sum of the four most significant contributions:

$$\dot{\varpi}_{btotal} = \dot{\varpi}_{bsecular} + \dot{\varpi}_{btidal} + \dot{\varpi}_{bGR} + \dot{\varpi}_{brotational} , \quad (3.5)$$

where, assuming synchronous rotation,

$$\dot{\varpi}_{btidal} = \frac{15}{2} k_{2b} \left(\frac{R_b}{a_b}\right)^5 \left(\frac{M_\star}{M_b}\right) f_2(e_b) n_b \,, \tag{3.6}$$

$$\dot{\varpi}_{bGR} = \frac{3 n_b^3}{1 - e_b^2} \left(\frac{a_b}{c}\right)^2 \,, \tag{3.7}$$

and

$$\dot{\varpi}_{brotational} = \frac{k_{2b}}{2} \left(\frac{R_b}{a_b}\right)^5 \frac{n_b^3 a_b^3}{G m_b} \left(1 - e_b^2\right)^{-2} \,. \tag{3.8}$$

The eccentricity function, $f_2(e_b)$ is given by

$$f_2(e_b) = \left(1 - e_b^{-2}\right)^{-5} \left(1 + \frac{3}{2} e_b^2 + \frac{1}{8} e_b^4\right) . \tag{3.9}$$

Tidal dissipation occurs primarily within the inner planet, and leads to continual decrease of the inner planet's semi-major axis through

$$\dot{a_b} = -e_b^2 a_b \frac{21\pi}{2 P_b} \frac{k_{2b}}{Q_b} \frac{M_\star}{m_b} \left(\frac{R_b}{a_b}\right)^5 \,. \tag{3.10}$$

When a system of the type modeled above is subjected to tidal friction, it evolves to a stationary configuration or a "fixed point" within ~ 3 circularization timescales (see Mardling 2007 for an in-depth discussion). Formally, a secular fixed point can be characterized by simultaneously aligned (or anti-aligned) apses and identical precession rates of the orbits. In other words, in the frame that precesses with the orbits, the system in stationary. It then follows (in the limit of large Q_b) that when $\dot{\varpi}_{btotal} = \dot{\varpi}_{csecular}$, we have

$\dot{e}_b = \dot{e}_c = 0$. When a fixed-point system is subjected to tidal dissipation (that is, has a finite Q_b) the eccentricities of both orbits decay slowly, and the system remains quasi-stationary.

To second order in eccentricity, the tidal luminosity of a spin-synchronous planet is given by (e.g. Peale & Cassen 1978)

$$\frac{dE_b}{dt} = \frac{21}{2}\frac{k_{2b}}{Q_b}\frac{GM_\star^2 n_b R_b^5 e_b^2}{a_b^6}\,, \tag{3.11}$$

Note that if $e_b > 0$ then the planet cannot be fully spin synchronized. Further, if the planet is a fluid body, it will be unable to maintain a permanent quadropole moment, and will therefore not reside in spin-orbit resonance. The pseudo-synchronization theory of Hut (1981); see also Goldreich & Peale (1966) can be used to calculate the spin frequency (which for small e_b approaches n_b)

$$\frac{\Omega_{\rm spin}}{n_b} = \frac{1 + \frac{15}{2}e_b^2 + \frac{45}{8}e_b^4 + \frac{5}{16}e_b^6}{(1 + 3e_b^2 + \frac{3}{8}e_b^4)(1 - e_b^2)^{3/2}} \tag{3.12}$$

The analysis of Levrard et al. (2007), furthermore, indicates that this spin asynchronicity of the planet will cause the tidal luminosity to exceed that given by the above formula by a small amount.

The tidal Love number, k_{2b} parameterizes the degree of central condensation in the fluid inner transiting planet. The mass distribution in turn affects the total tidal luminosity through Equation (11) and contributes to the orbital precession rate through $\dot{\varpi}_{btidal}$. The quantity k_{2b} therefore provides an explicit connection between the interior sturcture and energetics of the planet on the one hand, and the orbital dynamics on the other. If the

density distribution, $\rho(r)$, in a planet is available, then calculation of k_{2b} is straightforward (Sterne 1939):[1]

$$k_{2b} = \frac{3 - \eta_2(R_{Pl})}{2 + \eta_2(R_{Pl})}\,, \tag{3.13}$$

where $\eta_2(R_{Pl})$ is obtained by integrating an ordinary differential equation for $\eta_2(r)$ radially outward from $\eta_2(0) = 0$.

$$r\frac{d\eta_2}{dr} + \eta_2^2 - \eta_2 - 6 + \frac{6\rho}{\rho_m}(\eta_2 + 1) = 0\,, \tag{3.14}$$

where ρ_m is the mean density interior to r and $R_{\rm Pl}$ is the outer radius of the planet. It is noteworthy that a fixed-point system has no hysteresis. In other words, the above calculation indirectly yields a measurement of the current k_2, independent of the system's dynamical history, or the structural evolution of the planets.

3.4 Application to the HAT-P-13 Planetary System

The theory discussed above finds an ideal application in the context of HAT-P-13. As discussed in Bakos et al (2009, hereafter B09) this system was discovered in a wide-field photometric survey, and was later confirmed and characterized with high-precision radial velocities. The system contains an inner, transiting, jovian-mass planet, "b", and an outer body, "c", with $M_c \sin(i_c)$ close to the giant-planet brown dwarf boundary. The $V = 10.65$ G4V parent star, formerly best known as GSC 3416-00543, was essentially unstudied prior to the photometric detection of its inner planet, and so, as a result, all quoted planetary

[1] note that the quantity $k_{2,1}$ defined in Sterne (1939) is the *apsidal motion constant*, that is, $k_{2b}/2$ in our notation

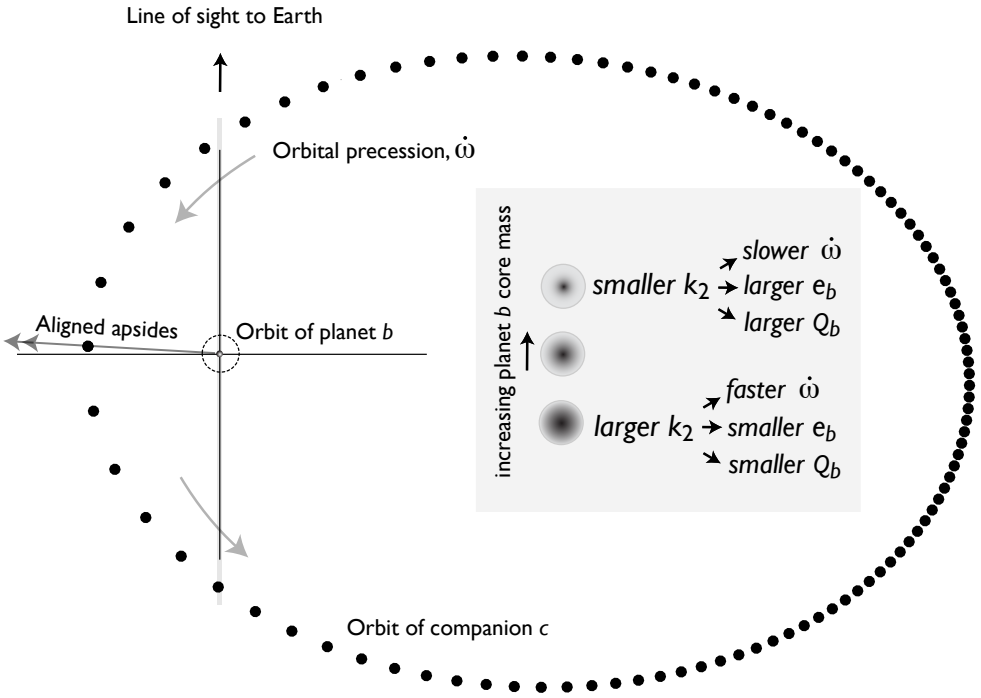


Figure 3.1: A representation of the orbital architecture of the HAT-P-13 system to scale. The inset schematic illustrates the dependencies of k_{2b}, Q_b, $\dot{\varpi}$ and e_b on the mass of the planet's heavy element core.

and stellar properties are drawn from the B09 discovery paper. For reference, we note that B09 derive an inner planet mass of $m_b = 0.851^{+0.029}_{-0.046}\,M_{\rm Jup}$, a period $P_b = 2.91626 \pm 0.00001$ days, and an eccentricity (measured from a fit to 32 radial velocity measurements) of $e_b = 0.021 \pm 0.009$. The outer companion has $M_c \sin(i_c) = 15.2 \pm 1.0\,M_{\rm Jup}$, period $P_c = 428.5 \pm 3.0$ days, and eccentricity $e_c = 0.691 \pm 0.018$. To within the significant observational uncertainty, the apsidal lines of the two planets are aligned. The parent star has $M_\star = 1.22^{+0.05}_{-0.10}\,M_\odot$, and the planetary radius is $R = 1.28 \pm 0.079 R_{\rm Jup}$.

The short orbital period and non-zero eccentricity of HAT-P-13b suggest that tidal circularization should be highly effective over the presumed multi-Gyr age of the star. In this case, dissipative secular evolution will have brought the system to a fixed point configuration, which to high accuracy satisfies the constraint given by

$$\dot{\varpi}_{btotal}(e_b, k_{2b}) = \dot{\varpi}_{csecular}\,. \tag{3.15}$$

We note that although perturbations from potential companions below detectability may modify our solution for the eccentricity slightly, the system as currently characterized from the observations is fully consistent with such a configuration.

Our approach is illustrated schematically in Figure 1, and operates as follows. For given $M_{\rm Pl}$, $T_{\rm eff}$ and $R_{\rm Pl}$ (all of which are strongly constrained by the observations) we compute planetary interior structure and evolution models with a descendant of the Berkeley stellar evolution code (Henyey et al. 1964). This program assumes that the standard equations of stellar structure apply, and it has been used extensively in the study of both extrasolar and solar system giant planets (see e.g. Pollack et al. 1996, Bodenheimer et al. 2003, Hubickyj

et al. 2005, and Dodson-Robinson et al. 2008 for descriptions of the method and its input physics). Energy sources within the planet include gravitational contraction, cooling of the interior, and tidal heating of the interior, which we assume occurs at adiabatic depth. It is important to keep in mind that the issue of where tidal heat actually gets deposited in gaseous bodies is not well understood, and is a topic of ongoing research (Ogilvie & Lin, 2004, Wu 2005). In the context of our model, the planet is considered to be quasi-steady-state. Namely, due to tidal heating of the interior, the planet is not subject to secular cooling, and therefore is not contracting. At the planetary surface, the luminosity is composed of two components: the internal luminosity generated by the planet, $L_{int} = dE/dt$, and the energy absorbed from the stellar radiation flux and re-radiated (insolation). Pure molecular opacities are used in the radiative outer layers of the planet (Freedman et al. 2008). A given evolutionary sequence starts at a radius of roughly 2 $R_{\rm Jup}$ and ends at an age of 4.5 Gyr. The resulting planetary radius is highly insensitive to the chosen age. It is assumed that the planet arrived at its present orbital position during or shortly after formation, that is, at an age of $< 10^7$ yr.

We used the code to delineate a range of plausible models for HAT-P-13b, all of which are required to match the observed planetary mass, $M = 0.85 M_{\rm Jup}$, and the inferred planetary effective temperature, $T_{\rm eff} = 1649\,{\rm K}$. Our models are divided into three sequences, which are required to match the observed one-sigma lower limit on the planetary radius, $R_{\rm Pl} = 1.20\,R_{\rm Jup}$ (sequence 1), the best-fit radius $R_{\rm Pl} = 1.28\,R_{\rm Jup}$ (sequence 2), and the one-sigma upper limit on the observed radius, $R_{\rm Pl} = 1.36\,R_{\rm Jup}$ (sequence 3). The structurally relevant unknown parameters are the planet's solid core mass, M_{core}, and the total

tidal luminosity, $L_{int} = dE/dt$. By computing a variety of models in this 2-dimensional parameter space, we can pin down the ($M_{\rm core}$, L_{int}) pairs that generate planets that satisfy a given choice of ($M_{\rm Pl}$, $R_{\rm Pl}$, $T_{\rm eff}$). A range-spanning aggregate of the models is listed in Table 1.

Also listed in Table 1 are the tidal Love numbers, k_{2b} (obtained from $\rho(r)$ via Equations 13 and 14), the tidal quality factors, Q_b (obtained from equation 11), and the fixed point eccentricities, e_b that enable the satisfaction of equation 15. Quoted errors on Q_b and e_b are obtained by using B09's reported uncertainties on the observed planetary and orbital properties, and adopting the assumptions that the error distributions are both normally distributed and uncorrelated across parameters. Note that the only time the assumption of tidal heating is inferred is in the calculation of Q. The Love number is determined entirely from the orbital configuration.

Given our aggregate of models, we can consider the effect of the Love number, k_{2b} on the orbital architecture in more detail. If we completely ignore the orbital precession induced by the planet's tidal and rotational bulges, and adopt B09's best-fit measurements of all relevant orbital parameters other than e_b, the equilibrium inner planet eccentricity is $e_b^{eq} = 0.0336$. We verified with a fully numerical 3-body simulation (Chambers 1999), that includes the relativistic precession, that this value is indeed an excellent approximation to the fixed-point eccentricity. If we assume a Love number, $k_{2b} = 0.3$, and include the precession due to planetary rotational and tidal bulges, the equilibrium eccentricity drops to $e_b^{eq} = 0.0161$, a difference, Δe_b, that is eminently detectable. Therefore, with the precise estimate of planet b's eccentricity, that will emerge from secondary transit timing and

Table 3.1: Stationary Orbital and Interior Models of HAT-P-13b

Core Mass ($M_\oplus$)	R (R_{Jup})	dE/dt (erg/s)	k_2	Q	e
0	1.20	3.49 $\times 10^{25}$	0.425	228561 $\pm$ 65014	0.0165 $\pm$ 0.0032
0	1.28	1.78 $\times 10^{26}$	0.38	42054 $\pm$ 14864	0.0143 $\pm$ 0.0036
0	1.36	5.6 $\times 10^{26}$	0.34	12875 $\pm$ 4396	0.0129 $\pm$ 0.0030
40	1.20	2.485 $\times 10^{26}$	0.297	30868 $\pm$ 8603	0.0195 $\pm$ 0.0036
40	1.28	8.45 $\times 10^{26}$	0.26	8924$\pm$ 2798	0.0176 $\pm$ 0.0037
40	1.36	1.93 $\times 10^{27}$	0.228	3959 $\pm$ 1154	0.0162 $\pm$ 0.0034
80	1.20	1.028 $\times 10^{27}$	0.217	6810 $\pm$ 2049	0.0221 $\pm$ 0.0049
80	1.28	2.53 $\times 10^{27}$	0.187	3036 $\pm$ 929	0.021 $\pm$ 0.0041
80	1.36	4.96 $\times 10^{27}$	0.163	1535 $\pm$ 453	0.0193 $\pm$ 0.0038
120	1.20	3.25 $\times 10^{27}$	0.159	1967 $\pm$ 618	0.0246 $\pm$ 0.0046
120	1.28	6.88 $\times 10^{27}$	0.135	997 $\pm$ 273	0.0226 $\pm$ 0.0039
120	1.36	1.27 $\times 10^{28}$	0.116	563 $\pm$ 182	0.0216 $\pm$ 0.0042

additional radial velocity measurements, it will be possible to make solid inferences about planet b's core mass and internal luminosity from direct measurement of the tidal Love number, k_{2b}. In the context of this system, it is important to note that the orbital precession of the planets is quite slow ($\sim 10^{-3}$ deg/yr). As a result, k_{2b} cannot be measured directly from transit light curves, as described in (Ragozzine & Wolf, 2009), and must be inferred from the equilibrium eccentricity.

To further illustrate this idea, we obtained a series of equilibrium eccentricity values as a function of k_{2b}. The results from these calculations are shown in figure 2. Modeling the errors as described above, the mean trend of planet b's eccentricity can be approximated by

a fourth-order polynomial as

$$e_b^{eq} \approx 0.0334 - 0.0985 k_{2b} + 0.188 k_{2b}^2 - 0.184 k_{2b}^3 + 0.069 k_{2b}^4. \quad (3.16)$$

The plotted errors of the equilibrium eccentricities are the standard deviations obtained from each sample of e_b^{eq}'s for a given k_{2b}. We stress that these error bars will shrink very significantly with improved observational measurements obtainable from photometry, timing, and radial velocity.

In addition to the mean trend and errors, specific regions in (e_b, k_{2b}) space, occupied by a set of interior models with 0, 40, 80, and 120 $M_\oplus$ core masses (see Table 1) are also marked. These regions are represented as four quadrilaterals overlaying the graph. The corners of each quadrilateral correspond to the combination of the Love number for a given model, specific to a radius of 1.2 R_{Jup} or 1.36 R_{Jup}, and 1-σ bounds on its equilibrium fixed point eccentricity as determined by Equation 15. An increased core mass tends to lower the Love number. Accordingly, the left-most quadrilateral on Figure 2 corresponds to the 120 $M_\oplus$ core model, while the right-most quadrilateral represents the core-less model.

Recall again, that we have neglected the precession induced by the star's tidal bulge and rotation. While unlikely, if these effects shall turn out to be important, they will lead to a further decrease in the equilibrium eccentricity of planet b. In this case, the star's Love number, $k_{2\star}$, will also enter the calculation. In addition, we are assuming a co-planar configuration. The validity of this assumption will be tested by forthcoming transit timing measurements, and a further clue will be provided by measurement of the alignment of the inner planet's orbit with the stellar equator via the Rossiter-McLaughlin effect. If the

mutual inclination of the orbits is significant, a similar calculation can still be performed to obtain a modified fixed-point eccentricity, and thus a somewhat different Love number.

Given the non-zero eccentricity of planet b, a natural question emerges: how long will planet b remain slightly eccentric despite tidal dissipation? To answer this question, we performed a tidally dissipated secular integration, similar to those discussed in Batygin et al. (2009) using an artificially low tidal quality factor of $Q_b = 10$ to speed up the proceedings. This integration revealed that the *e*-folding time for planet b's eccentricity is $\tau \approx 5.78(Q_b/10) \times 10^5$ years, or approximately 6 Gyr for $Q_b = 10^5$. As a consequence, we expect that the orbital configuration of the system *has* evolved somewhat during the current lifetime of the star. This timescale also places the low values of Q_b, e.g. $Q_b < 10,000$, that are currently admitted by the observations into disfavor.

3.5 Conclusion

Our analysis indicates that the HAT-P-13 system has the near-immediate potential to give startlingly detailed information about the density structure and the efficiency of tidal dissipation in the interior of an extrasolar planet. When high-precision (yet fully feasible) refinements of the orbital parameters are obtained, we will gain a precise and accurate measurement of the tidal quality factor, Q_b, of HAT-P-13b - similar to those that we currently have for the solar system giant planets. Furthermore, it is seems reasonable to assume that additional examples of systems that contain a transiting planet at a well characterized tidal fixed point will soon emerge from the ongoing photometric and doppler velocity surveys.

We therefore encourage immediate observational effort to obtain an improved charac-

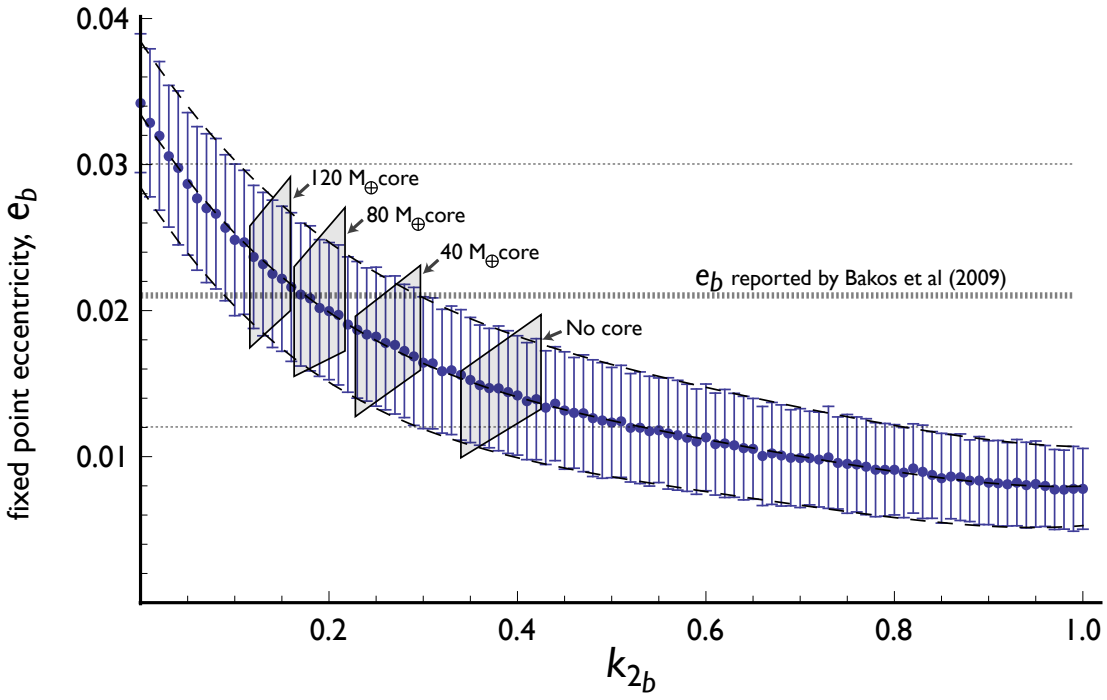


Figure 3.2: Stationary eccentricity of HAT-P-13b as a function of its Love number, k_{2b}, with error bars. Each blue dot represents the sample mean of the computed fixed point eccentricities. The dashed lines are best-fit fourth order polynomials. The four quadrilaterals are the approximate regions of the (e_b, k_{2b}) space occupied by each of the models presented in Table 1.

terization of the HAT-P-13 system, and we reiterate the importance of the wide-field surveys (such as HAT Net) that can locate transiting planets orbiting the *brightest available* stars in the sky.

Acknowledgments We thank Darin Ragozzine and Dan Fabrycky for useful discussions and suggestions for the manuscript. We also thank the anonymous referee for insightful comments. This research is based in part upon work supported by NASA Grant NNX08AH82G (PB) and by the National Science Foundation CAREER program under Grant No. 0449986 (GL).

Bibliography

Bakos, G. A., et al. 2009, arXiv:0907.3525

Batygin, K., Laughlin, G., Meschiari, S., Rivera, E., Vogt, S., & Butler, P. 2009, The Astrophysical Journal, 699, 23

Bodenheimer, P., Laughlin, G., & Lin, D. N. C. 2003, The Astrophysical Journal, 592, 555

Butler, R. P., Marcy, G. W., Fischer, D. A., Brown, T. M., Contos, A. R., Korzennik, S. G., Nisenson, P., & Noyes, R. W. 1999, The Astrophysical Journal, 526, 916

Chambers, K. C. 1999, Bulletin of the American Astronomical Society, 31, 1547

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, The Astrophysical Journal Letters, 529, L45

Dodson-Robinson, S. E., Bodenheimer, P., Laughlin, G., Willacy, K., Turner, N. J., & Beichman, C. A. 2008, The Astrophysical Journal Letters, 688, L99

Fabrycky, D. C. 2009, IAU Symposium, 253, 173

Freedman, R. S., Marley, M. S., & Lodders, K. 2008, The Astrophysical Journal Supplement, 174, 504

Goldreich, P., & Peale, S. J. 1966, Nature, 209, 1117

Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, The Astrophysical Journal Letters, 529, L41

Henyey, L. G., Forbes, J. E., & Gould, N. L. 1964, The Astrophysical Journal, 139, 306

Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2005, Icarus, 179, 415

Hut, P., 1981, Astronomy & Astrophysics, 99, 126

Jordán, A., & Bakos, G. Á. 2008, The Astrophysical Journal, 685, 543

Levrard, B., Correia, A. C. M., Chabrier, G., Baraffe, I., Selsis, F., & Laskar, J. 2007, Astronomy & Astrophysics, 462, L5

Mardling, R. A. 2007, Monthly Notices of Royal Astronomical Society, 382, 1768

Mayor, M., & Queloz, D. 1995, Nature, 378, 355

Mayor, M., et al. 2009, arXiv:0906.2780

Murray, C. D., & Dermott, S. F. 1999, Solar System Dynamics (Cambridge: Cambridge Univ. Press) (MD99)

Ogilvie, G. I., & Lin, D. N. C. 2004, The Astrophysical Journal, 610, 477

Peale, S. J., & Cassen, P. 1978, Icarus, 36, 245

Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., & Greenzweig, Y. 1996, Icarus, 124, 62

Ragozzine, D., & Wolf, A. S. 2009, The Astrophysical Journal, 698, 1778

Sterne, T. E. 1939, Monthly Notices of Royal Astronomical Society, 99, 451

Wu, Y., & Goldreich, P. 2002, The Astrophysical Journal, 564, 1024

Wu, Y. 2005, The Astrophysical Journal, 635, 674

Chapter 4

Early Dynamical Evolution of the Solar System: Pinning Down the Initial Condition of the Nice Model

Originally published as:
Batygin, K., & Brown, M. E. 2010, The Astrophysical Journal, 716, 1323

4.1 Abstract

In the recent years, the "Nice" model of solar system formation has attained an unprecedented level of success in reproducing much of the observed orbital architecture of the solar system by evolving the planets to their current locations from a more compact configuration. Within the context of this model, the formation of the classical Kuiper belt requires a phase during which the ice giants have a high eccentricity. An outstanding question of this model is the initial configuration from which the Solar System started out. Recent work has shown that multi-resonant initial conditions can serve as good candidates, as they naturally prevent vigorous type-II migration. In this paper, we use analytical arguments, as well as self-consistent numerical N-body simulations to identify fully-resonant initial conditions, whose dynamical evolution is characterized by an eccentric phase of the ice-giants, as well

as planetary scattering. We find a total of eight such initial conditions. Four of these primordial states are compatible with the canonical "Nice" model, while the others imply slightly different evolutions. The results presented here should prove useful in further development of a comprehensive model for solar system formation.

4.2 Introduction

The question of how the solar system formed dates back centuries. The last decade, however, has seen a considerable amount of progress made on this issue. Notably, the development of the Nice model (Tsiaganis et al. 2005) has proven to be a milestone. The scenario foretold by the Nice model is as follows: the giant planets form in a compact configuration, and driven by planetesimal scattering (Fernandez & Ip 1984, Malhotra 1995) begin migrating divergently. Eventually, Jupiter and Saturn cross their mutual 2:1 mean-motion resonance (MMR), which results in an acquisition of eccentricities for both planets. Subsequently, the whole outer solar system undergoes a brief period of dynamical instability, during which Uranus and Neptune are scattered to their current orbits.

There are a few aspects to the success of the Nice model. First and foremost, it has been able to replicate the architecture of the secular dynamics of the outer solar system (Morbidelli et al. 2009a). Second, it provides a semi-quantitative description of the formation of the Kuiper belt (Levison et al. 2008). Third, the inward flux of planetecimals during the phase of dynamical instability allows for chaotic capture of Jupiter's and Neptune's Trojan populations (Morbidelli et al. 2005, Nesvorniý et al. 2007). Finally, if the resonance crossing between Jupiter and Saturn is timed appropriately, the global mayhem provides a

natural trigger for Late Heavy Bombardement (LHB) (Gomes et al. 2005). There are other observational constraints that should be reproduced in a model for the solar system's formation, such as the dynamical structure of the inner solar system, and considerable progress has already been made in this direction (Brasser et al. 2009, Morbidelli et al. 2009b). At the same time, it is also crucial to explore the unobservable aspect of the model, namely the initial conditions. This is the purpose of our study.

In this paper, we consider various multi-resonant configurations as possible initial conditions for the Nice model. We investigate the early stages of dynamical evolution for a large number of candidate systems and show that only eight configurations appear to be consistent with the formation of the Kuiper belt in the framework of the Nice model. It is noteworthy that running simulations of the Nice model to completion is computationally expensive. Consequently, as we seek to examine a large array of initial conditions, we are forced to utilize early dynamical events, namely planetary scattering, as proxies for successful formation of the solar system. In this manner, we limit the duration of each simulation to only a few tens of millions of years. Within the context of our integrations, this is long enough for the system to pass through the epoch of dynamical instability, but not long enough to scatter away all of the planetesimals which end up on long-term unstable orbits.

The plan of the paper is as follows: in section 2, we explain how multi-resonant configurations prevent type-II migration, and our approach to their assembly. In section 3, we discuss the evolution scenarios of the considered systems, estimate the amplitudes of eccentricity jumps in relevant cases, and present the results of N-body simulations. We

discuss our results and conclude in section 4.

4.3 Multi-Resonant Configurations

One of the important differences between the solar system and the majority of the detected aggregate of extra-solar planetary systems is the lack of a close-in giant planet. This difference suggests that while it is common for planets to migrate to small orbital radii, some mechanism was at play in the early solar system which prevented vigorous orbital decay. One such mechanism, which is both efficient and reasonable, is resonant capture (Masset & Snellgrove 2001).

When a newly formed gaseous planet reaches a critical mass of $\sim 1M_J$, it opens a gap in the proto-planeatary disk. Incidentally, the planet continues to interact with the disk via various resonances. Summed together, the resonant torques from a given side of the disk, somewhat counter-intuitively, push the planet away from that side. As a result, the planet positions itself at a point in the gap where all torques cancel, and moves inward together with the disk on the viscous time-scale, in a process termed Type-II migration (Morbidelli & Crida 2007). Simultaneously if another planet, whose semi-major axis is larger, is migrating inwards faster, it will eventually encounter a mean-motion resonance. Under a large spectrum of circumstances, converging orbits can lead to capture into a mean-motion resonance, ensuring that the two planets' period ratio remains constant for extended periods of time. In fact, for slow enough migration rates and low enough eccentricities, resonant capture is certain (Peale 1986). When this happens, the gas between the two planets drains as the gaps overlap. Consequently, the torque balance on the resonant pair

results from gas interior to the inner planet and that exterior to the outer planet. This leads to a drastic reduction of the migration rate (Lee & Peale 2002). Furthermore, if the inner planet is more massive than the outer planet, as is the case with Jupiter and Saturn, the migration of the resonant pair can be halted altogether or even reversed (Morbidelli & Crida 2007).

Numerical studies of Jupiter and Saturn submerged in a gaseous proto-planetary disk suggest precisely the above scenario. Indeed, Saturn's migration is considerably faster than Jupiter's because of its lower mass and inability to fully open a clean gap. The pioneering results of Masset & Snellgrove (2001) showed that locking Jupiter and Saturn in the 3:2 MMR can effectively halt the pair's migration. The somewhat more precise numerical experiments of Morbidelli & Crida (2007) confirmed this and also showed that capture into 2:1 and 5:3[1] MMR's are viable outcomes, depending on where Saturn forms relative to Jupiter. The work of Pierens & Nelson (2008) however suggests that while capture into 2:1 and 5:3 MMR's is certainly possible, in a number of cases Saturn eventually breaks away and continues its inward migration until it is captured in the 3:2 MMR. Collectively, the above mentioned results suggest that 3:2 MMR is indeed a likely initial configuration of Jupiter & Saturn, although there is not enough evidence to decisively rule out the 2:1 MMR or the 5:3 MMR as initial conditions. Consequently, for the sake of completeness, we consider all three of these resonances as possible starting configuratoins for Jupiter and Saturn.

By extension of the above scenario, the ice giants, which are believed to form after

[1]It must be noted that capture into 5:3 MMR is less probable, since it is a second order resonance. Furthermore, even if Jupiter and Saturn are captured, subsequent motion can be unstable (Morbidelli & Crida 2007).

Jupiter and Saturn, behave in a qualitatively similar way. Namely, as they migrate from the outer disk inwards, they too become trapped in MMR's. Consequently, at the epoch of the disappearance of the gas, we are left with a multi-resonant system, in which each planet is in resonance with its neighbors. Morbidelli et al. (2007) performed hydrodynamical simulations of this process with the purpose of identifying such configurations that are long-term stable. Considering only systems where Jupiter and Saturn are locked in a 3:2 MMR, they were able to find two stable fully-resonant states.

Our approach to assembling multi-resonant systems follows that of Lee & Peale (2002). In addition to Newtonian N-body interactions, each planet is subject to semi-major axis decay

$$\frac{\dot{a}}{a} = K, \tag{4.1}$$

and eccentricity damping

$$\frac{\dot{e}}{e} = 10^2 \left(\frac{\dot{a}}{a} \right), \tag{4.2}$$

where a is semi-major axis, e is the eccentricity, and K is an adjustable migration frequency. In our simulations, we keep K the same for all planets, ensuring always convergent migration. In accord with Lee & Peale (2002), a Bulirsch-Stoer integration method (Press et al. 1992) was used. In contrast with the full hydrodynamical simulations, this method is simpler and computationally cheaper, allowing us to sample a large array of systems. Additionally, given the problem's straight-forward nature, it is unlikely that a configuration found using this approach cannot be obtained using other methods.

In the context of these simulations, all four giant planets were introduced simultane-

ously on planar circular orbits, slightly outside of their desired resonant locations, with the more massive ice giant on the outermost orbit. If capture into the desired resonances did not occur, we varied K. As pointed out in Morbidelli et al. (2007), the sequence in which planets get captured may be important, since changing the order can change the librating resonant angles. To avoid this degeneracy, we always set the initial orbits such that Saturn would get captured first and Neptune last. To avoid confusion, we shall always refer to the outermost planet as Neptune, although in the Nice-model, the orbits of the ice giants may switch places. After each desired configuration was achieved, the 5-body system was subjected to a 100Myr dynamical stability test, using the mercury6 software (Chambers 1999). Note that 100Myr is highly conservative, given that our full dynamical evolution simulations only last $\sim$30Myr. However, these stability integrations show that the same initial conditions are also applicable for scenarios where the global instability occurs somewhat later than what is considered in this work. Table (1) lists all stable multi-resonant configurations that we generated.

Table 4.1: Multi-resonant Initial Conditions

J:S	S:U	U:N
3:2	2:1	3:2
3:2	2:1	4:3
3:2	3:2	3:2
3:2	3:2	4:3
3:2	**3:2**	**5:4**
3:2	**4:3**	**3:2**
3:2	**4:3**	**4:3**
5:3	2:1	3:2
5:3	2:1	4:3
5:3	2:1	5:4
5:3	2:1	6:5
5:3	3:2	3:2
5:3	3:2	4:3
5:3	**3:2**	**5:4**
5:3	**3:2**	**6:5**
5:3	**4:3**	**3:2**
5:3	**4:3**	**4:3**
5:3	4:3	5:4
2:1	2:1	3:2
2:1	2:1	4:3
2:1	3:2	3:2
2:1	3:2	4:3
2:1	4:3	3:2
2:1	**4:3**	**4:3**

All of the stable multi-resonant initial conditions considered in this study. The bold lines represent the configurations that proved to be compatible with a Nice model-like evolution.

4.4 Dynamical Evolution

Having found a large array of stable multi-resonant systems, we now need to determine which of these configurations can resemble the current state of the solar system, having dynamically evolved. There are two constraints of interest here. First is the structure of the secular dynamics of the giant planets. Recently, Morbidelli et al. (2009a) showed that a smooth migration scenario, such as the one envisioned by Malhotra (1995), is incompatible with the observed eccentricities and inclinations of the giant planets. A mean motion resonance crossing event by itself is also insufficient because it does not excite the inclinations to a necessary degree or reproduce the amplitudes of the g_5 and g_6 secular eigenmodes correctly. To create the current eccentricities, inclinations, and eigenmode amplitudes, encounters must have happened between an ice giant and a gas giant. Furthermore, if the instability took place after the inner solar system was already intact, encounters must have taken place between an ice giant and both gas giants to cause Jupiter to "jump", in order to prevent slow secular resonance sweeping of the inner solar system (Brasser et al. 2009).

The second criterion of interest is the formation of the Kuiper belt, particularly the classical region. The transport mechanism of planetesimals to this region, proposed by Levison et al. (2008) relies on overlapping mean motion resonances. Namely, when Neptune's eccentricity exceeds ~ 0.2, its exterior MMR's widen enough to overlap, and motion of all particles in the region becomes highly chaotic. This allows for planetesimals to execute a random-walk and invade the classical Kuiper belt region. With time, as Neptune's eccentricity decays due to dynamical friction (Stewart & Withrill 1988), the resonances become narrower, and the particles occupying the classical region, no longer chaotic, remain on

their orbits forever.

Using the two constraints described above as proxies for a successful formation scenario, we look for a subset of our generated initial conditions that result in evolutions which encompass both a scattering event and a transient high-eccentricity phase of the outer ice giant. In the Nice model, these two constraints are practically always satisfied simultaneously, since the ice giants tend to switch places and the scattering event is intimately tied to the high eccentricities.

Our dynamical evolution simulations include the four outer planets and a disk of ~ 3000 equal-mass planetesimals, while the mass of the inner solar system is added to the sun. The radial surface density of the planetesimal swarm was assumed to have a power law structure: $\Sigma \propto r^{-k}$, where $k \in (1, 2)$. Consequently, the set of initial-value problems at hand is controlled by three parameters: Jupiter's semi-major axis, which due to resonant relations controls the semi-major axes of the other planets, the planetesimal disk's mass, m_{disk}, and k. As it turns out, the actual value of k has little effect on whether a given initial condition gives rise to a scattering event and an eccentric Neptune. Rather, it controls how fast the planets grind through the disk. As a result, we allow it to float randomly from simulation to simulation. Also, an advantage of multi-resonant initial conditions lies in that with enough simulations, it is possible to determine a unique combination of (a_J, m_{disk}), since Jupiter and Saturn must scatter enough particles to arrive to their current 5:2 commensurability.

In our simulations, the inner edge of the planetesimal disk was placed ~ 1AU outside of Neptune's orbit. This forces migration, driven by planetesimal scattering, to begin shortly after the start of the simulation. In other words, we do not attempt to time the onset of

instability with LHB, as was done in Gomes et al. (2005). There is another implication of a relatively close inner edge. If the planetesimal swarm is nearby when the instability begins, the planetary orbits penetrate deeply into the disk, and the resulting dynamical friction plays a stabilizing effect (Levison, personal communication). Intuitively, it makes sense to place the inner edge of the disk where the dynamical lifetime of planetesimals equals the lifetime of the gaseous nebula. While a 1 AU separation is approximately correct (Gomes et al. 2005), in the future it may be a worthwhile exercise to determine this boundary precisely for each multi-resonant initial condition. The outer edge of the disk was placed at 30 AU with the purpose of eventually halting Neptune's migration (Tsiaganis et al. 2005, Levison et al. 2008).

In order to not force the scattering-driven migration fictitiously, but still keep the computational cost down, we ignore self-gravity of the planetesimal disk. The resulting Hamiltonian takes the form:

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} - G \sum_{i=1}^{5} m_i \sum_{j=i+1}^{N} \frac{m_j (x_i - x_j)}{|x_i - x_j|^2} \tag{4.3}$$

where p is momentum, x is position, m is mass and G is the universal gravitational constant. To further diminish the computational cost, after a particle was scattered beyond 500AU, it was removed from the simulation. It must be noted that planetesimal-planetesimal interactions may in reality be important, since they give rise to an effective viscosity in the swarm (Levison, personal communication). Thus, more careful validation of our results should preferentially include these thorny effects. We used a hybrid Bulisch-Stoer/Wisdom-Holman algorithm of the mercury6 software for all integrations. We consistently used a time-step

of $\tau = 300$d, and checked all successful simulations with a smaller $\tau = 60$d time-step to ensure that the observed instability is not a numerical artifact. In all such checks, the evolutions were practically indistinguishable, which assures that integrals of motion are sufficiently conserved. Finally, our simulations only cover a few tens of millions of years, since the focus here is on distinguishing between initial conditions that give rise to gas-giant/ice-giant scattering and ones that don't. As a result, the long-term evolution of the system after the instability is unexplored.

4.4.1 Initial Conditions with Jupiter and Saturn in a 3:2 MMR

Let us first consider a family of initial conditions, listed in table (1), where Jupiter and Saturn are in a 3:2 MMR. This family of initial conditions was previously studied in some detail by Morbidelli et al. (2007). Consequently, this section's results are partially reproductions. Using a hydrodynamical model, Morbidelli et al. (2007) found six multi-resonant configurations, two of which they determined to be long-term stable. These are the configurations listed in table (1) where both Jupiter & Saturn and Saturn & Uranus pairs are in 3:2 MMR's while Uranus & Neptune are in either 4:3 or 5:4 MMR's. There are two more compact configurations listed in table (1) which we determined to be stable, although the counterparts of these configurations put together by Morbidelli et al. (2007) were unstable[2]. These configurations are the two where Jupiter & Saturn are in a 3:2 MMR, Saturn & Uranus are in a 4:3 MMR, and Uranus & Neptune are either in a 3:2 MMR or 4:3 MMR.

As discussed in Morbidelli et al. (2007), if Jupiter and Saturn start out in a 3:2 MMR,

[2]Note that our simplified migration model is computationally cheaper than that of Morbidelli et al. (2007), allowing us to run more trials to arrive at stable configurations.

the instability is triggered by their encounter with the 5:3 MMR. While this is a second-order resonance, small jumps in Jupiter's and Saturn's eccentricities go a long way, especially in highly compact configurations. Unfortunately, in this case it is difficult to conclusively determine which configurations will result in evolutions with scattering events a-priori. Thus, we must rely solely on numerical integrations to explore the various evolutionary outcomes of these initial conditions.

After an initial run of 20 integrations for each initial condition of the family listed in table (1), we ruled out the configurations where Saturn & Uranus are in a 2:1 MMR as well as the configuration where all planet pairs are in 3:2 MMR's because all evolutions were characterized by smooth migration. We subjected the remaining four configurations to 30 additional integrations and found that the only configuration which does not result in ice-giant/gas-giant scattering is the one where Saturn & Uranus are in a 3:2 MMR and Uranus & Neptune are in a 4:3 MMR. The evolutions of the remaining initial conditions are presented in figures (1) - (3), and their final orbital parameters are entered into table (4).

For the initial condition in which Saturn and Uranus are in a 3:2 MMR, 10% of the integrations were successful with 57% of them exhibiting close encounters between an ice giant and *both* gas giants. The same fractions for the two configurations where Saturn and Uranus are initially in a 4:3 MMR are 20% & 30% and 27% & 50% for the case where Uranus $ Neptune are in a 3:2 MMR and 4:3 MMR respectively.

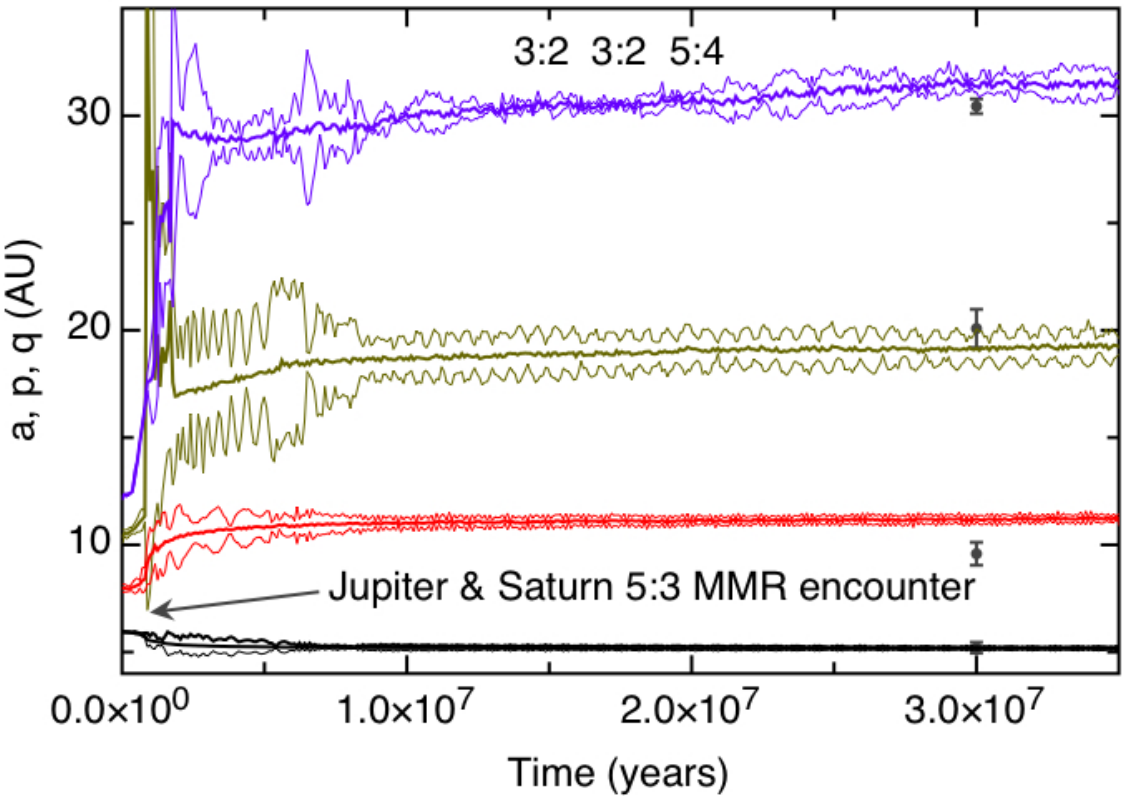


Figure 4.1: Dynamical evolution of the initial configuration where initially Jupiter & Saturn are in a 3:2 MMR, Saturn & Uranus are in a 3:2 MMR and Uranus & Neptune are in a 5:4 MMR (as labeled). Curves depicting the semi-major axes, perihelion, and apohelion of each planet are labeled. The current semi-major axes, perihelia and apohelia of the current solar system are plotted as grey points for comparison. In this model, the global instability is brought forth by Jupiter & Uranus encountering a mutual 5:3 MMR.

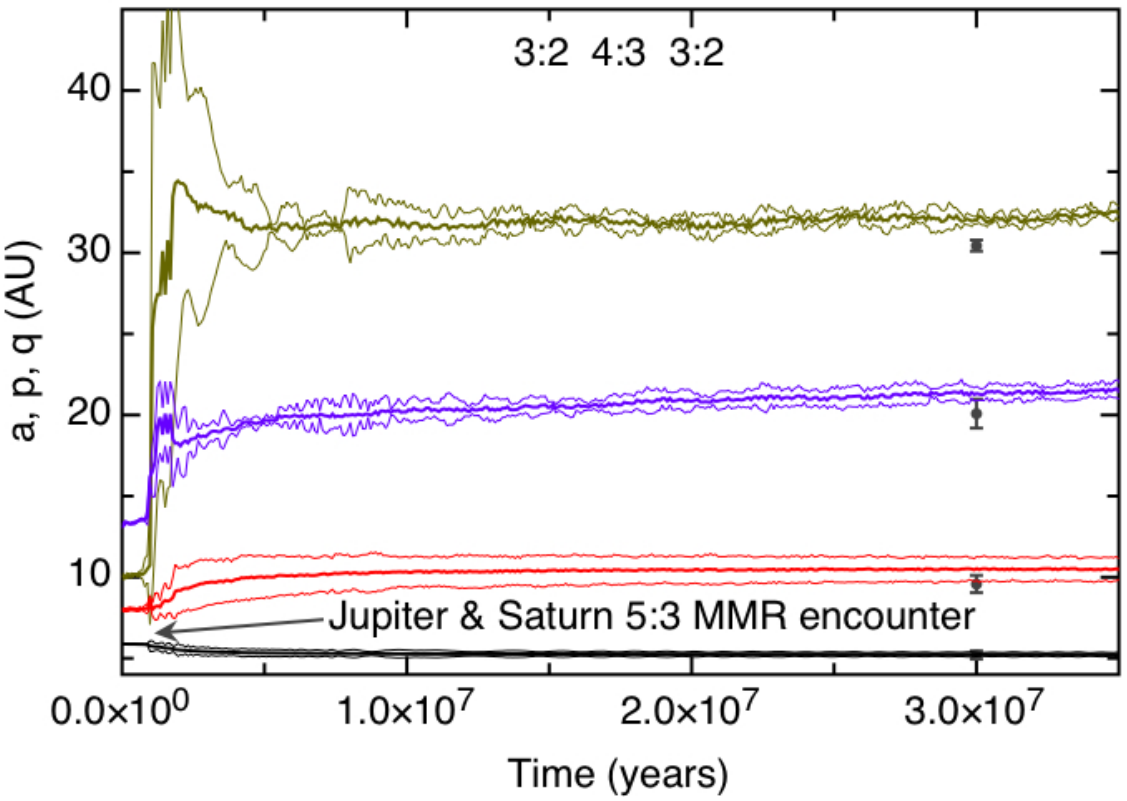


Figure 4.2: Same as fig.1 except initially, Saturn & Uranus are in a 4:3 MMR and Uranus & Neptune are in a 3:2 MMR.

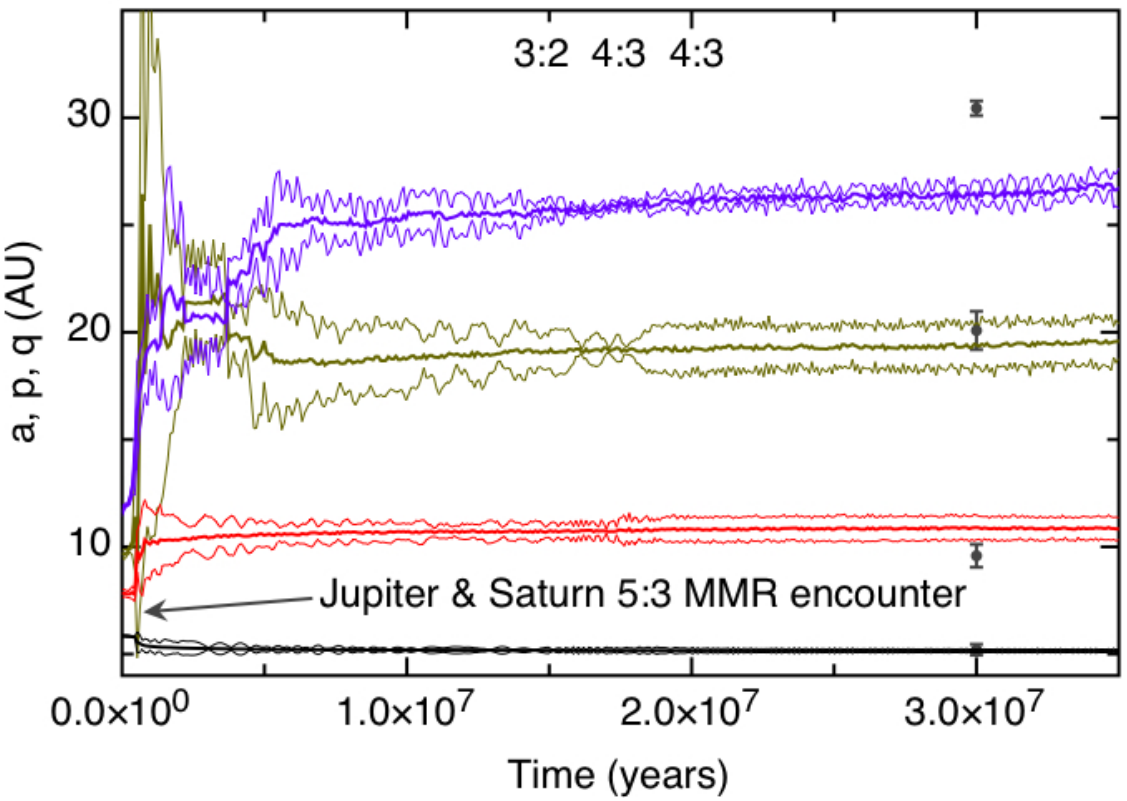


Figure 4.3: Same as fig.1 except initially, both Saturn & Uranus and Uranus & Neptune are in 4:3 MMR's.

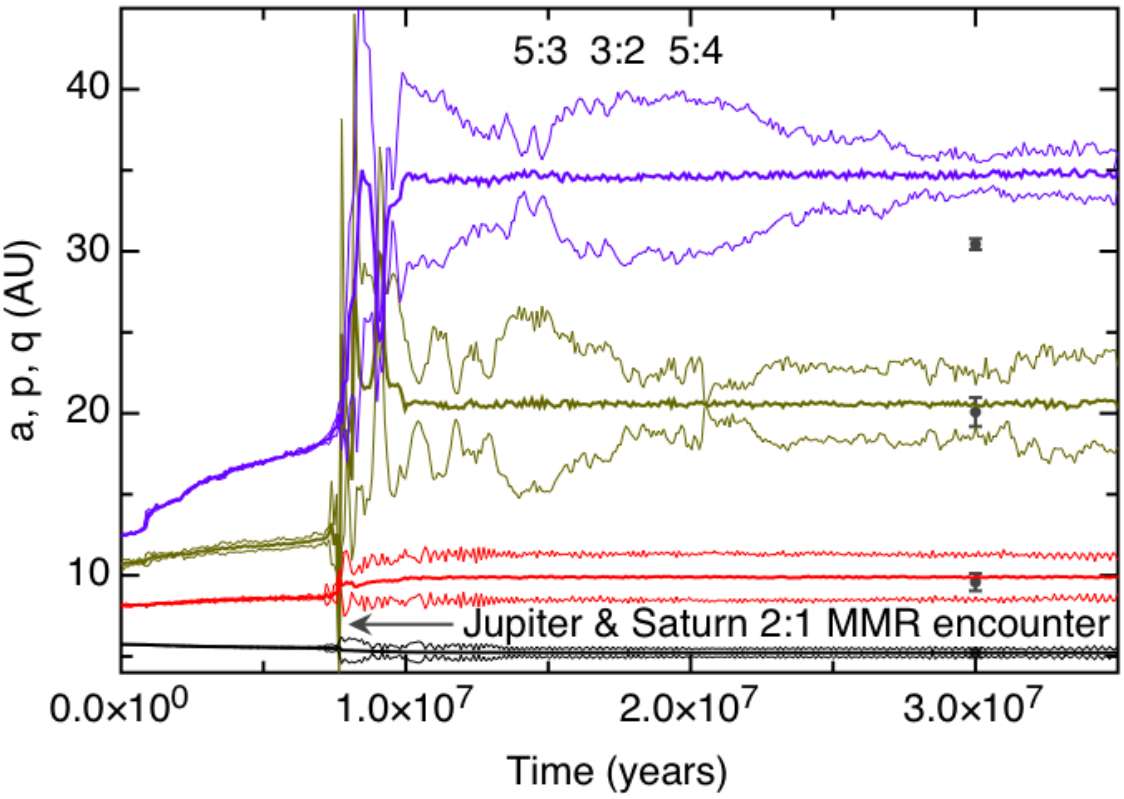


Figure 4.4: Dynamical evolution of the initial configuration where initially Jupiter & Saturn are in a 5:3 MMR, Saturn & Uranus are in a 3:2 MMR and Uranus & Neptune are in a 5:4 MMR. In this model, the global instability is brought forth by Jupiter & Saturn encountering a mutual 2:1 MMR, just as in the classical Nice model. All else is as in fig.1.

4.4.2 Initial Conditions with Jupiter and Saturn in a 5:3 MMR

We now move on to the next family of initial conditions. To begin with, we take the same approach as above. Stable multi-resonant configurations of this family are listed as the second set of entries in table (1). We simulated the evolutions of these systems with 20 integrations each. After completion, a clear boundary between initial conditions that result in smooth migration and those that result in scattering developed. Namely, all setups where Saturn & Uranus are initially in a 2:1 MMR were characterized by smooth evolutions. A similar scenario describes the fate of initial conditions where Saturn & Uranus are in a 3:2 MMR while Uranus & Neptune are in a 3:2 or a 4:3 MMR. However in the same context, if Uranus and Neptune start out in a 5:4 or a 6:5 MMR, ice giant/gas giant scattering as well as transient phases of high eccentricities are present. Particularly, for the configuration where Uranus & Neptune start out in a 5:4 MMR, 20% of the integrations were successful with 50% of them exhibiting close encounters between an ice giant and both gas giants. For the configuration where Uranus & Neptune start out in a 6:5 MMR, also 20% of the integrations were successful, but none of the solutions exhibited ice giant encounters with both gas giants.

In the subset of initial conditions where Saturn & Uranus are in a 4:3 MMR, the configurations with Uranus & Neptune in a 3:2 MMR and a 4:3 MMR can serve as good candidates for solar system formation, but the configuration with Uranus & Neptune in a 5:4 MMR consistently leads to ejections. 10% of the integrations with Uranus & Neptune initially in a 3:2 MMR were successful, all of them exhibiting scattering with both ice giants. In the context of the configuration with Uranus & Neptune initially in a 4:3 MMR,

15% of the integrations were successful, while 33% of them lead to encounters of an ice giant with both gas giants.

Examples of successful evolutions that start from the initial conditions described above are presented in figures (4) - (7), with final orbital parameters entered into table (4). Note that the scattering event in the evolution of the configuration where Saturn & Uranus are initially in a 3:2 MMR and Uranus & Neptune are in a 5:4 MMR (fig. 4) is considerably more violent than that in most other examples. This is because majority of close encounters here is between Jupiter and Uranus, while in most other simulations, Saturn is responsible for scattering. It is important to understand that this is not a unique feature of the particular initial condition. We have observed similar phenomena in simulations of other setups as well.

Note that in a scenario where Jupiter and Saturn start out in a 5:3 MMR, the instability is brought on by their crossing of the 2:1 MMR, just as in the classical Nice model. Much effort has been put into fine-tuning the classical Nice model's initial conditions (Tsiaganis et al. 2005, Morbidelli et al. 2005, Gomes et al. 2005, Levison et al. 2008). What matters most, however, are the locations of the planets when Jupiter and Saturn are crossing the 2:1 MMR. Let us now examine if the classical Nice model is compatible with any multi-resonant initial conditions from this family.

We begin our calculation by measuring the semi-major axes of the four planets at the Jupiter/Saturn 2:1 MMR crossing in figure (2a) of Levison et al. (2008). The values are listed in the second column of Table (2). Between encounters with MMR's, migrations of Jupiter and Saturn are mostly due to scattering of planetesimals. Malhotra (1995) showed

Table 4.2: Analytical Calculation of Planetesimal-Driven Migration

Planet	f	a (AU) (J:S 2:1)	a (AU) (J:S 5:3)	a (AU) (5:4 4:3 4:3)
Jupiter	0.35	5.45	5.69	5.7
Saturn	0.2	8.74	7.97	8.02
Uranus	0.08	12.61	9.84	9.71
Jupiter	0.15	18.36	12.34	11.77

Inputs and results of the analytical calculations of planetesimal-driven migration (section 3.2). f is the effective scattering efficiency, inferred from numerical simulations. Although it varies from run to run, the variation is not too great. The second column, a (AU) (J:S 2:1), lists the positions of the planets at the time of Jupiter/Saturn 2:1 MMR crossing, as inferred from the results of Levison et al. (2008). The next column, a (AU) (J:S 5:3), list the semi-major axes of the planets with Jupiter & Saturn nominally in the 5:3 MMR, traced back using equation (4) from the previous column. The last column, a (AU) (5:4 4:3 4:3), lists the semi-major axes of the 5:4 4:3 4:3 multi-resonant configuration, assembled as discussed in section 2. Note the quantitative similarity between the two right-most columns. This leads us to believe that the 5:4 4:3 4:3 multi-resonant configuration, as well as other compact configurations from the same family are compatible with the classical Nice model.

from conservation of angular momentum that this process obeys a relation, which upon integration can be written as

$$\Delta \log a \simeq f \frac{m_s}{m_{planet}} \tag{4.4}$$

where f is an empirically determined "efficiency," listed in Table (2), and m_s is the total scattered mass. When applied to Jupiter and Saturn simultaneously, this relation can be used to "backtrace" the system's migration, by roughly estimating the starting semi-major axes of Jupiter and the scattered mass m_s. Setting $\Delta a_S = (5/3)^{2/3} a_J^i - (2)^{2/3} a_J^f$ and $\Delta a_J = a_J^f - a_J^i$ with $a_J^f = 5.45$ AU yields a total scattered mass of $37 m_\oplus$ and $a_J^i = 5.69$ AU.

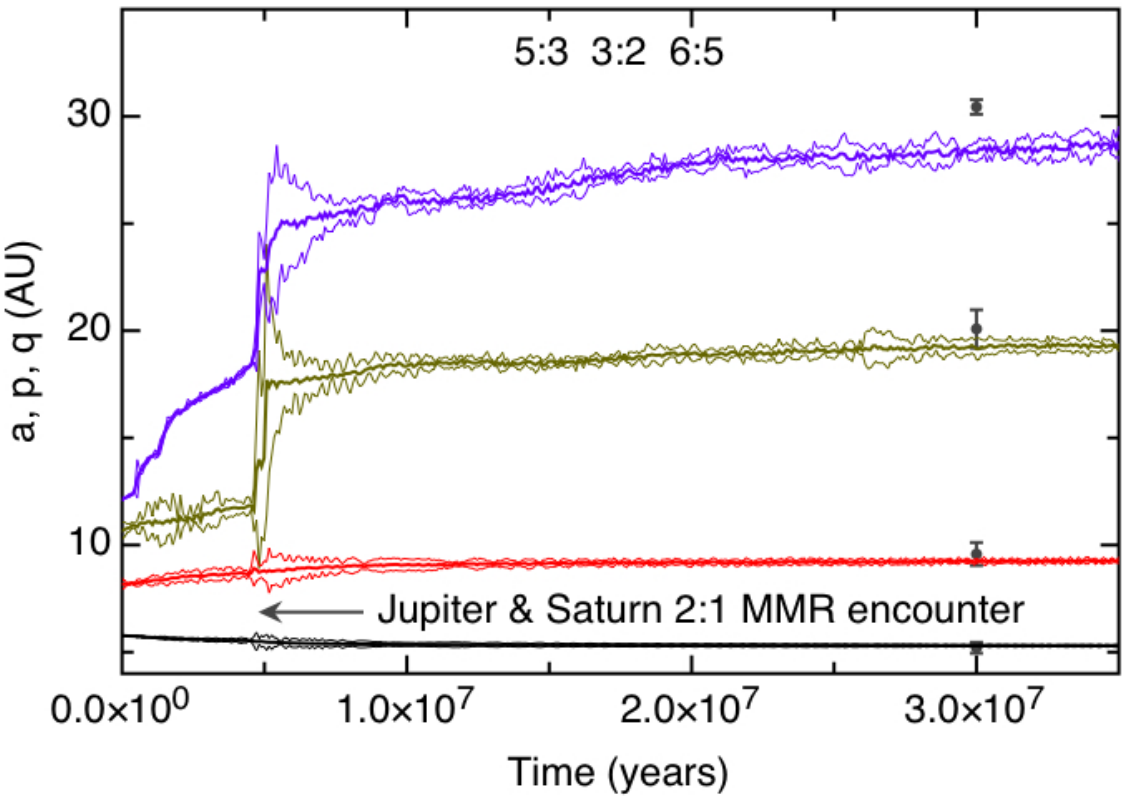


Figure 4.5: Same as fig.4 except initially, Saturn & Uranus are in a 3:2 MMR and Uranus & Neptune are in a 6:5 MMR.

Neglecting high-order resonant encounters, we then use the calculated scattered mass and apply equation (4) to Uranus and Neptune to determine their original positions. The back-traced initial conditions are listed in the third column of table (3). Incidentally, these initial conditions are close to a multi-resonant configuration where Saturn & Uranus and Uranus & Neptune are both in 4:3 MMR's. Recall that this initial condition is indeed one of the setups that consistently exhibit scattering. However, given the similarities in dynamical evolutions among the successful initial conditions of this family, at this level of accuracy, it is probably safe to say that all four of them are compatible with the classical Nice model results.

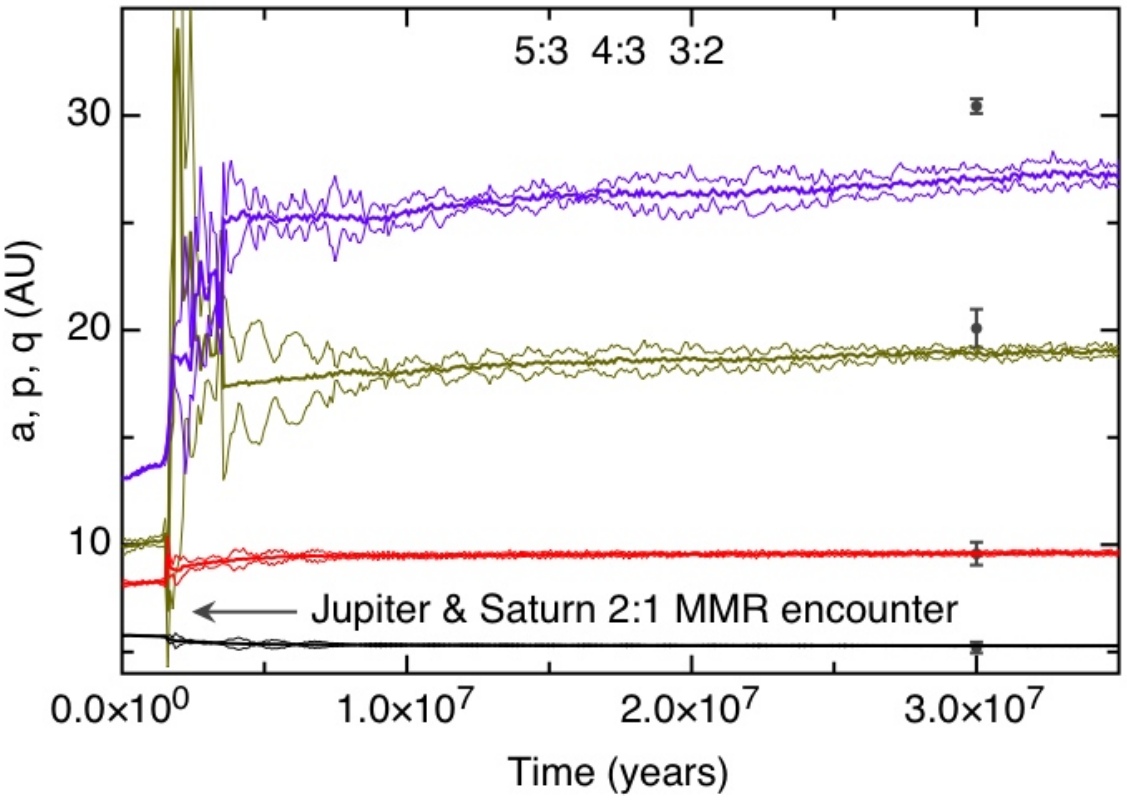


Figure 4.6: Same as fig.4 except initially, Saturn & Uranus are in a 4:3 MMR and Uranus & Neptune are in a 3:2 MMR.

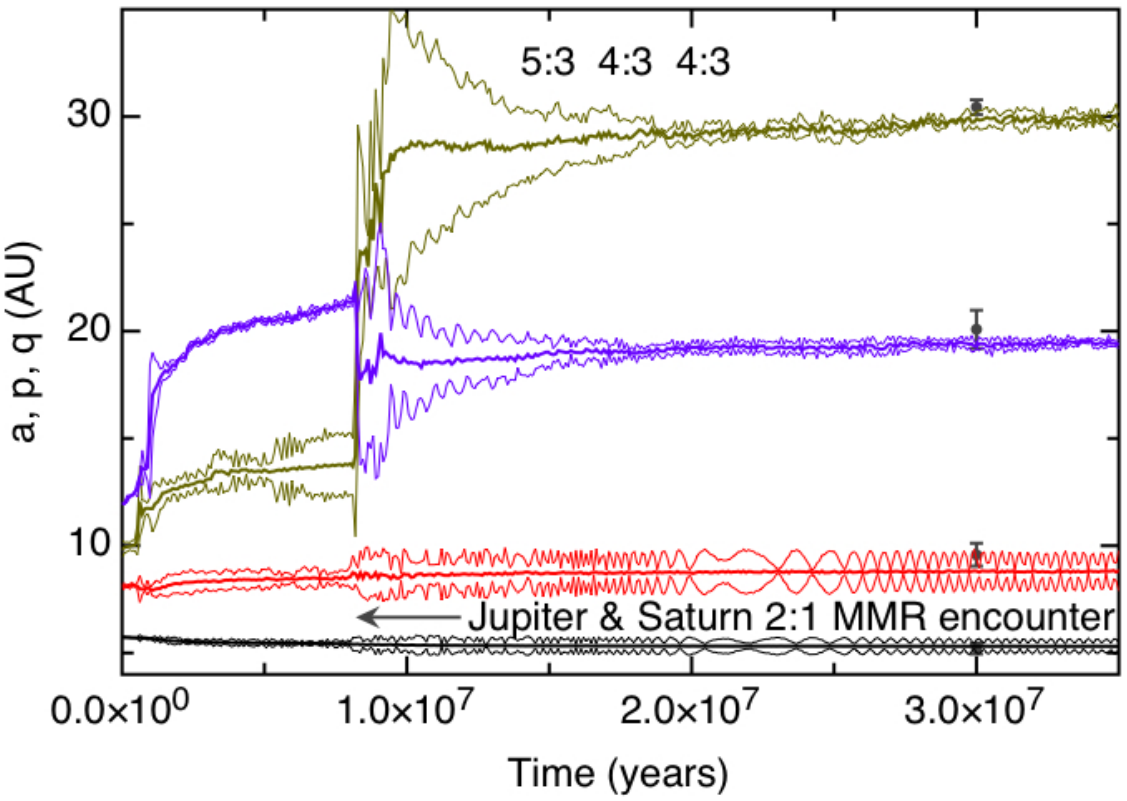


Figure 4.7: Same as fig.4 except initially, both Saturn & Uranus and Uranus & Neptune are in 4:3 MMR's.

4.4.3 Initial Conditions with Jupiter and Saturn in a 2:1 MMR

Let us now consider the final family of initial conditions, listed in table (1), where Jupiter and Saturn are initially in a 2:1 MMR. Unlike the scenario of the classical Nice model (Tsiaganis et al. 2005), there are no major resonances to cross for Jupiter and Saturn between the 2:1 and the 5:2 MMR's. Consequently, a different mechanism, involving different resonances, is needed to create the instability. Thommes et al. (2008) considered the dynamical evolution of a system where Jupiter & Saturn are in a 2:1 MMR, Saturn & Uranus are in a 3:2 MMR, and Uranus & Neptune are in a 4:3 MMR. In such a system, the instability is triggered by Uranus and Neptune crossing a 7:5 MMR. Due to a weaker, second-order nature of this resonance, the eccentricity increase is rather small. Incidentally in this particular system, this is enough for the ice giants to cross orbits and scatter off of each other, but not off of one of the gas giants[3]. It appears that somewhat larger eccentricities are needed. Testing each initial condition with a large number of numerical simulations, as discussed above, is rather time-consuming. Consequently, it is worthwhile to quantify the amplitudes of eccentricity jumps due to various resonance crossings before-hand if possible. For this set of initial conditions, under the assumption of adiabatic migration, the eccentricity jumps are deterministic and can be estimated analytically (Henrard 1982).

Following the treatment of (Peale 1986, see also Murray & Dermott 1999), we consider

[3]Another scenario present in the integrations of Thommes et al. (2008) is the escape of one of the ice giants. In this case, the remaining ice giant is left with a high eccentricity, but there are only 3 planets left in the system.

the planar internal first-order $j:(j-1)$ resonant Hamiltonian

$$\begin{aligned} H_{res} &= -\frac{G^2 m_\odot^2 m^3}{2\Lambda^2} - \frac{G^2 m_\odot^2 m'^3}{2\Lambda'^2} \\ &- \frac{G^2 m_\odot^2 m m'^3}{\Lambda'^2} f(a/a') \sqrt{\frac{2\Gamma}{\Lambda}} \cos(j\lambda' + (1-j)\lambda + \gamma) \\ &- \Gamma\dot{\gamma}_{sec} + \Lambda\dot{\lambda}_{sec} - \Gamma'\dot{\gamma}'_{sec} + \Lambda'\dot{\lambda}'_{sec} \end{aligned} \tag{4.5}$$

where λ is the mean longitude, $\gamma = \varpi$ is the longitude of perihelion, $\Lambda = (m\, m_\odot)/(m + m_\odot)\sqrt{G(m_\odot + m)a}$ & $\Gamma = \Lambda(1 - \sqrt{1-e^2})$ are their respective Poincaré conjugate momenta, and the prime designates the outer planet. The secular changes in mean longitude and longitude of perihelion are accounted for by the last four terms, while $f(a/a')$ arises from the classical expansion of the planetary disturbing potential and is a function of Laplace coefficients and their derivatives. The expressions for $f(a/a')$ are presented in Appendix B of (Murray & Dermott 1999). Under a series of variable transformations (see Peale 1986 for derivation), this Hamiltonian can be rewritten to take a simpler form. Let us introduce the constants α, β and ϵ:

$$\alpha = (j-1)n^* - jn'^* + \dot{\gamma}_{sec}, \tag{4.6}$$

$$\beta = \frac{3}{2}\left[\frac{(j-1)^2}{ma^2} + \frac{j^2}{m'a'^2}\right], \tag{4.7}$$

$$\epsilon = n^{3/2} f(a/a') \frac{a^2}{a'} \frac{m'}{m_\odot} \sqrt{m}, \tag{4.8}$$

where n^* is sum of the Keplerian mean motion and the secular change in mean longitudes. It is important to note that these expressions are not strictly constant, since semi-major axis

changes. However, β is only weakly dependent on semi-major axis (Peale 1976), and in the case of ϵ, variations due to the cosine term dominate, so the assumption of constant coefficients is sound (Murray and Dermott 1999). Relative to the original Hamiltonian, we scale the momentum as:

$$\Phi = \Gamma \left(\frac{2\beta}{\epsilon} \right)^{2/3}. \tag{4.9}$$

The corresponding conjugate angle ϕ is simply the cosine argument in equation (5), although if $\epsilon > 0$, we also need to add π to the expression (Murray & Dermott 1999). That said, the transformed Hamiltonian takes the form

$$H_{res} = \delta\Phi + \Phi^2 - 2\sqrt{2\Phi}\cos(\phi). \tag{4.10}$$

This Hamiltonian is parameterized by

$$\delta = \alpha \left(\frac{4}{\epsilon^2 \beta} \right)^{1/3}, \tag{4.11}$$

which is a measure of the perturbed object's proximity to exact resonance. Finally we note that this Hamiltonian is most easily visualized in terms of polar coordinates, so we introduce the new mixed canonical variables $x = \sqrt{2\Phi}\cos\phi$ and $y = \sqrt{2\Phi}\sin\phi$ (Henrard & Lamaitre 1983). The Hamiltonian now takes the form

$$H_{res} = \frac{\delta(x^2 + y^2)}{2} + \frac{(x^2 + y^2)^2}{4} - 2x. \tag{4.12}$$

Upon application of Hamilton's equations, we see that the stationary points of the above

Hamiltonian are described by the equation

$$x^3 + \delta x - 2 = 0. \tag{4.13}$$

For resonant encounters aided by divergent migration, $\delta < 3$ initially. In this case, the existence of a separatrix is ensured, and there are three real fixed points, all of which lie on the x-axis. Two of these points are always negative, and the more negative one is unstable, as it lies on the intersection of the inner and the outer branches of the separatrix. This is crucial to the estimation of eccentricity jumps during resonant encounters.

If migration is slow enough for δ to be approximately constant over one period of motion, the action, defined as

$$J = \oint \Phi d\phi = \oint x dy \tag{4.14}$$

is an adiabatic invariant (Peale 1986). In other words, it is constant except during separatrix crossing. Furthermore, when the separatrix is far away, the trajectories of the circulating orbits in (x, y) space are circles to a good approximation. Consequently, we can write $J = 2\pi\Phi$ (Murray & Dermott 1999).

When two planets approach commensurability, a wide separatrix is seen as shrinking down on the orbit of the perturbed planet in (x, y) space. When the inner branch of the separatrix engulfs the planetary orbit, the process of resonance crossing is characterized by the planet switching to the separatrix's outer circulating branch. The outer branch has a wider radius, thus the increase in action. However, during this switch, the perturbed planet

must necessarily pass through the unstable stationary point described above. Consequently, the calculation is as follows: knowing the action prior to the resonant encounter, we can determine the value of δ at the transition using equation (13). Recall however, that δ also parameterizes the Hamiltonian, and therefore determines the shape of the separatrix, while the area engulfed by the outer branch corresponds to the new action (see supplemental material of Tsiaganis et al. 2005 for an intuitive discussion). It can be shown that the actions before and after resonance crossing are related by

$$J_i + J_f = -2\pi\delta. \tag{4.15}$$

Thus, the new eccentricity can be easily backed out.

The above analysis can also be applied to external resonances. In this case, γ in the cosine argument of the Hamiltonian (5) is replaced by γ', and its factor $\sqrt{2\Gamma/\Lambda}$ is replaced by $\sqrt{2\Gamma'/\Lambda'}$, since we are now concerned with an e' resonance. Accordingly, we change the scaling factors to

$$\alpha = (j-1)n^* - jn'^* + \dot{\gamma}'_{sec}, \tag{4.16}$$

$$\epsilon = n'^{3/2} f(a/a') a' \frac{m}{m_\odot} \sqrt{m'}, \tag{4.17}$$

while β remains the same. Note also that any indirect terms in the expansion of the disturbing function must be accounted for in $f(a/a')$. Under these transformations, Hamiltonian (10) still applies, and so does the subsequent analysis (Murray and Dermott 1999).

The resulting estimates of eccentricity jumps for various first-order resonant encounters between Uranus & Neptune and Saturn & Uranus are listed in Table (3). As can be

Table 4.3: Analytical Estimates of Eccentricities After a Resonant Encounter

Resonance	e_N	e_U
3:2	0.031	0.037
4:3	0.027	0.031
5:4	0.024	0.028
Resonance	e_U	e_S
2:1	0.062	0.022
3:2	0.098	0.018
4:3	0.084	0.016
5:4	0.075	0.015

Results of analytical estimations of eccentricity jumps during first-order resonant encounters. The top three lines represent resonant encounters between Uranus & Neptune, while the bottom four lines are that between Saturn & Uranus. All planets were taken to have initial eccentricities of $e_{init} = 0.01$.

seen from these calculations, all first-order resonant encounters between Uranus and Neptune produce rather small eccentricity jumps. Therefore, we disfavor them as good options for triggering instability scenarios in which encounters with Saturn take place. Numerical integrations performed in the two previous sections are suggestive of this as well. We therefore rule out the configurations where Saturn and Uranus are in a 2:1 MMR. Resonant encounters between Saturn and Uranus, however, are a different story: in all cases, Uranus acquires an eccentricity comparable to 0.1. Simulations reveal that the configurations where Saturn and Uranus are in a 3:2 MMR do not result in strong instabilities. This is because the system is given a chance to encounter high-order MMR's between Uranus and Neptune and spread out before crossing the 2:1 MMR between Saturn & Uranus. Furthermore, in order to increase the chances of Uranus/Saturn orbital crossing, it helps to start the two planets in the most compact stable resonance - namely the 4:3. From here, the degeneracy lies in whether Uranus and Neptune start out in a 3:2 or 4:3 MMR. Out of 30

numerical simulations performed for each configuration, we only observed ice-giant/gas-giant scattering events in the evolutions of the system where Uranus and Neptune are in the 4:3 MMR. Particularly, 23% were successful, and in 57% of the successful integrations, an ice giant exhibited encounters with both gas giants. Figure (8) shows the dynamical evolution of this configuration with time, while table (4) lists the final values of semi-major axes, eccentricities, and inclinations for the planets.

Aside from Saturn's close encounter with an ice-giant, Jupiter's and Saturn's migration is dominated by scattering of planetesimals. As a result, equation (4) approximates their evolution well. Similarly to the previous section, when applied to Jupiter and Saturn simultaneously, this relation can be used to roughly estimate the starting semi-major axis of Jupiter and the initial m_{disk}. Setting $\Delta a_S = (2)^{2/3} a_J^i - (5/2)^{2/3} a_J^f$ and $\Delta a_J = a_J^f - a_J^i$ with $a_J^f = 5.2$ AU yields a total scattered mass of $49 m_\oplus$ and $a_J^i = 5.5$ AU. These values are in good agreement with numerical integrations. For instance, the evolution presented in figure (1) started had the parameters: $a_J^i = 5.5$ AU and $m_{disk} = 51 m_\oplus$.

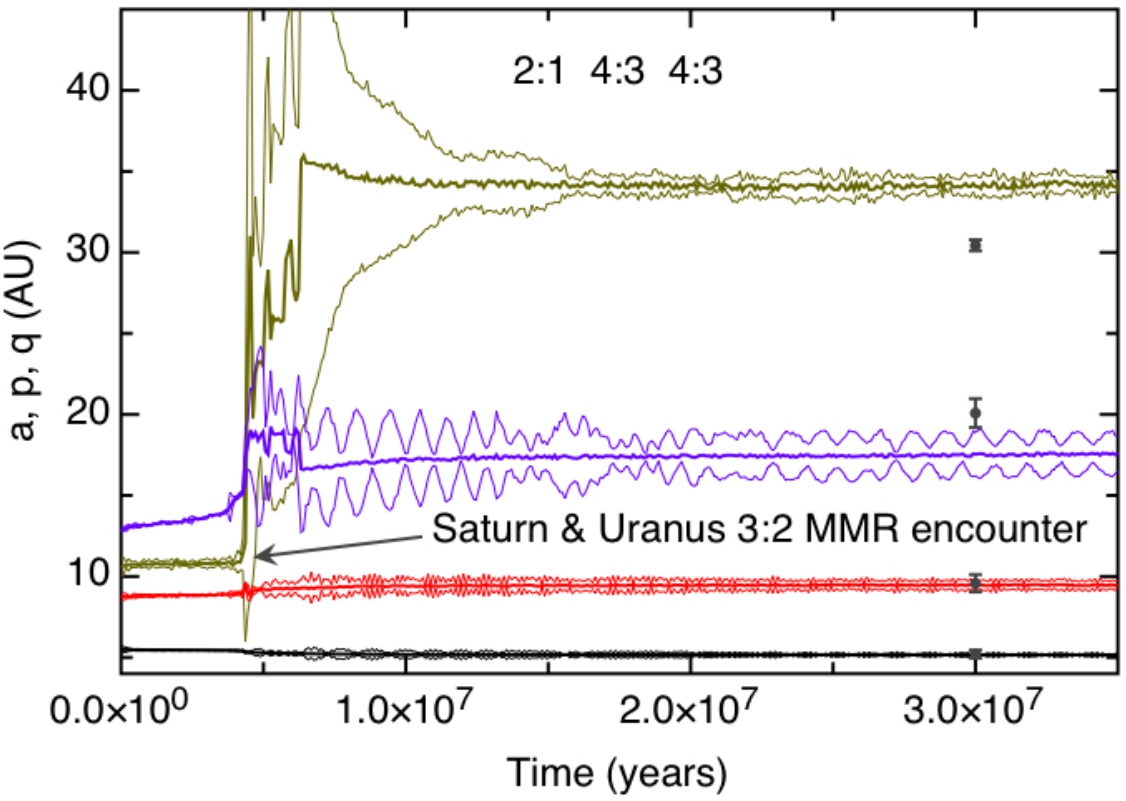


Figure 4.8: Dynamical evolution of the initial configuration where initially Jupiter & Saturn are in a 2:1 MMR, while both Saturn & Uranus and Uranus & Neptune are in a 4:3 MMR's. In this model, the global instability is brought forth by Saturn & Uranus encountering a mutual 3:2 MMR. All else is as in fig.1.

4.5 Discussion

The calculations presented in this work aim to place constraints on the early dynamical state of the solar system. We began by constructing a large array of multi-resonant systems. We then considered each one of these systems individually and tested them against two inter-related constraints: scattering of an ice-giant from a gas-giant and the outer ice giant undergoing a transient phase of high eccentricity. Both events are needed to reproduce the current dynamical architecture of the solar system.

We showed numerically that three configurations in which Jupiter & Saturn start out in a 3:2 MMR follow evolutionary tracks which are compatible with our constraints. The three systems are those in which (1) Saturn & Uranus are in a 3:2 MMR while Uranus & Neptune are in a 5:4 MMR, (2) Saturn & Uranus are in a 4:3 MMR while Uranus & Neptune are in a 3:2 MMR and (3) Saturn & Uranus are in a 4:3 MMR while Uranus & Neptune are also in a 4:3 MMR. Two of these configurations were previously thought to be unstable (Morbidelli et al. 2007).

We showed that the classic Nice model can be reproduced from multi-resonant initial conditions where Juipter & Saturn start out in a 5:3 MMR. The four particular primordial states that we found were those where initially, (1) Saturn & Uranus are in a 3:2 MMR while Uranus & Neptune are in a 5:4 MMR, (2) Saturn & Uranus are in a 3:2 MMR while Uranus & Neptune are in a 6:5 MMR, (3) Saturn & Uranus are in a 4:3 MMR while Uranus & Neptune are in a 3:2 MMR, and (4) Saturn & Uranus are initially in a 4:3 MMR while Uranus & Neptune are also in a 4:3 MMR.

Finally, we used an analytical technique to rule out a large portion of the generated

multi-resonant systems, in which Jupiter and Saturn are initially in the 2:1 MMR, based on an argument that the eccentricities generated by any Uranus/Neptune resonant encounters are too small. Simultaneously, we showed that the considered constraints can be satisfied by the dynamical evolution whose initial condition has Jupiter & Saturn locked in a 2:1 MMR, and the other pairs of the planets in 4:3 MMR's.

The calculations presented here are intended in part as a point of departure for future research. There is indeed a large array of unexplored issues. For instance, in the case of the system where Jupiter & Saturn start out in a 2:1 MMR, it is not clear if a single scattering event alone is enough to correctly reproduce the secular dynamics of Jupiter and Saturn, or if a resonant encounter is also required. In the initial condition where Jupiter and Saturn start out in a 5:3 MMR, dynamical stability may pose an issue (Morbidelli & Crida 2007). A further criterion of interest is LHB. While we do not attempt to time the onset of instability with LHB in these simulations, it is certainly fair to ask if the dynamical evolutions presented here are compatible with a long quiescent period preceding any resonant encounters. Two related issues immediately follow. First, will changing the placement of the inner boundary of the planetesimal swarm qualitatively change the process responsible for the onset of the instability of the dynamical evolutions? Second, how will the effective viscosity that arises from the self-gravity of the disk affect our results? A large-scale numerical modeling effort will be instrumental in providing these answers.

There is certainly room for broader study of the current setup of the problem as well. In this work, we have restricted ourselves to four-planet multi-resonant configurations. Certainly, the idea of initially forming more than two ice giants is not unreasonable (Ford &

Chiang 2007, see however Levison & Morbidelli 2007). Although the results of Morbidelli et al. (2007) suggests that additional planets in a compact multi-resonant system compromise dynamical stability, more work is needed to obtain a good handle on this part of the problem.

In conclusion, the determination of a small subset of initial conditions allows for a much more efficient survey of the parameter space. In this work, we have taken a step in this direction. We must keep in mind that the system at hand is highly chaotic, and must in the end be studied numerically. The resulting determinations are often probabilistic rather than conclusive, however the results are certainly bound to gain statistical weight as the number of completed simulations increases. Thus, while much progress is yet to be made, additional research carries great value since a solid understanding of initial conditions plays an unavoidably important role in further development of a comprehensive model for solar system's formation.

Acknowledgments We thank Hal Levison, Alessandro Morbidelli, Ramon Brasser, Gregory Laughlin and Darin Ragozzine for useful discussions.

Table 4.4: Orbital Elements at the End of Simulations

			a (AU)	e	i (deg.)
3:2 3:2 5:4 configuration	$k = 1.93$	$m_{disk} = 91M_{\oplus}$			
Jupiter			5.2	0.013	0.17
Saturn			11.2	0.025	0.18
Uranus			19.2	0.017	0.9
Neptune			31.5	0.018	1.3
3:2 4:3 3:2 configuration†	$k = 1.0$	$m_{disk} = 82M_{\oplus}$			
Jupiter			5.2	0.027	0.31
Saturn			10.5	0.068	0.5
Uranus			21.5	0.022	0.9
Neptune			32.5	0.011	0.85
3:2 4:3 4:3 configuration†	$k = 1.41$	$m_{disk} = 75M_{\oplus}$			
Jupiter			5.15	0.018	0.55
Saturn			10.8	0.05	1.15
Uranus			19.6	0.036	1.63
Neptune			26.7	0.043	2.95
5:3 3:2 5:4 configuration	$k = 1.50$	$m_{disk} = 60M_{\oplus}$			
Jupiter			5.22	0.073	0.37
Saturn			9.9	0.109	1.24
Uranus			20.39	0.122	2.66
Neptune			34.89	0.034	0.65
5:3 3:2 6:5 configuration	$k = 1.35$	$m_{disk} = 63M_{\oplus}$			
Jupiter			5.3	0.011	0.07
Saturn			9.28	0.016	0.27
Uranus			19.23	0.008	0.08
Neptune			28.51	0.022	0.57
5:3 4:3 3:2 configuration	$k = 1.85$	$m_{disk} = 64M_{\oplus}$			
Jupiter			5.29	0.004	0.49
Saturn			9.64	0.013	1.5
Uranus			18.99	0.016	0.69
Neptune			27.38	0.022	0.22
5:3 4:3 4:3 configuration†	$k = 1.75$	$m_{disk} = 58M_{\oplus}$			
Jupiter			5.3	0.02	0.25
Saturn			8.8	0.09	0.14
Neptune			19.7	0.01	0.65
Neptune			30.4	0.007	1.86
2:1 4:3 4:3 configuration†	$k = 1.9$	$m_{disk} = 51M_{\oplus}$			
Jupiter			5.16	0.016	0.08
Saturn			9.48	0.029	0.13
Uranus			17.57	0.06	0.76
Neptune			34.34	0.004	0.6

Orbital elements of solar system analogues, resulting from different initial conditions, at the end of the dynamical evolution simulations, presented in figures 1 - 8. Simulations where the ice giants switched places are labeled with a †. The disk mass used in each simulation, as well as the disk's power law index k are also given.

Bibliography

Henrard, J. 1982, Celestial Mechanics, 27, 3

Henrard, J., & Lamaitre, A. 1983, Celestial Mechanics, 30, 197

Tsiganis, K., Gomes, R., Morbidelli, A., & Levison, H. F. 2005, Nature, 435, 459

Gomes, R., Levison, H. F., Tsiganis, K., & Morbidelli, A. 2005, Nature, 435, 466

Morbidelli, A., Levison, H. F., Tsiganis, K., & Gomes, R. 2005, Nature, 435, 462

Malhotra, R. 1995, The Astronomical Journal, 110, 420

Fernandez, J. A., & Ip, W.-H. 1984, Icarus, 58, 109

Nesvorný, D., Vokrouhlický, D., & Morbidelli, A. 2007, The Astronomical Journal, 133, 1962

Brasser, R., Morbidelli, A., Gomes, R., Tsiganis, K., & Levison, H. F. 2009, arXiv:0909.1891

Morbidelli, A., Brasser, R., Tsiganis, K., Gomes, R., & Levison, H. 2009, AAS/Division for Planetary Sciences Meeting Abstracts, 41, #55.03

Masset, F., & Snellgrove, M. 2001, Monthly Notices of Royal Astronomical Society, 320, L55

Peale, S. J. 1976, Annual Reviews of Astronomy & Astrophysics, 14, 215

Peale, S. J. 1986, Satellites, 159

Morbidelli, A., & Crida, A. 2007, Icarus, 191, 158

Lee, M. H., & Peale, S. J. 2002, The Astrophysical Journal, 567, 596

Pierens, A., & Nelson, R. P. 2008, Astronomy & Astrophysics, 482, 333

Morbidelli, A., Tsiganis, K., Crida, A., Levison, H. F., & Gomes, R. 2007, The Astronomical Journal, 134, 1790

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Cambridge: University Press, —c1992, 2nd ed.,

Chambers, J. E. 1999, Monthly Notices of Royal Astronomical Society, 304, 793

Morbidelli et al 2009 arXiv:0909.1886

Levison, H. F., Morbidelli, A., Vanlaerhoven, C., Gomes, R., & Tsiganis, K. 2008, Icarus, 196, 258

Stewart, G. R., & Wetherill, G. W. 1988, Icarus, 74, 542

Thommes, E. W., Bryden, G., Wu, Y., & Rasio, F. A. 2008, The Astrophysical Journal, 675, 1538

Levison (2009), personal communication

Murray, C. D., & Dermott, S. F. 1999, Solar System Dynamics 1999, Cambridge University Press

Ford, E. B., & Chiang, E. I. 2007, The Astrophysical Journal, 661, 602

Levison, H. F., & Morbidelli, A. 2007, Icarus, 189, 196

Chapter 5

Retention of a Primordial Cold Classical Kuiper Belt in an Instability-Driven Model of Solar System Formation

Originally published as:
Batygin, K., Brown, M. E., & Fraser, W. C. 2011, The Astrophysical Journal, 738, 13

5.1 Abstract

The cold classical population of the Kuiper belt exhibits a wide variety of unique physical characteristics, which collectively suggest that its dynamical coherence has been maintained through out the solar system's lifetime. Simultaneously, the retention of the cold population's relatively unexcited orbital state has remained a mystery, especially in the context of a solar system formation model, that is driven by a transient period of instability, where Neptune is temporarily eccentric. Here, we show that the cold belt can survive the instability, and its dynamical structure can be reproduced. We develop a simple analytical model for secular excitation of cold KBOs and show that comparatively fast apsidal precession and nodal recession of Neptune, during the eccentric phase, are essential for preservation of an unexcited state in the cold classical region. Subsequently, we confirm

our results with self-consistent N-body simulations. We further show that contamination of the hot classical and scattered populations by objects of similar nature to that of cold classicals has been instrumental in shaping the vast physical diversity inherent to the Kuiper belt.

5.2 Introduction

The quest to understand the origins of the solar system dates back centuries. The last two decades, however, have seen a renewed interested in the problem, as the discovery of the Kuiper belt (Jewitt & Luu, 1993) has provided important new clues to the physical processes that took place during the early stages of our solar system's evolution. The continued acquisition of new information gave rise to a multitude of new formation models (see Morbidelli et al. (2008) for a comprehensive review). Among the newly proposed scenarios, an instability model, termed the "Nice" model (Tsiganis et al., 2005; Gomes et al., 2005; Morbidelli et al., 2005), has been particularly successful in reproducing the observed properties of planetary orbits and the Kuiper belt (Levison et al., 2008).

Within the context of the narrative told by the Nice model, planets start out in a multi-resonant configuration (Morbidelli et al., 2007; Batygin & Brown, 2010), and, driven by planetesimal scattering, begin migrating divergently (Fernandez & Ip, 1984). Eventually, the planets encounter a low-order mean motion resonance, which results in a transient period of instability. During this period, the ice giants scatter outwards and settle roughly onto their current semi-major axes but with high eccentricities (Tsiganis et al., 2005; Thommes et al., 2008). Neptune's excited eccentricity gives rise to a chaotic sea between its exterior

3:2 and 2:1 mean motion resonances (MMRs), allowing planetesimals to random-walk into the "classical" region (Levison et al., 2008). Subsequently, as the planets circularize due to dynamical friction (Stewart & Wetherill, 1988), the scattered and resonant populations of the Kuiper belt are sculpted.

An outstanding problem within the Nice model lies in the formation of the cold classical population of the Kuiper belt, which is the central theme of this study. The cold population is distinctive from the rest of the Kuiper belt in a number of ways. First and foremost, as its name suggests, the orbital distribution is dynamically unexcited. When Neptune scatters planetesimals, it tends to pump up their inclinations to tens of degrees. Yet the cold population resides on nearly co-planar orbits, with inclinations not exceeding $\sim 5^\circ$ (Brown, 2001; Gladman et al., 2008). The eccentricities of the cold population, on average, also tend to be diminished in comparison with the hot population, but the division there is not as apparent. Figure 1 shows the eccentricities of the current aggregate of observed Kuiper Belt objects (KBOs) between 30 and 60AU. Cold classical objects, whose inclinations are below 5° are plotted as black dots, while all other objects with inclinations above 5° are plotted as blue dots. Note that cold population's eccentricity distribution is not monotonic in semi-major axes. Between 42AU and 45AU, planar KBOs have roughly isentropic eccentricities. However, low-eccentricity objects progressively disappear beyond 45AU. We refer to this feature of the Kuiper belt as the "wedge" (see Figure 1).

A second distinction is the colors of cold classical KBOs. In general, the Kuiper belt exhibits a vast diversity of colors, from neutral gray to deep red. Within this range, cold classical KBOs readily stand out as clump of exclusively red material (Trujillo & Brown,

2002; Lykawka & Mukai, 2005). In a similar manner, the size distribution of the cold population differs significantly from that of the hot classical population (Fraser et al., 2010). Finally, the fraction of binaries present in the cold population is uniquely large (Stephens & Noll, 2006). Moreover, it has been shown that the wide binaries of the cold population in particular, would have been disrupted by encounters with Neptune (Parker & Kavelaars, 2010), and thus must have never been scattered. While it is difficult to interpret each of these observational facts as conclusive evidence for a particular history, their coherence suggests that the cold classicals are a unique population whose dynamical similarity has been maintained through the dramatic evolution of the outer solar system (Morbidelli & Brown, 2004).

A number of formation mechanisms for the cold classical population have been suggested. Within the context of a smooth migration scenario (Malhotra, 1995; Murray-Clay & Chiang, 2005; Hahn & Malhotra, 2005), a primordially cold population can in principle escape dynamical excitation. However, other drawbacks of the smooth migration scenario, such as the inability to reproduce secular architecture of the planets and difficulties in forming the hot classical belt, render it unlikely (Morbidelli et al., 2009; Brasser et al., 2009). Levison et al. (2008) advocated a similar emplacement history for the cold classicals as the hot classicals (i.e. via MMR overlap). The cold population that is produced in such simulations however, is not cold enough and not physically distinct from the hot population. Subsequently, Morbidelli et al. (2008) showed that if a local, cold population is implemented into the orbital solution of Levison et al. (2008), it will have the same orbital distribution as the implanted population after the instability, so the problem remains. Thus,

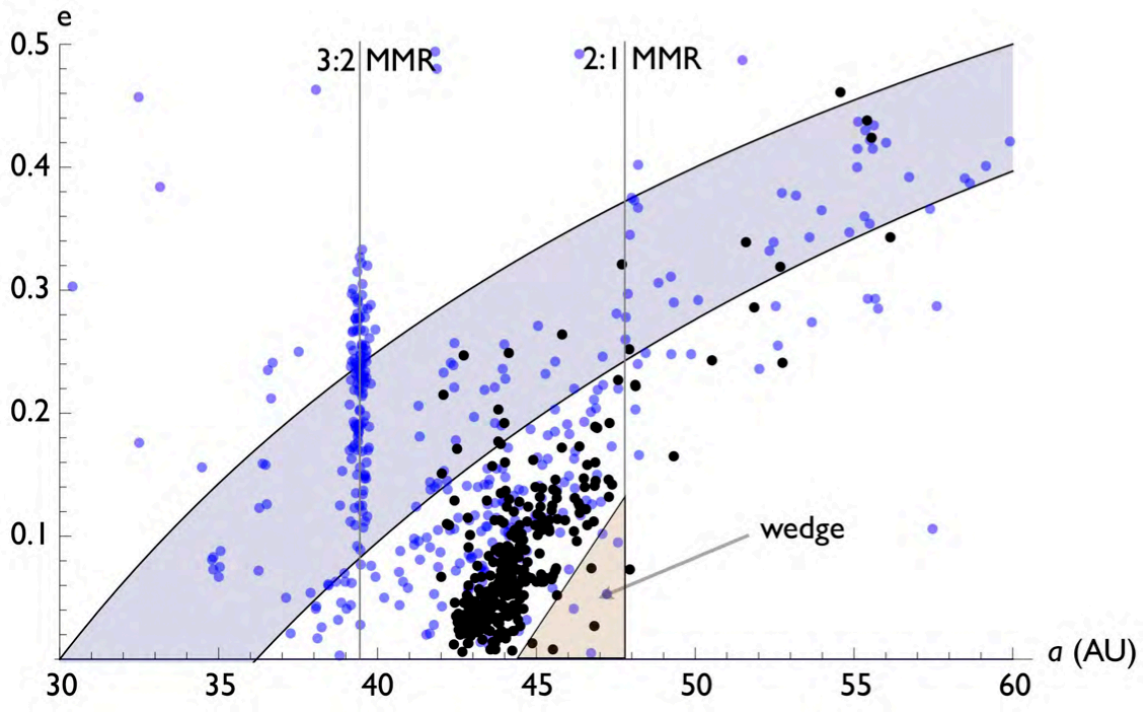


Figure 5.1: Semi-major axis vs. eccentricity of he observed Kuiper belt. The black points denote objects with inclinations below $i < 5^\circ$ i.e. the cold classical population. The blue points represent all other objects with $i > 5^\circ$. The filled curves represent the scattered disk region and the major mean-motion resonances are shown as solid lines. The triangle, adjacent to the 2:1 MMR depicts the wedge structure, inherent to the cold classical population.

no coherent picture of the formation of the cold population exists.

In this work, we show that in-situ formation of the cold population is consistent with an instability model and all observed dynamical properties of the population, including the low inclinations and the wedge (shown in Figure 1) can be formed. The aim here is not to replicate the Kuiper belt and the orbits of the planets in a complex N-body simulation, but rather to identify the dynamical processes responsible for the sculpting of the region. The plan of our paper is as follows: in section 2, we construct an analytical model for secular excitation of a primordially unexcited belt, and thus derive the conditions for retention of dynamically cold orbits. Moreover, we show that the now-fossil wedge is a result of

a temporary slow-down in orbital precession. In section 3, we perform self-consistent N-body simulations that confirm our analytical results. Numerical simulations show that while the cold population can remain undisrupted, similar objects immediately interior to the 3:2 MMR get scattered all over the Kuiper belt. We conclude and discuss our results in section 4.

5.3 Secular Excitation of the Cold Kuiper-Belt

Here, we seek to develop a simplified analytical model that describes the long-term interactions between Neptune and an initially dynamically cold population of KBOs, residing between its exterior 3:2 and 2:1 MMRs, during a transient phase of high eccentricity. Prior to the instability, the planets sit in a compact configuration on near-circular orbits. As long as the orbital separation between the planets and KBOs remains large, their mutual interactions are extremely weak, so the KBOs maintain their dynamically cold orbits. Consequently this period is unimportant to the problem at hand.

When planet-planet resonance crossing (or some other mechanism) causes the instability, the gain in semi-major axes and acquisition of high eccentricities and inclination of the planets takes place on a time scale that is considerably shorter than Neptune's apsidal precession period (i.e. less than a million years or so). As a result, it can be viewed as instantaneous within the context of a secular approximation. Thus, in an orbit-averaged sense, it is as if Neptune suddenly appears at 30 AU with a high e and i and begins interacting with the KBOs. Since we seek to show that after the transient phase of high eccentricity, the KBOs can end up on dynamically cold orbits, we must restrict Neptune

from penetrating the region beyond 40 AU. This places the maximum eccentricity, attainable by Neptune, below $e_{max} < (4/3 - 1) = 1/3$. This is however a weak constraint, since an eccentric, inclined Neptune can still cause large modulations in the eccentricities and inclinations of the KBOs on a secular time-scale (Murray and Dermott, 1999). Let us now develop a mathematical model for these secular interactions.

We begin by modeling Neptune's evolution. In our model, we take the mass of the cold KBOs to be negligible, so they have no effect on Neptune's orbit (this is not necessarily true, at all times, for other Kuiper belt populations). The lack of mass in the primordial cold belt is a requirement for our model that brings up concerns about its formation. We shall discuss this in some detail in section 4. Since we seek to retain the majority of the local population, and we know that the mass of the current cold classical population is much less than that of the Earth, this is a reasonable assumption. The other planets, as well as the massive component of the Kuiper belt, will cause apsidal and nodal precession of Neptune's orbit, which we write as $g = \langle \dot{\varpi}_N \rangle$ and $f = \left\langle \dot{\Omega}_N \right\rangle$ respectively. Note that we are only accounting for the average precessions. We express dynamical friction as exponential decay of e & i with constant time-scales τ_e and τ_i. These time-scales are different, and their numerical values in N-body simulations tend to be of order $\sim 10^7$ years (Levison et al., 2008). We neglect the modulation of Neptune's e & i by the other planets. In other words, we only retain the free elements.

In terms of complex Poincarè variables ($x = e \exp(\imath\varpi), y = i \exp(\imath\Omega)$), we can formu-

late the first-order Lagrange's equations for Neptune as follows:

$$\frac{dx_n}{dt} = \imath g x_n - \frac{x_n}{\tau_e} \qquad \frac{dy_n}{dt} = \imath f y_n - \frac{y_n}{\tau_i}. \tag{5.1}$$

It is trivial to show that these equations admit the solutions

$$x_n = e_n^0 \exp[(\imath g - 1/\tau_e)t] \qquad y_n = i_n^0 \exp[(\imath f - 1/\tau_i)t], \tag{5.2}$$

where e_n^0 and i_n^0 are the initial (maximum) eccentricity and inclination of Neptune respectively and $\imath = \sqrt{-1}$. Here, we take $e_n^0 = 0.25$ and $i_n^0 = 10^\circ$, in accord with results of numerical simulations (Tsiganis et al., 2005; Levison et al., 2008; Batygin & Brown, 2010). In our simple model, the secular evolution of the KBOs is dictated entirely by Neptune's evolution. In the spirit of Laplace-Lagrange secular theory, we only retain terms up to second order in eccentricity and inclination in the disturbing function of the KBOs to ensure a decoupled, analytical solution. The resulting first-order Lagrange's equations read (Wu and Goldreich, 2002)

$$\frac{dx_{kbo}}{dt} = \imath A x_{kbo} + \imath A_n x_n \qquad \frac{dy_{kbo}}{dt} = \imath B y_{kbo} + \imath B_n y_n, \tag{5.3}$$

where A, A_n, B, and B_n are constants that depend only on the planetary masses and semi-major axes ratios of Neptune to KBOs (e.g. Ch.7 of Murray and Dermott (1999)). Note

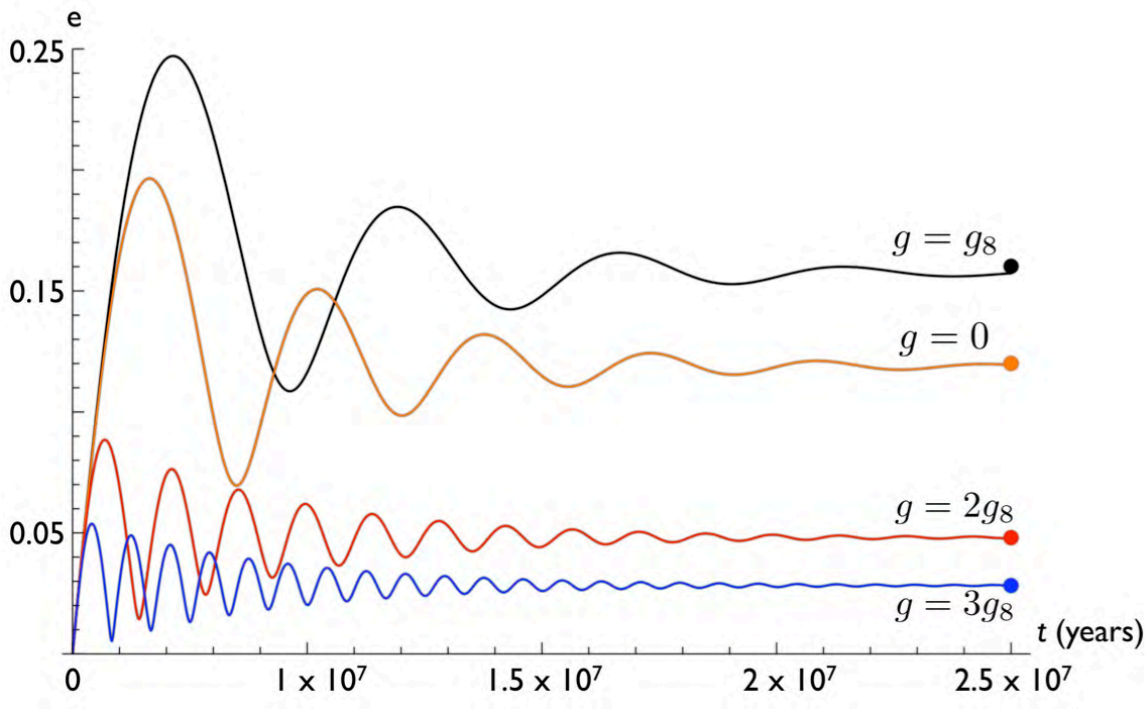


Figure 5.2: Secular excitation of a KBO at $a = 45$AU, as dictated by equation (4). In these solutions, we chose $a_N = 30$AU, $e_n^0 = 0.25$, and $\tau_e = 4$Myr. The final eccentricities (given by equation 8) are plotted as dots. Note that a low final eccentricity requires a comparatively fast precession.

that in the free precession terms, (A,B), the presence of other planets can also be accounted for with ease.

From here, let us focus only on the eccentricity evolution, since the derivation of the inclination evolution follows an identical procedure. Setting the initial orbital state vector the KBO to zero ($[x, y] = \vec{0}$), the solution to the above equation reads

$$x_{kbo} = \frac{e_n^0 \tau_e A_n (\exp[\imath A t] - \exp[(\imath g - 1/\tau_e)t])}{A\tau_e - g\tau_e - \imath}. \tag{5.4}$$

The controlling parameter in this solution is Neptune's precession, g. Four solutions, for a KBO at $a = 45$AU, with different g's are presented in Figure 2. A natural unit of g is

the g_8 eigenfrequency of the Laplace-Lagrange secular solution for the solar system, which physically corresponds to Neptune's average precession rate in the current solar system (see Murray and Dermott (1999)). Incidentally, the same unit can be used for the nodal recession rate in the inclination solution, since quantitatively $g_8 \approx -f_8 \approx 0.65"/yr$. As can be seen in Figure 2, varying g leads to dramatically different results. In particular, if low-eccentricities are to be retained, g must significantly exceed g_8.

After a sufficient amount of time, when Neptune's eccentricity has decayed away (i.e. $t \gg \tau_e$), the second exponential in the numerator of equation (4) can be neglected. Such a solution represents a precessing KBO with a constant eccentricity. Accordingly, the time dependence of the solution only governs its angular part. Since we are solely interested in the final orbits of the KBOs, we must extract only the radial part of the solution. Let us write the $t \gg \tau_e$ solution as an exponential of an arbitrary number, ξ:

$$\exp(\xi) = \frac{e_n^0 \tau_e A_n \exp[\imath A t]}{A\tau - g\tau - \imath}. \tag{5.5}$$

Solving for ξ, and complex-expanding the logarithm, we have

$$\begin{aligned} \xi &= \ln\left(\frac{e_n^0 \tau_e A_n}{\sqrt{1+\tau_e^2(g-A)^2}}\right) \\ &+ \imath \arg\left(-\frac{e_n^0 \tau_e A_n \exp[\imath A t]}{\imath - \tau_e(A-g)}\right). \end{aligned} \tag{5.6}$$

The argument of the logarithm in the above equation is the radial part of the complex solution, which corresponds to the final eccentricity of the KBO, with an equivalent expression

for the inclination:

$$e_{kbo}^{final} = \frac{e_n^0 \tau_e A_n}{\sqrt{1+\tau_e^2(g-A)^2}}$$
$$i_{kbo}^{final} = \frac{i_n^0 \tau_i B_n}{\sqrt{1+\tau_i^2(f-B)^2}}. \tag{5.7}$$

In principle, we could have arrived at the same answer by complex-expanding the solution and taking the square root of the sum of the squares of the real and imaginary parts, although the intermediate expressions would have been considerably more messy.

The above equations can be simplified even further by considering their limiting regimes. If the decay time-scale is much longer than the beat frequency $(g-A, f-B)$, we can Taylor expand the equations to first order in $(1/\tau)$ around zero. The answer then becomes independent of τ.

$$e_{kbo}^{final} \simeq e_n^0 \frac{A_n}{g-A} \qquad i_{kbo}^{final} \simeq i_n^0 \frac{B_n}{f-B} \tag{5.8}$$

This procedure is equivalent to assuming that $\tau_e^2(g-A)^2 \gg 1$ or $\tau_i^2(f-B)^2 \gg 1$ and throwing away the 1 under the square root in the denominator[1]. It is clear from Figure 2, where the approximate solutions are plotted as big dots, that quantitative agreement with the "full" solution (equation 4) is excellent, in the parameter regime of interest.

Figures 3 and 4 show the secular excitation of initially cold KBOs eccentricities and inclinations between the 3:2 and 2:1 MMRs with $g = 0, g_8, 2g_8$ and $3g_8$ as solid lines. These

[1]Alternatively, if the decay time-scales are short, we are in the non-adiabatic regime, where the solutions become $e_{kbo}^{final} \simeq e_n^0 \tau_e A_n$ and $i_{kbo}^{final} \simeq i_n^0 \tau_i B_n$

solutions suggest that if one is to retain an eccentricity below $e < 0.1$ and inclination below $i < 5^\circ$, Neptune's average orbital precession and nodal recession rates must have exceeded $\sim 3g_8$ during the eccentric phase. The enhanced precession is primarily a consequence of Uranus. When Neptune scatters, it arrives somewhat closer to the sun than its current orbit and migrates to ~ 30AU by scattering KBOs (here, we have implicitly omitted this effect by stating that the coefficients A, A_n, B, and B_n are constant). Thus, at the time of scattering, the semi-major axis ratio of Neptune to Uranus may be lower, leading to an enhanced precession. Additionally, the mass contained in the Kuiper belt may also play a role in inducing secular precession of Neptune.

The solution described above gives eccentricities that monotonically decrease with semi-major axes. However, as already discussed above, the observed cold population exhibits a somewhat different behavior, with low-eccentricity objects progressively disappearing in the vicinity of the 2:1 MMR. This dynamically unique structure (i.e. the wedge - see Figure 1) is an essential feature to any proposed formation mechanism for the cold classicals.

A wedge-like structure cannot be reproduced by a sweeping 2:1 MMR in an instability-driven formation model. Unlike the smooth migration scenario, where resonant capture is possible (Henrard and Lemaitre, 1983; Malhotra, 1995; Ketchum et al., 2011), when Neptune is eccentric, the chaotic motion, that arises from resonant splitting (Wisdom, 1980), leads to an effective randomization of the eccentricities (Quillen, 2006). In other words, the KBOs that are temporarily captured do not form a coherent structure such as the wedge. An alternative scenario for formation of the wedge is one where the local population ends

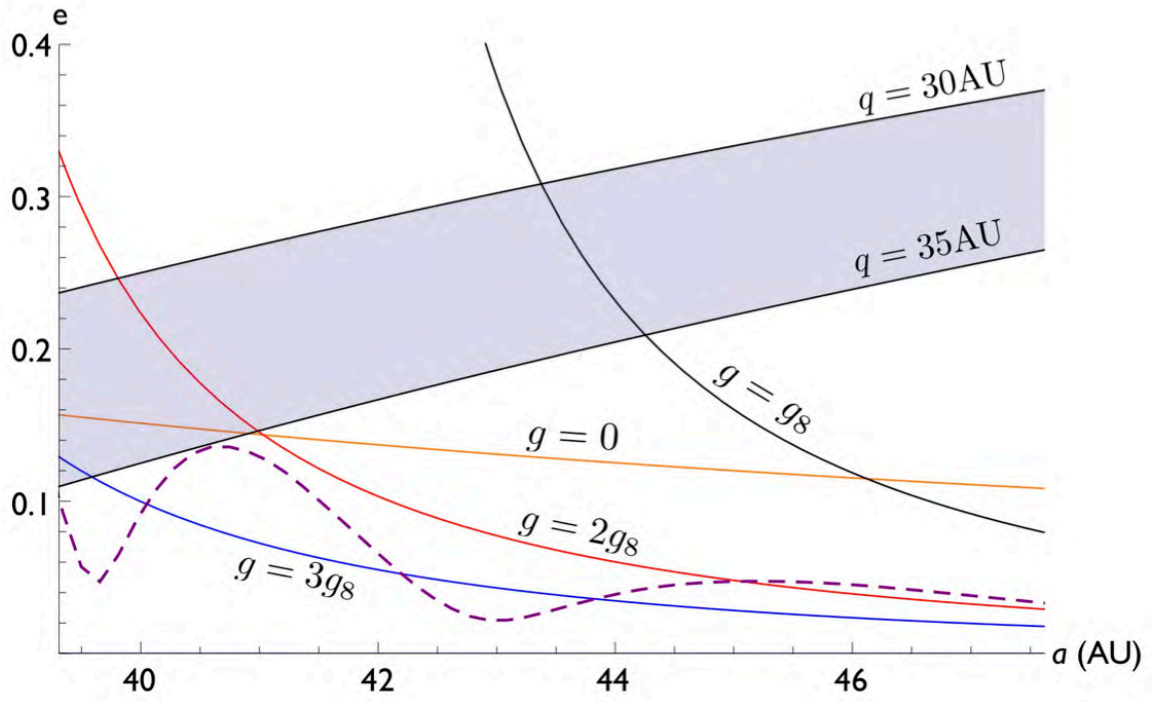


Figure 5.3: Post-excitation (final) eccentricities in the cold region of the Kuiper belt. Solutions with $g = 0, g_8, 2g_8$ and $3g_8$ are presented as solid lines. Note that in order to retain nearly-circular orbits $g \gtrsim 3g_8$ is required. The dashed line represents a solution where Neptune's precession rate is not kept constant. The shaded region corresponds to the scattered disk.

at 45AU, and the wedge is a result of an extended scattered disk with $q \sim 40$AU (Gladman et al., 2008). It is unlikely, however, that in the extended scattered disk scenario, the low inclinations of scattered objects could be preserved.

Here, we propose the formation of the wedge to be a consequence of secular perturbations. Thus, we seek to modify the above secular solution such that it yields eccentricities that are not monotonically decreasing with semi-major axes in the region of interest. As already described above, the controlling parameter in the secular solution is g. So far, we have kept g constant. However, since Neptune scatters numerous KBOs during its circularization, and the orbits of other planets (particularly Uranus) are changing as well, one would expect Neptune's precession to vary considerably, in a chaotic manner.

It is difficult to predict the exact nature of this variation without a detailed calculation, so here we consider an extreme case as a proof of concept. Namely, we set $g = 4g_8$ at all times, except $\tau < t < 1.1\tau$, where we set $g = 0$. Note that the precession of Neptune need not necessarily stop. We are choosing $g = 0$, rather than a diminished precession rate (such as, say $g = g_8$) merely for the sake of argument. An analytical solution is attainable in a similar fashion as above, by breaking up the integration into 3 separate time intervals. If g is not held constant through out Neptune's circularization, the final eccentricity and inclination take on a different character. Qualitatively, this can be understood as follows: when Neptune stops precessing, it starts to induce considerable oscillations in eccentricities of KBOs; however, once the precession becomes rapid again, the modulation stops and the eccentricities become frozen-in. These solutions are plotted as dashed curves in Figures 3 and 4. The details of the non-monotonic solution depend on when and for how long Neptune's precession is halted, and change further if the precession is merely slowed down, rather than stopped. Furthermore, the dashed curves in Figures 3 and 4 shift to larger semi-major axes if the free precessions of the KBOs (A, B) are enhanced. While it is understood that these calculations do not reproduce the cold classical population in detail, they do show that primordially unexcited objects can retain cold orbits in face of dynamical excitation, and coherent structure can be formed in the context of a purely secular solution.

5.4 Numerical Simulations

Having motivated in-situ formation of the cold classical population with analytical arguments, we now turn to numerical N-body simulations for confirmation of the above results

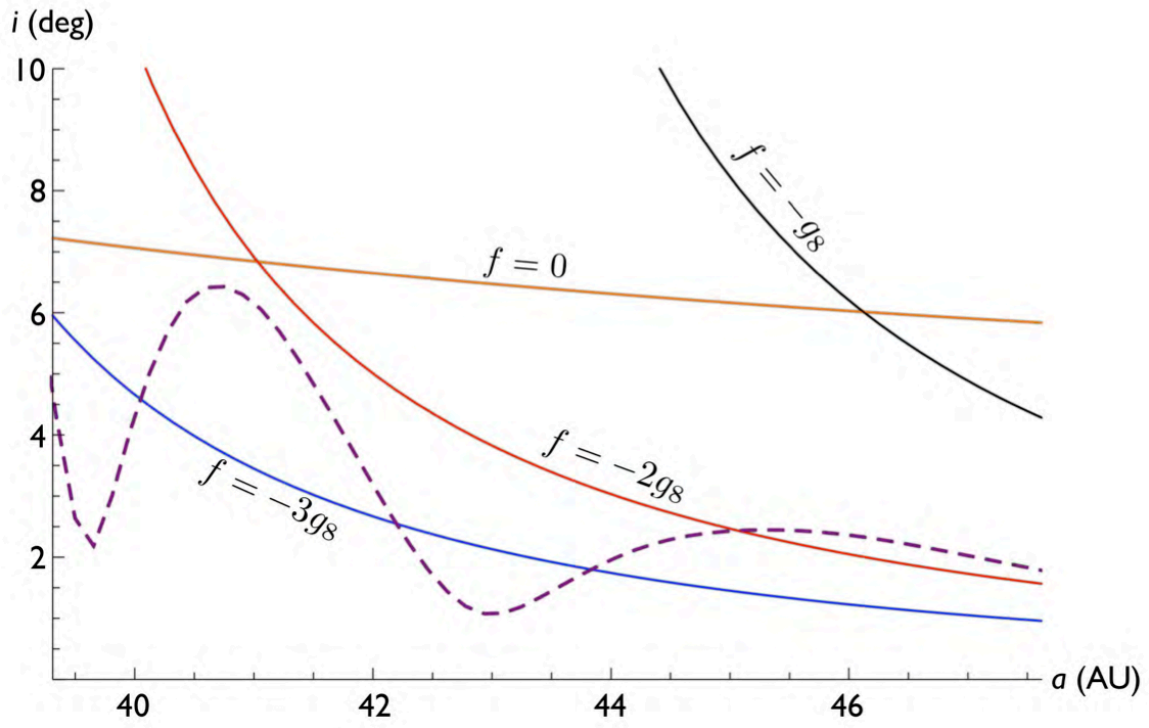


Figure 5.4: Post-excitation (final) inclinations in the cold region of the Kuiper belt. Solutions with $f = 0, g_8, 2g_8$ and $3g_8$ are presented as solid lines. Note that in order to retain $i \lesssim 5^\circ$ in the $42 - 45$AU region $f \lesssim -3g_8$ is required. The dashed line represents a solution where Neptune's nodal recession rate is not kept constant. Note that the quantitative character of the solution here is subtly different from the eccentricity solution (Figure 3). This is because B involves a Laplace coefficient of the first kind, while A involves that of the second kind.

and inclusion of omitted physics (such as close-encounters, mean-motion resonances, and higher-order secular terms in the disturbing function). In this study, the integrations were performed using the *mercury6* integration software package (Chambers, 1999) utilizing the "hybrid" algorithm. The disk was composed of two components. The massive planetesimal swarm, containing 3000 particles, resided between the immediate stability boundary of the initial multi-resonant configuration and $\sim$ 35AU. This was followed by a disk of another 3000 mass-less particles that extended to 60 AU. Thus we are assuming that a significant density gradient exists, in the vicinity of Neptune's final orbit, such that the mass

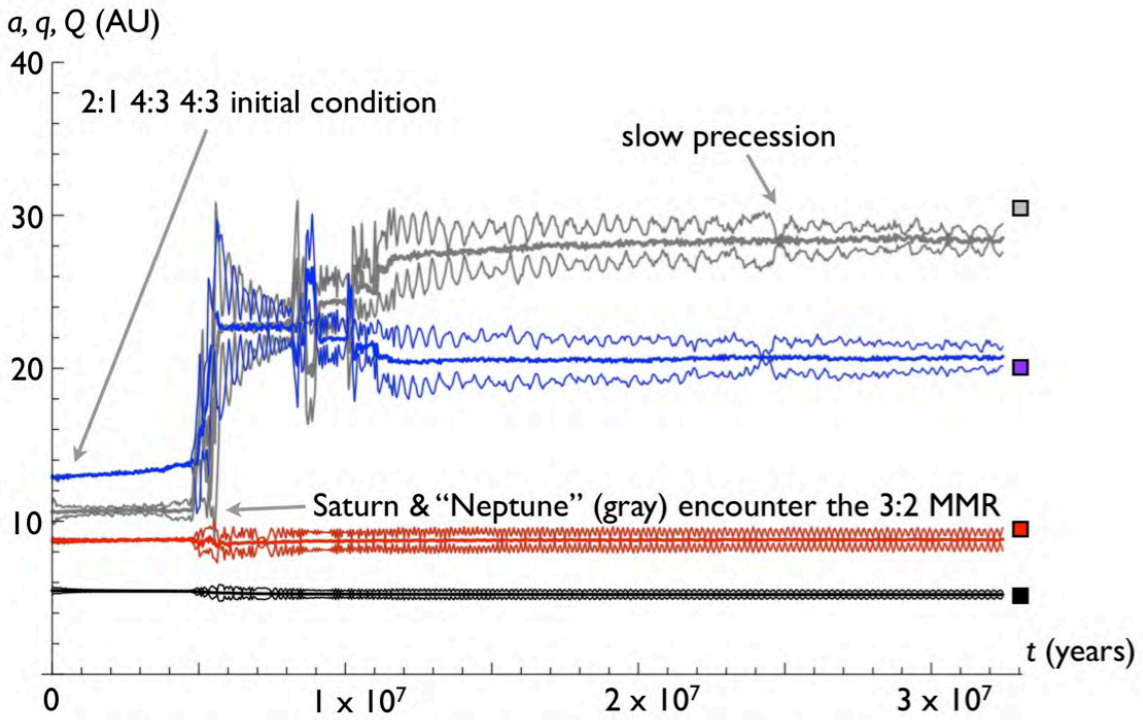


Figure 5.5: Orbital evolution of planets. The system starts out in the (2:1 4:3 4:3) initial condition, and undergoes a brief period of instability when Neptune (gray) and Saturn (Red) encounter a mutual 3:2 MMR. At $t \approx 22$Myr, Neptune's precession rate temporarily slows down and sculpts the wedge. The boxes on the right of the plot correspond to actual semi-major axes of the giant planets. An evolved remnant planetesimal disk of this simulation is presented in Figures 7 & 8.

in the outer disk is insufficient to drive Neptune's migration. However, the numbers of particles were chosen due to considerations of computational cost and are not intended to be representative of the relative fraction of bodies in the planetesimal disk in any way. The initial conditions were drawn from the eight multi-resonant states that were identified by Batygin & Brown (2010) as being compatible with an instability formation model. The planetesimals were initialized on near-coplanar, near-circular orbits ($e \sim \sin i \sim 10^{-3}$). The self-gravity of the planetesimal swarm was neglected to reduce computational cost of the experiments, as 30 permutations of each initial condition were integrated.

Batygin & Brown (2010) used the presence of scattering events between an ice giant and a gas giant, followed by a transient phase of high eccentricity, as a proxy for whether successful formation of the classical Kuiper belt can occur. Further constraints on the initial conditions can be placed by considering the reproduction of the outer solar systems' secular eigenmodes. Particular difficulty has been found in ensuring that the amplitude of Jupiters g_5 mode is larger than that of the g_6 mode (Morbidelli et al., 2009). Having completed all of the integrations, we checked the relative amplitudes of the g_5 and the g_6 modes in all solutions. Surprisingly, we found that despite a transient period of instability and gas giant/ice giant scattering, the (3:2 3:2 4:3)[2] and the (5:3 4:3 4:3) initial conditions did not reproduce the secular architecture of the planets, in neither this set nor in the set of integrations of Batygin & Brown (2010). If Jupiter and Saturn were indeed initially locked in the 3:2 MMR, as hydrodynamic simulations suggest (Masset & Snellgrove, 2001; Morbidelli & Crida, 2007; Pierens & Nelson, 2008), only the (3:2 3:2 5:4) and (3:2 4:3 4:3) initial conditions are left as a viable options for the starting state of the solar system.

As already discussed in section 2, interactions between the cold outer disk and the outer-most ice-giant are largely independent of the starting condition, since scattering in a successful simulation always sets the planets onto orbits that are close to that of the current solar system, but with moderate eccentricities. Consequently, we did not restrict our analysis to any particular initial condition. Out of our set of 180 integrations, in 8 cases, primordially cold objects were able to retain unexcited orbits in addition to the gas-giant eigenmodes being reproduced correctly. Here, we focus on two representative integrations:

[2]In our notation, each pair of numbers represents an MMR in the multi-resonant initial condition. For example, (3:2 3:2 4:3) corresponds to a starting state where Jupiter & Saturn, as well as Saturn & Uranus are in 3:2 MMRs, while Uranus & Neptune are in a 4:3 MMR.

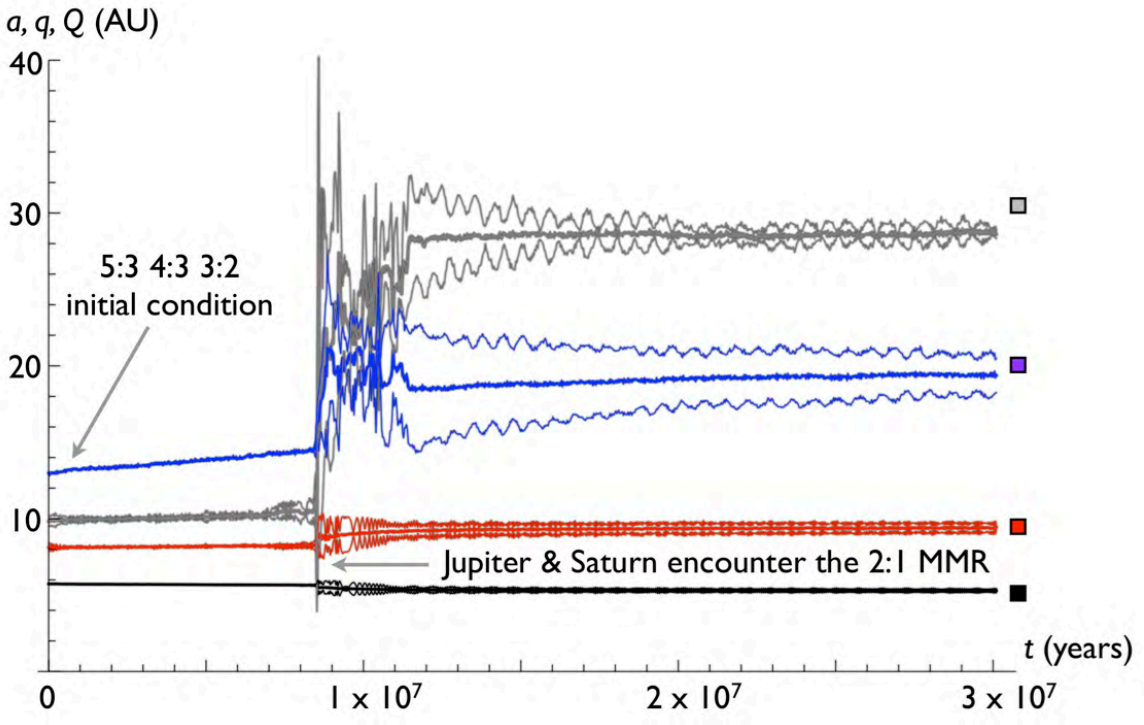


Figure 5.6: Orbital evolution of planets. The system starts out in the (5:3 4:3 3:2) initial condition, and undergoes a brief period of instability when Saturn (Red) and Jupiter (Black) encounter a mutual 2:1 MMR. The boxes on the right of the plot correspond to actual semi-major axes of the giant planets. An evolved remnant planetesimal disk of this simulation is presented in Figures 9 & 10.

one starting from the (2:1 4:3 4:3) initial condition (Figure 5) and another starting from the (5:3 4:3 3:2) initial condition (Figure 6). In both cases, the cold classical population is produced, but the wedge is only formed in the simulation that starts from the (2:1 4:3 4:3) initial condition (although it is somewhat smaller than its observed counterpart). Note, that the formation of the wedge has little to do with the initial condition - rather, its production is a random process. Similarly, the exact degree of excitation of the cold population's inclinations is sensitively dependent on the details of Neptune's evolution, which is chaotic. Thus, the fact that the wedge is reproduced in one simulation and the degree of excitation of the inclinations is reproduced in another are unrelated results.

A vast majority ($> 90\%$) of the objects in the cold classical region (i.e. $a \gtrsim 42$AU) are retained in our simulations on stable orbits. On the contrary, only about a few thousandth of the particles in the inner disk are emplaced onto stable orbits in the Kuiper belt region. This implies that in order to self-consistently study the formation of the Kuiper belt, $N \gg 3000$ is needed. Unfortunately, the required resolution is not computationally feasible. However, the problem can still be addressed by the use of "tracer" simulations, an approach already utilized in the context of Kuiper belt formation by Levison et al. (2008).

In a tracer simulation the planets and planetesimals are not self-consistently evolved in time. Rather, the evolution of the planets is pre-loaded from a master simulation and the planetesimals, which are treated as test particles, are evolved subject only to gravitational interactions with the planets. At the beginning of a tracer simulation, the tracer disk is initialized to have the same distribution as the massive component of the planetesimal disk. Consequently, at all times during the integration, the tracer particles also have an identical orbital distribution to that of the massive planetesimals. Each simulation was seeded with 100 test particles and integrated on Caltech's *PANGU* super-computer. We employed the Bulirsch-Stoer algorithm (Press et al., 1992) in our tracer integrations.

We performed 200 tracer simulations for each of the evolutions presented in Figures 5 & 6. This amounts to evolving a primordial disk of ~ 26000 particles, including the outer belt. After the ~ 30 Myr simulations were completed, approximately $\sim 7\%$ of the particles that originated interior to 35AU had semi-major axes in the range $35\text{AU} < a < 60\text{AU}$, shown as pale blue dots in Figures 7 - 10. We further cloned the populations[3] of

[3]At the end of the simulations, there was only statistically significant structure in the $a, e.i$ distributions. The orbital angles took on random values during scattering.

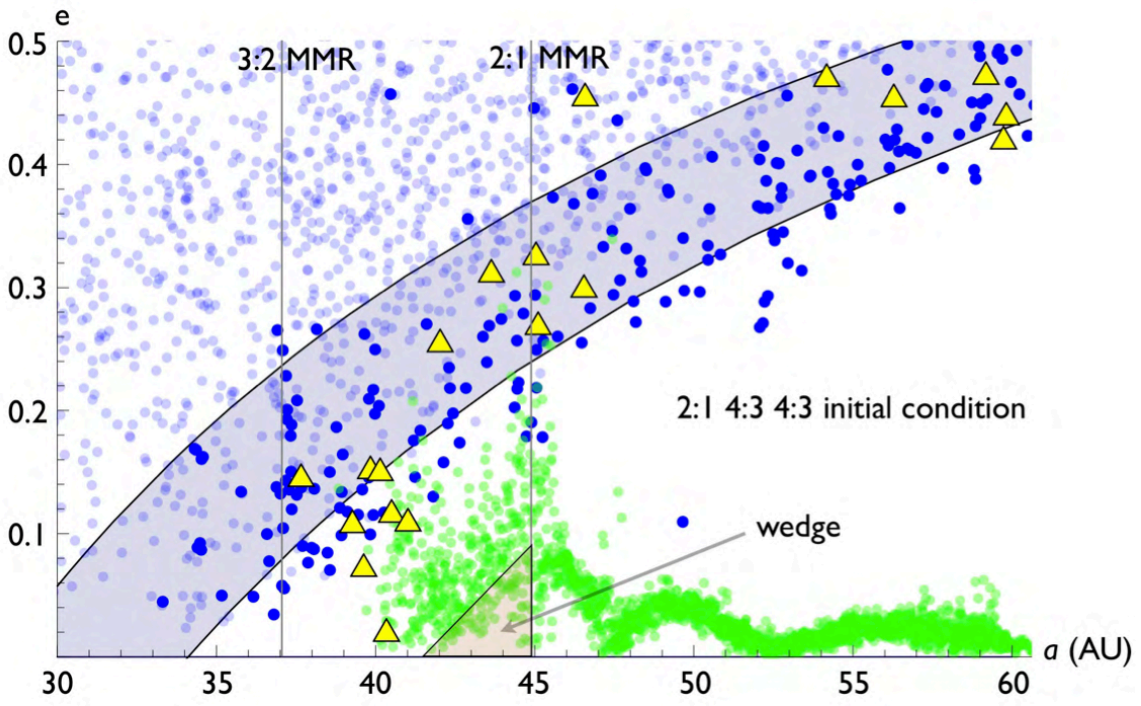


Figure 5.7: Eccentricity distibution of the remnant planetesimal disk of the simulation that starts from the (2:1 4:3 4:3) multi-resonant state (see Figure 5). The pale blue dots show objects that originated interior to $\sim$ 35AU, 30Myr after the beginning of the simulation. The dark blue dots represent objects that originated interior to 35AU, but are stable over 500Myr. Green dots represent the test-particles that originate between 40 & 60 AU. Yellow triangles represent test particles that originated between 35 and 40 AU. A wedge that is somewhat similar to the observed one (see Figure 1) forms in this simulation, as a result of a temporary slow-down in Neptune's precession rate (see Figure 11). Note that in this simulation, the classical Kuiper belt region lies between $\sim$ 37 and $\sim$ 45AU, as Neptune's final semi-major axis is $a \sim 28.3$AU. However, the aim here is to elucidate the physical mechanisms, rather than reproduce the actual Kuiper belt. The shaded region corresponds to the scattered disk.

tracer particles in the Kuiper belt region to effectively increase the number of implanted hot classical, scattered, and resonant particles by another factor of 6. The resulting Kuiper Belt, including the test-particles that originate beyond 35AU, was then evolved for an additional 500 Myr to ensure that all unstable particles have time eject. At the end of the 500Myr, only $\sim 5\%$ of the implanted objects, that were present at the end of the 30Myr simulations, ended up on stable orbits. Consequently, the cumulative fraction of objects that are implanted into

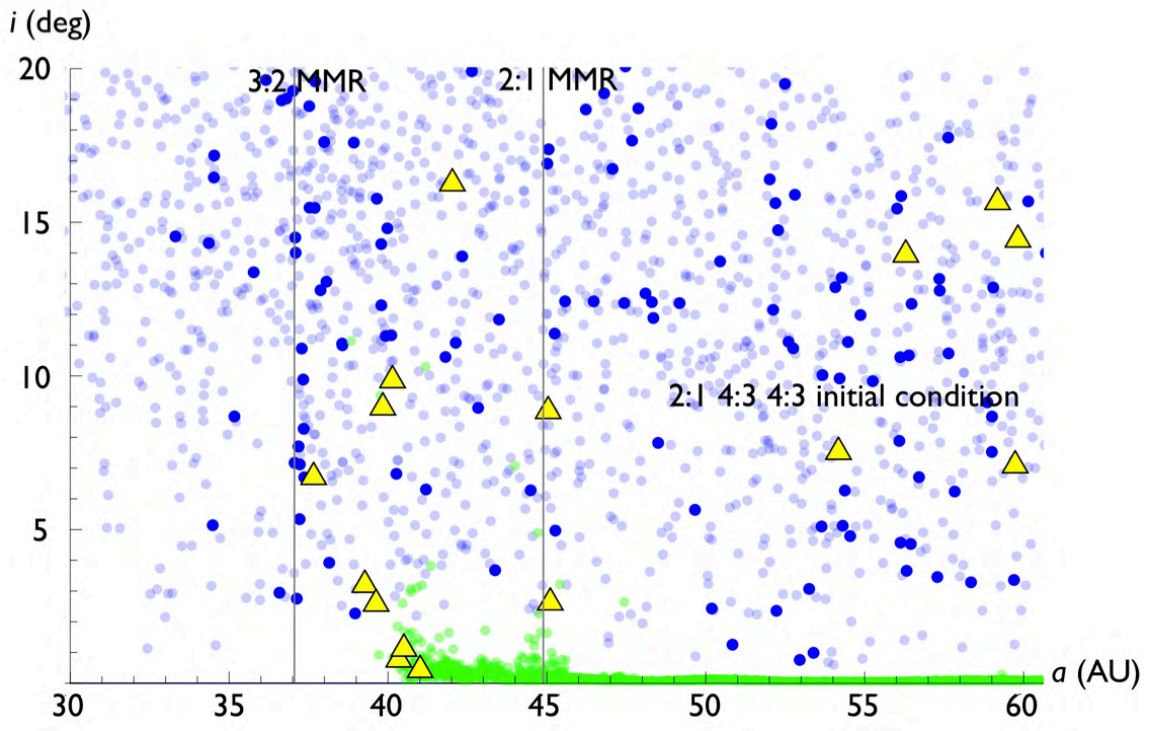


Figure 5.8: Inclination distribution of the remnant planetesimal disk of the simulation that starts from the (2:1 4:3 4:3) multi-resonant state (see Figure 5). The pale blue dots show objects that originated interior to $\sim$ 35AU, 30Myr after the beginning of the simulation. The dark blue dots represent objects that originated interior to 35AU, but are stable over 500Myr. Green dots represent the test-particles that originate between 40 & 60 AU. Yellow triangles represent test particles that originated between 35 and 40 AU.

the Kuiper belt from the inner disk is $\sim 0.3\%$. The stable objects are shown as dark blue dots in Figures 7 - 10.

Note that in our simulations, the resonant populations are considerably diminished in number. This is largely a cost of performing self-consistent simulations with planetesimals that are unrealistically massive. Every time Neptune scatters a KBO, its resonances jump unrealistically far, disturbing the resonant KBOs, leading to their eventual ejection (Murray-Clay & Chiang, 2006). In a suite of customized simulations where the instability still occurs, but planets are analytically guided to their final orbits, and gravity is softened (Levison et al., 2008), Neptune's MMR's end up overpopulated. This leads one to believe

that the true parameter regime of Neptune's migration resided somewhere between what is presented in this work and that of Levison et al. (2008) (Morbidelli, personal communication).

Although both of the integrations presented here produce a cold classical belt, it is immediately apparent that the wedge is only produced in the integration that starts from the (2:1 4:3 4:3) initial condition, although again the process has little do with the choice of initial condition. Furthermore, from Figure 7, it can be readily inferred that the production of the wedge must be a secular effect since the structure in this simulation extends beyond the 2:1 MMR, i.e. the unswept region. Note that owing to the enhanced free precession of the KBOs (due to presence of a massive Kuiper belt), the wedge structure is shifted to the right, compared with analytical estimates presented in the previous section.

In the context of these integrations, we are further able to confirm that the formation of the wedge is due to a considerable slowdown in Neptune's precession. During circularization in the integration that starts from a (5:3 4:3 3:2) initial condition, Neptune precession is always roughly $g \approx 4.7 g_8$, while it is eccentric. On the contrary, in the integration that starts from the (2:1 4:3 4:3) initial condition, Neptune's precession rate varies considerably ($1.6 g_8 \lesssim g \lesssim 3.7 g_8$) between 23Myr and 25Myr (Figure 11). The presence of a mechanism for the successful formation of the wedge, from a local population is an important argument for confirmation of the in-situ formation of the cold classical population in the context of an instability model.

It is noteworthy that in the results of the simulation, the wedge appears much less coherent, at semi-major axes interior to the 2:1 MMR. This is a consequence of eccentric

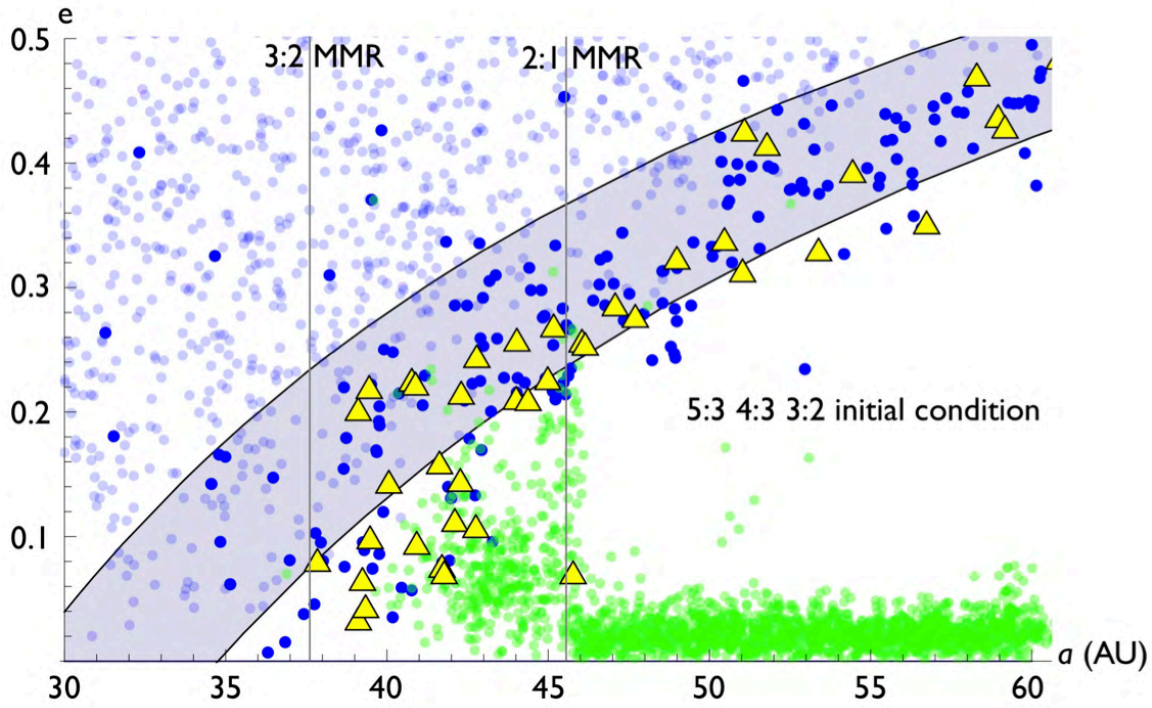


Figure 5.9: Eccentricity distribution of the planetesimal disk of the simulation that starts from the (5:3 4:3 3:2) multi-resonant state (see Figure 5). The pale blue dots show objects that originated interior to $\sim$ 35AU, 30Myr after the beginning of the simulation. The dark blue dots represent objects that originated interior to 35AU, but are stable over 500Myr. Green dots represent the test-particles that originate between 40 & 60 AU. Yellow triangles represent test particles that originated between 35 and 40 AU. Note that the wedge does not form in this simulation, because Neptune's precession never slows down, while it is eccentric. Note that in this simulation, the classical Kuiper belt region lies between $\sim$ 38 and $\sim$ 46AU, as Neptune's final semi-major axis is $a \sim$ 29AU. However, the aim here is to elucidate the physical mechanisms, rather than reproduce the actual Kuiper belt. The shaded region corresponds to the scattered disk.

resonant sweeping. Because of Neptune's considerable eccentricity, the KBO multiplet and the Neptune multiplet of the resonance overlap even for small KBO eccentricities. This allows the KBO to randomly explore the phase-space occupied by both sections of the resonance. However, as Neptune's eccentricity is monotonically decreasing, so is the phase space volume occupied by Neptune's multiplet of the 2:1 MMR, making capture impossible (Quillen, 2006). Moreover, because of different precession rates, the nominal location of

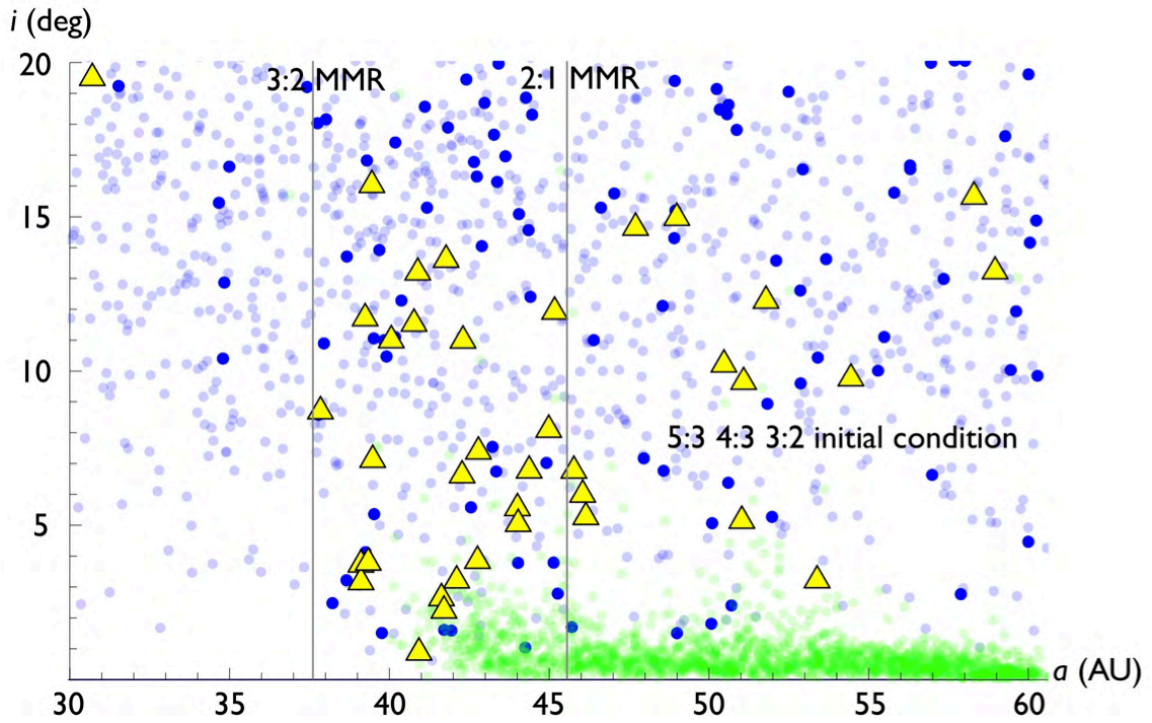


Figure 5.10: Inclination distribution of the planetesimal disk of the simulation that starts from the (5:3 4:3 3:2) multi-resonant state (see Figure 5). The pale blue dots show objects that originated interior to ~ 35AU, 30Myr after the beginning of the simulation. The dark blue dots represent objects that originated interior to 35AU, but are stable over 500Myr. Green dots represent the test-particles that originate between 40 & 60 AU. Yellow triangles represent test particles that originated between 35 and 40 AU.

Neptune's multiplet of the resonance lags (i.e smaller semi-major axis) that of the KBO. Thus, if a KBO exits the resonance shortly after it enters, it tends to get transported closer to the sun, since it enters at the KBO multiplet and exits at the Neptune multiplet. The change in semi-major axes, however is only the resonant splitting width, so it is rather small ($\delta a < 0.1$AU). This randomization of the orbital elements causes the inner part of the wedge to appear less coherent in Figure 7.

It is finally worth noting that although KBOs that become the cold classical population are able to roughly retain their primordial orbital distribution, the objects between 35AU and 40AU inevitably get scattered by Neptune during the instability. Indeed, in both sim-

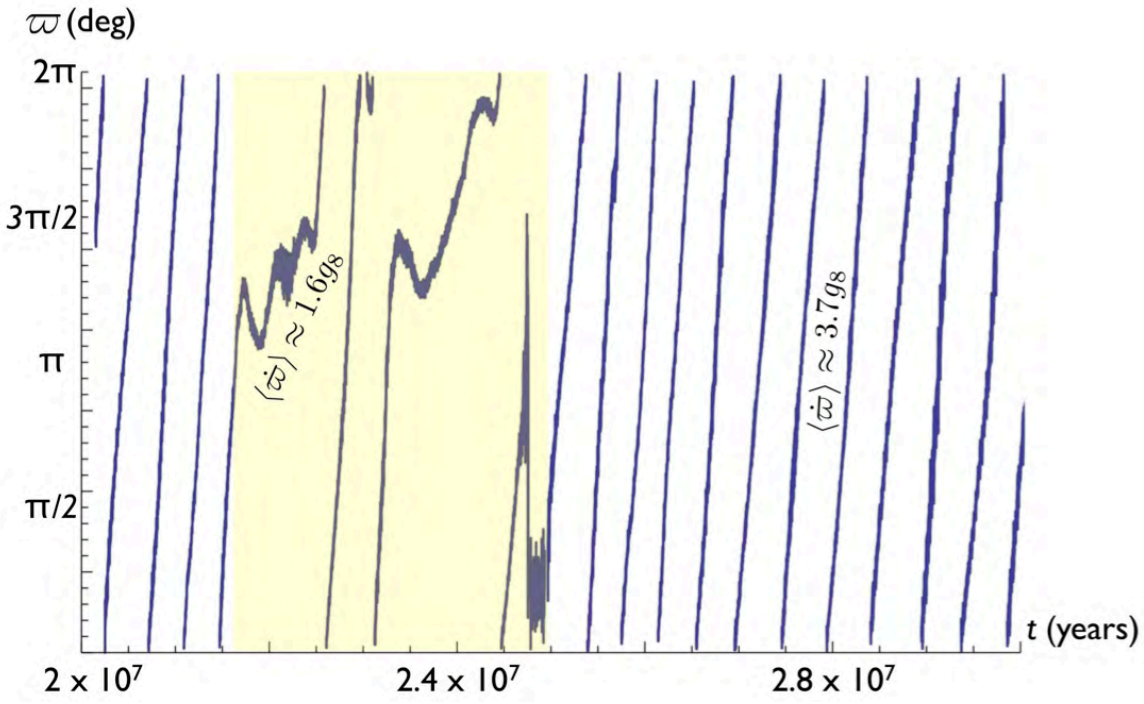


Figure 5.11: Precession of Neptune's longitude of perihelion in the simulation that originates from the (2:1 4:3 4:3) multi-resonant initial condition (see Figure 5). Most of the time, Neptune's precession rate exceeds its current value by a factor of a few. However, the precession rate slows down considerably at $t \approx 22$Myr. The wedge forms as a result of the highlighted slowdown in Neptune's apsidal precession.

ulations presented here, the scattered cold classicals (shown as yellow triangles in Figures 7-10) join the scattered disk as well as the hot classical population, while some particles get trapped in resonances temporarily, during their evolution.

The fact that these lifted objects mostly get emplaced onto stable orbits is suggestive that the results of intrusion of inclined populations by cold-classical like objects, that took place during the instability, should still be observable today. In other words, the in-situ formation scenario for cold-classicals presented here predicts that that a class of objects, occupying the same unique color region as the cold classicals, should be present in the excited populations.

5.5 Discussion

In this paper, we present a self-consistent dynamical model for the evolution of a primordial cold classical population of the Kuiper belt, in the context of an instability-driven formation scenario for the solar system. We show, from simple analytical considerations, that the cold belt can survive the transient period of dynamical instability, inherent to the planets. In order for a primordially cold population of KBOs to maintain an unexcited state, the average apsidal precession and nodal recession rates of Neptune during the transient phase of instability must have been considerably faster than what is observed in today's solar system. Simultaneously, successful formation of the wedge (see Figure 1) requires that the apsidal precession rate drops by a factor of a few for a short period of time. Numerical integrations presented in this work confirm the results of the analytical calculations and reveal a particular result, where the formed cold population and the wedge closely resemble their observed counter-parts. The dynamical evolution of cold classicals we propose here is in close agreement with the uniqueness of cold classical's physical characteristics.

In situ formation of cold classicals brings to light the issue of truncation of the classical belt near the 2:1 MMR. In the chaotic capture mechanism proposed by Levison et al. (2008), the outer edge comes about naturally, as the 2:1 MMR sculpts the belt. In our solution, however, a cold belt that extends further out is surely possible. Thus, we are forced to attribute the proximity of the edge and the 2:1 MMR to a mere coincidence. Another question of interest is the fate of primordially cold binaries in the $35 - 40$AU region. It is likely that many of these binaries will get disrupted by close-encounters with Neptune, although the exact fraction will depend on the details of Neptune's evolution. Consequently,

an in-depth analysis of the evolution of the scattered cold KBOs may open up an avenue towards further constraining the orbital history of Neptune.

Although in-situ formation of cold classicals resolves a pressing dynamical problem within the Nice model, it gives rise to a new issue that requires attention. Namely, the outstanding question of importance is planetesimal formation beyond ~ 35AU, given the steep size distribution of the cold classicals. In other words, how is the formation of planetesimals, up to ~ 200 km in size accomplished in such a low-density environment?

Although the answer to this question is by no means trivial, one possible solution to this problem lies within the context of streaming instabilities (Youdin & Goodman, 2005). Streaming instabilities have already been suggested as the dominant formation process in the cold classical population, as gravitational collapse has been shown to yield wide binaries (Nesvorný et al., 2010). Importantly, in the proposed picture, planetesimal formation is a threshold process, that "turns on," only when gas drag accumulates a critical amount of dust in a given location within the solar nebula. Thus, one can in principle envision a system where most of the dust gets carried inward of ~ 35AU by gas drag, but infrequently, the dust surface density reaches a critical value in the outer nebula, causing a few, but sizable planetesimals to be born. Such a scenario would likely result in a very sharply decreasing surface density profile in the outer nebula, at the epoch of disappearance of the gas. As a result, this picture would imply the existence of a steep density gradient in the primordial planetesimal disk, such as the one we require in our model, consistent with preventing Neptune's extended migration.

Another possibility for the formation process is hierarchical coagulation, where plan-

etesimal growth is accomplished by collisions among smaller objects in a quiescent environment (Kenyon, 2002). Particularly, it has been suggested that if aided by turbulent concentration, hierarchical coagulation could yield the desired mass of the cold classical population (Cuzzi et al., 2010). In fact, even if the original mass of the cold belt exceeded its current value, erosion by collisional grinding could in principle be invoked to reduce the overall mass. However this process may prove problematic in reproducing the observed wide binary fraction of the cold belt (Nesvorny et al., 2011).

Whatever the formation process for the cold classicals is, the results presented here have considerable implications. First and foremost, the successful retention of the cold-classical population in the context of an instability-driven model, fixes the most significant drawback of the Nice model. Second, our scenario suggests that the cold classical Kuiper belt is the only population of objects in the outer solar system that has not been transported away from its formation site. Furthermore, assuming that collisional grinding has played a negligible role in the cold population's evolution (as suggested by the observed binary fraction (Nesvorny et al., 2011)), the cold classical population essentially yields the surface density of the solar nebula at $a \sim 45$AU, since the majority of the KBOs are retained in place. This potentially makes the cold classicals a unique laboratory for the study of surface processes as well as the chemistry of the primordial solar nebula. Third, based upon the results of the numerical simulations, we expect that objects that are physically similar to the cold classicals should be scattered throughout the Kuiper belt and as a result may explain the spectral similarity between cold classicals and some objects at higher inclinations.

In conclusion, it appears that quantitative evaluation of planetesimal formation beyond

~ 35AU is required to draw a complete picture of the of the in-situ formation and evolution scenario for the cold classicals. However, the considerable improvement of the model for early dynamical evolution of the Kuiper belt presented here supports the overall validity of the hypothesis.

Acknowledgments We thank Alessandro Morbidelli, Hal Levison, Darin Ragozzine and Peter Goldreich for useful conversations.

Bibliography

Batygin, K., & Brown, M. E. 2010, The Astrophysical Journal, 716, 1323

Brasser, R., Morbidelli, A., Gomes, R., Tsiganis, K., & Levison, H. F. 2009, Astronomy & Astrophysics, 507, 1053

Brown, M. E. 2001, Icarus, 151, 190

Chambers, J. E. 1999, Monthly Notices of Royal Astronomical Society, 304, 793

Cuzzi, J. N., Hogan, R. C., & Bottke, W. F. 2010, Icarus, 208, 518

Fernandez, J. A., & Ip, W.-H. 1984, Icarus, 58, 109

Fraser, W. C., Brown, M. E., & Schwamb, M. E. 2010, Icarus, 210, 944

Hahn, J. M., & Malhotra, R. 2005, The Astronomical Journal, 130, 2392

Henrard, J., & Lamaitre, A. 1983, Celestial Mechanics, 30, 197

Levison, H. F., Morbidelli, A., Vanlaerhoven, C., Gomes, R., & Tsiganis, K. 2008, Icarus, 196, 258

Jewitt, D., & Luu, J. 1993, Nature, 362, 730

Morbidelli, A., Levison, H. F., & Gomes, R. 2008, The Solar System Beyond Neptune, 275

Morbidelli, A., & Crida, A. 2007, Icarus, 191, 158

Morbidelli, A., Tsiganis, K., Crida, A., Levison, H. F., & Gomes, R. 2007, The Astronomical Journal, 134, 1790

Murray-Clay, R. A., & Chiang, E. I. 2006, The Astrophysical Journal, 651, 1194

Gomes, R., Levison, H. F., Tsiganis, K., & Morbidelli, A. 2005, Nature, 435, 466

Gladman, B., Marsden, B. G., & Vanlaerhoven, C. 2008, The Solar System Beyond Neptune, 43

Kenyon, S. J. 2002, Publications of the Astronomical Society of the Pacific, 114, 265

Ketchum, J. A., Adams, F. C., & Bloch, A. M. 2011, The Astrophysical Journal, 726, 53

Lykawka, P. S., & Mukai, T. 2005, Earth Moon and Planets, 97, 107

Morbidelli, A., Levison, H. F., Tsiganis, K., & Gomes, R. 2005, Nature, 435, 462

Morbidelli, A., & Brown, M. E. 2004, Comets II, 175

Morbidelli, A., Brasser, R., Tsiganis, K., Gomes, R., & Levison, H. F. 2009, Astronomy & Astrophysics, 507, 1041

Murray-Clay, R. A., & Chiang, E. I. 2005, The Astrophysical Journal, 619, 623

Masset, F., & Snellgrove, M. 2001, Monthly Notices of Royal Astronomical Society, 320, L55

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

Nesvorný, D., Youdin, A. N., & Richardson, D. C. 2010, The Astronomical Journal, 140, 785

Nesvorny, D., Vokrouhlicky, D., Bottke, W. F., Noll, K., & Levison, H. F. 2011, arXiv:1102.5706

Pierens, A., & Nelson, R. P. 2008, Astronomy & Astrophysics, 482, 333

Parker, A. H., & Kavelaars, J. J. 2010, The Astrophysical Journal Letters, 722, L204

Thommes, E. W., Bryden, G., Wu, Y., & Rasio, F. A. 2008, The Astrophysical Journal, 675, 1538

Tsiganis, K., Gomes, R., Morbidelli, A., & Levison, H. F. 2005, Nature, 435, 459

Trujillo, C. A., & Brown, M. E. 2002, The Astrophysical Journal Letters, 566, L125

Malhotra, R. 1995, The Astronomical Journal, 110, 420

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Cambridge: University Press, c1992, 2nd ed.,

Stephens, D. C., & Noll, K. S. 2006, The Astronomical Journal, 131, 1142

Stewart, G. R., & Wetherill, G. W. 1988, Icarus, 74, 542

Quillen, A. C. 2006, Monthly Notices of Royal Astronomical Society, 372, L14

Wisdom, J. 1980, The Astronomical Journal, 85, 1122

Wu, Y., & Goldreich, P. 2002, The Astrophysical Journal, 564, 1024

Youdin, A. N., & Goodman, J. 2005, The Astrophysical Journal, 620, 459

Chapter 6

Instability-Driven Dynamical Evolution Model of a Primordially 5 Planet Outer Solar System

Originally published as:
Batygin, K., Brown, M. E., & Betts, H. 2011, The Astrophysical Journal Letters, in press

6.1 Abstract

Over the last decade, evidence has mounted that the solar system's observed state can be favorably reproduced in the context of an instability-driven dynamical evolution model, such as the "Nice" model. To date, all successful realizations of instability models have concentrated on evolving the four giant planets onto their current orbits from a more compact configuration. Simultaneously, the possibility of forming and ejecting additional planets has been discussed, but never successfully implemented. Here we show that a large array of 5-planet (2 gas giants + 3 ice giants) multi-resonant initial states can lead to an adequate formation of the outer solar system, featuring an ejection of an ice giant during a phase of instability. Particularly, our simulations demonstrate that the eigenmodes which characterize the outer solar system's secular dynamics can be closely matched with a 5-planet

model. Furthermore, provided that the ejection timescale of the extra planet is short, orbital excitation of a primordial cold classical Kuiper belt can also be avoided in this scenario. Thus the solar system is one of many possible outcomes of dynamical relaxation and can originate from a wide variety of initial states. This deems the construction of a unique model of solar system's early dynamical evolution impossible.

6.2 Introduction

In the last two decades, discoveries of the Kuiper belt (Jewitt & Luu, 1993), as well as planets orbiting stars other than the Sun (Mayor & Queloz, 1995), have supplied the centuries-old quest to understand the formation of the solar system with fresh constraints and insights into physical processes at play. Among a multitude of newly proposed formation scenarios, the "Nice" model (Tsiganis et al., 2005; Gomes et al., 2005; Morbidelli et al., 2005) is particularly notable, as it has attained a considerable amount of success in reproducing the various observed features of the solar system. Within the context of the scenario envisioned by the Nice model, giant planets start their post-nebular evolution in a compact, multi-resonant configuration, and following a brief period of dynamical instability, scatter onto their current orbits (Morbidelli et al., 2007; Batygin & Brown, 2010; Walsh et al., 2011).

The first success of the Nice model lies in its ability to quantitatively reproduce the observed orbits of the giant planets, as well as their dynamical architecture (i.e. secular eigenmodes of the system) (Tsiganis et al., 2005; Morbidelli et al., 2009). Simultaneously, the brief instability, inherent to the model, provides a natural trigger to the Late Heavy

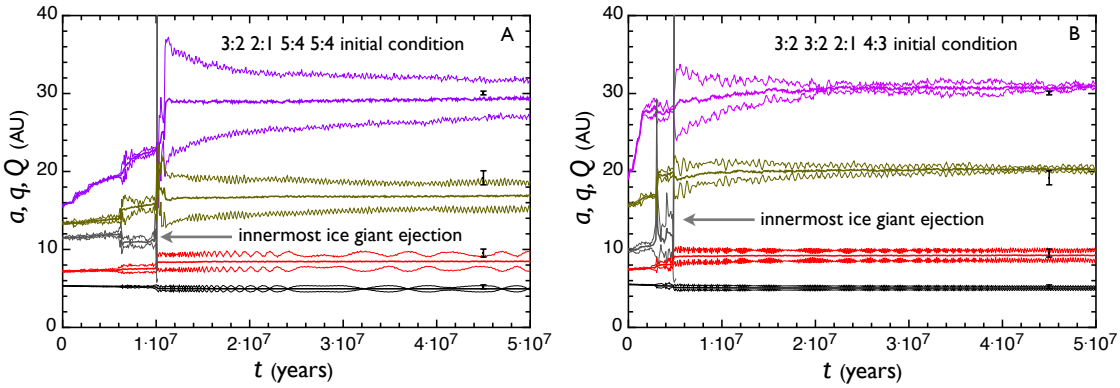


Figure 6.1: Orbital evolution of planets. Each planet's semi-major axis, as well as perihelion and apohelion distances are shown as functions of time. The actual perihelion and apohelion distances of the planets are also shown for comparison as black error bars. In both cases, the innermost ice-giant ejects during the transient phase of instability, leaving behind 4 planets, whose orbits resemble that of the solar system. See main text for a description of the initial conditions.

Bombardment (Gomes et al., 2005), as well as a transport mechanism for emplacement of dynamically "hot" Kuiper belt objects (KBOs) from inside of ~ 35AU (Levison et al., 2008). Meanwhile, it has been recently demonstrated that survival of a dynamically "cold" primordial population between Neptune's current 3:2 and 2:1 exterior mean-motion resonances (MMRs) is fully consistent with a Nice model-like evolution of the planets, implying an *in-situ* formation of the cold classical population of the Kuiper belt (Batygin et al., 2011). Finally, the presence of Jupiter's and Neptune's Trojan asteroids has been attributed to chaotic capture of planetesimals during the instability (Morbidelli et al., 2005; Nesvorný et al., 2007).

All successful realizations of the Nice model to date have been comprised exclusively of the four currently present giant planets. However, there exists no strong evidence that suggests that additional planets were not present in the solar system, at the epoch of the dispersion of the nebula. In fact, theoretical arguments, presented by Goldreich et al. (2004)

point to a possibility of initially forming as many as five ice giants, three of which get subsequently removed via ejections (see however Levison and Morbidelli (2007)). The dynamical sensibility of such a scenario is further strengthened by the fact that a considerable fraction of standard Nice model simulations result in an ejection of an ice-giant after an encounter with at least one of the gas giants.

In this paper, we explore an instability-driven dynamical evolution of a 5-planet system (2 gas giants + 3 ice giants) with an eye towards identifying a pathway towards reproduction of solar system-like dynamical architecture. In principle, the realm of possibility available to this study is enormous. Consequently, rather than performing a comprehensive parameter-search, here we limit ourselves to systems that contain an additional Uranus-like planet, with the aim of presenting a few proof-of-concept numerical experiments. The plan of the paper is as follows. In section 2, we describe our numerical setup. In section 3, we show that in our model, the planetary orbits, their secular eigenmodes, as well as various populations of the Kuiper belt are approximately reproduced. We conclude and discuss our results in section 4.

6.3 Numerical Experiments

The numerical setup of the simulations performed here was qualitatively similar to those presented by Batygin & Brown (2010) and Batygin et al. (2011). Particularly, the five giant planets were initialized in a compact, multi-resonant initial condition, surrounded by a massive planetesimal disk that extended between its immediate stability boundary and 30AU.

Two of the three ice-giants were taken to have the same mass as Uranus and were initially placed next to Saturn and as the outermost planet respectively. The middle ice-giant was taken to have Neptune's mass. In all simulations, Jupiter and Saturn started out in a 3:2 MMR, in accord with the results of hydrodynamical simulations of convergent migration of the planets in the solar nebula (Masset & Snellgrove, 2001; Morbidelli & Crida, 2007; Pierens & Nelson, 2008). The ice giants were also sequentially assembled into first order MMRs by applying dissipative forces, designed to mimic the presence of the nebula (Lee and Peale, 2002). Following resonant locking, each assembled multi-resonant initial condition was evolved in isolation for 10Myr, as an immediate test of orbital stability. This procedure yielded a total of 81 stable multi-resonant initial conditions.

The search for adequate dynamical evolutions was performed in two steps. First, we evolved 10 permutations of each initial condition, with planetesimal disks composed of $N = 1000$ planetesimals. Disk masses were chosen randomly between $M_{\rm disk}^{\rm min} = 25 M_\oplus$ and $M_{\rm disk}^{\rm max} = 100 M_\oplus$. The density profiles followed a power-law distribution, $\Sigma \propto r^k$ where the power-law index, k, was chosen randomly between $k_{\rm min} = 1$ and $k_{\rm max} = 2$. The planetesimals were initialized on near-circular orbits ($e \sim \sin i \sim 10^{-3}$). To reduce the already substantial computational cost, self-gravity of the planetesimal swarm was neglected.

Subsequently, we eliminated all initial conditions that did not yield any final systems that were comprised of 4 planets, reducing the number of viable initial conditions to 25. Then, an additional 30 permutations of these initial conditions were integrated with disks composed of $N = 3000$ planetesimals (but otherwise identical to those described above). Each integration was performed using the *mercury6* integration software package (Cham-

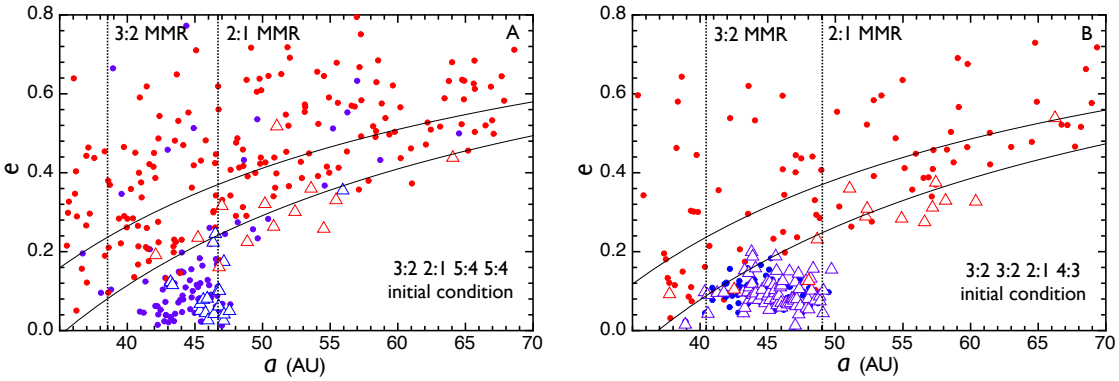


Figure 6.2: Eccentricity distribution of the remnant planetesimal disk. Red dots represent objects that have been dynamically emplaced, while the blue dots depict the locally formed cold classical belt at $t = 50$Myr. The red and blue triangles show objects whose orbits are stable on a 500Myr timescale. The scattered disk is shown with two solid curves and Neptune's exterior 3:2 and 2:1 MMR's are labeled with dashed lines. In the simulation presented in panel A of Figure (1), the cold belt suffers numerous encounters with the ejecting ice-giant, yielding considerable orbital excitation. Moreover, the inner cold belt is further dynamically depleted over 500Myr of evolution. In the simulation presented in panel B of Figure (1), there is only a single close encounter between the cold belt and the ejecting ice-giant, yielding a dynamically cold orbital structure.

bers, 1999) and spanned 50Myr[1]. The calculations were performed on Caltech's *PANGU* super-computer.

After their completion, simulations that were deemed successful were reintegrated with the use of tracer simulations (see Levison et al. (2008); Batygin et al. (2011)), to address the dynamical evolution of a locally formed population of KBOs. In particular, each run was supplemented with an additional disk of mass-less particles that resided in the cold classical region of the Kuiper belt (i.e. between the final exterior 3:2 and 2:1 MMRs of Neptune).

6.4 Results

Out of the 810 integrations that were initially performed, 214 ($\sim 25\%$) cases featured an ejection of a single ice-giant, yielding a system composed of 4 planets. Perhaps unsurprisingly, in most cases the ejected planet is the ice-giant that neighbors Saturn. Of the 750 simulations that were performed following the elimination of initial conditions, 33 evolutions resulted in orbits reminiscent of the solar system. Specifically, we searched for solutions where the Saturn-Jupiter period ratio exceed 2, while the final semi-major axes of the ice-giants were within 3AU of their observed counterparts. No strong requirements were placed on the planetary eccentricities and inclinations.

It is noteworthy that evolving the giant planets onto solar system-like orbits is insufficient for a simulation to be deemed successful for indeed, there are additional constraints that must be satisfied. The first orbital constraint is the reproduction of the secular architecture of the system. The secular orbital angular momentum exchange (i.e. eccentricity evolution) of a planetary system containing N secondaries can be approximately represented as a superposition of N eigenmodes, each corresponding to a fundamental frequency of the system (see Murray and Dermott (1999); Morbidelli et al. (2009)). Physically, the maximum eccentricity that a given planet attains in its secular cycle is equal to the sum of all of its corresponding eigenmode amplitudes. In the context of the traditional Nice model, some difficulty has been noted in correctly reproducing the dynamical character of Jupiter's and Saturn's eccentricity evolution. Particularly, it has been shown that smooth passage of Jupiter and Saturn through the 2:1 MMR has a tendency to under-excite the g_5 eccentricity

[1]Note that here, we make not attempt to time the onset of instability with the late heavy bombardment.

eigenmode as well as their mutual inclinations (Morbidelli et al., 2009). This difficulty can be overcome sometimes by invoking a close-encounter between an ice-giant and the gas giants. Consequently, we have checked, using Fourier analysis (see for example Morbidelli et al. (2009)), the relative strength of the g_5 and g_6 eigenmodes in all simulations whose orbital end-state resembled the solar system. We did not restrict the success criteria of our simulations to include the correct reproduction of the mean eccentricities of the planets and in some cases, the mean final eccentricities (namely those of the ice-giants) exceeded their observed counter-parts by as much as a factor of ~ 2. This is, however, likely an artifact of the coarse representation of the planetesimal disk and the resulting dynamical friction, that we employed in our calculations and should not be viewed as a major drawback.

In total, we found ten cases (corresponding to eight different initial conditions) where their amplitudes are satisfactorily reproduced. Specifically, in these ten cases, the amplitude of Jupiter's g_5 eigenmode exceeds that of the g_6 eigenmode, while the amplitudes are roughly the same for Saturn. We made no attempt at quantitatively matching the pair's inclination eigenmodes, however their reproduction does not appear to be problematic (Morbidelli et al., 2009). We also examined the amplitudes of the secular eigenmodes of Uranus and Neptune. Generally it appears that the dynamical architecture of the ice-giants is set in an essentially random manner, depending on the particular encounter history. Consequently, we decided to not use ice-giant secular architecture as a distinctive property in our analysis.

Successful formation of the Kuiper belt is another important constraint of the Nice model. Levison et al. (2008) have shown that the excited populations of the Kuiper belt are

naturally emplaced from inside of ~ 35AU during the instability. Given that this aspect of the planetary evolution is not particularly different between the 4-planet and the 5-planet scenarios, there is little reason to speculate that the dynamical pathway for formation of the resonant, scattered, and hot classical populations of the Kuiper belt will be inhibited. The same is likely to be true for chaotic capture of Trojan asteroids. The cold classical population of the Kuiper Belt, however, is a different story.

A series of observational dissimilarities between the cold classical population and the rest of the Kuiper belt (e.g. uniquely red colors (Trujillo & Brown, 2002; Lykawka & Mukai, 2005); strongly enhanced wide binary fraction (Stephens & Noll, 2006; Parker & Kavelaars, 2010)) suggest that the cold classicals formed in situ, and maintained dynamical coherence despite Neptune's temporary acquisition of high eccentricity and inclination (a characteristic orbital feature of the cold classical population is inclination that does not exceed $\sim 5^\circ$ (Brown, 2001)). In a recent study, Batygin et al. (2011) showed that local formation of the cold classicals is fully consistent with an instability-driven evolution of the planets, given favorable conditions during the instability. Particularly, Batygin et al. (2011) required the apsidal precession and nodal recession rates of Neptune to be comparatively fast to prevent secular excitation of the cold classical orbits, in addition to a sufficiently small apohelion distance of Neptune, to avoid orbital excitation due to close encounters.

In a 5-planet scenario, the retention of unexcited orbits of the cold population can be jeopardized by the ejecting planet. This possibility served as a premise for recalculation of the dynamical evolution of test particles in the cold classical region, with the aid of tracer simulations. We found that in only three of the ten simulations where the secular

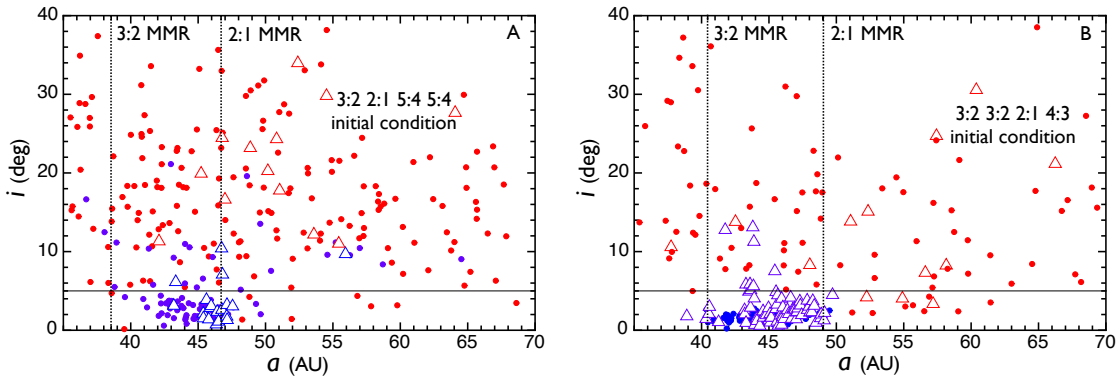


Figure 6.3: Inclination distribution of the remnant planetesimal disk. Red dots represent objects that have been dynamically emplaced, while the blue dots depict the locally formed cold classical belt at $t = 50$Myr. The red and blue triangles show objects whose orbits are stable on a 500Myr timescale. The tentative $i = 5°$ boundary between the cold and hot classical belts is shown with a solid line, while Neptune's exterior 3:2 and 2:1 MMR's are labeled with dashed lines. In the simulation presented in panel A of Figure (1), the cold belt suffers numerous encounters with the ejecting ice-giant, yielding considerable orbital excitation. Moreover, the inner cold belt is further dynamically depleted over 500Myr of evolution. In the simulation presented in panel B of Figure (1), there is only a single close encounter between the cold belt and the ejecting ice-giant, yielding a dynamically cold orbital structure.

eigenmodes were successfully reproduced, a primordial cold classical population of the Kuiper belt also retained unexcited orbits. Consequently, it appears that the retention of an unexcited cold belt is not the norm of the 5-planet scenario. This is not surprising, since dynamical excitation can only be avoided if ejection and close encounter timescales are sufficiently short. From a quantitative point of view, our simulations imply that dynamically cold orbits can only be sustained if the ejecting ice-giant spends $\sim 10^4$ years or less crossing the classical Kuiper belt region. As a result, we can expect that the incorporation of yet additional ice-giants into the primordial solar system (as suggested by Goldreich et al. (2004)) will further diminish the chances of reproducing the cold classical Kuiper belt, since all ejecting planets would have to do so very rapidly.

Two of the three successful simulations discussed above are presented in Figures (1-

3). Specifically, Figure (1) shows the orbital evolutions of the runs, while Figures (2) and (3) show the eccentricity and inclination distributions of planetesimals in the Kuiper belt region respectively. The red and blue dots, shown in Figures (2) and (3) depict the orbital distribution of emplaced and local planetesimal populations respectively at $t = 50$Myr. The red and blue triangles depict the planetesimals whose orbits remain stable at $t = 500$Myr.

The starting multi-resonant initial condition of the simulation presented in panels (A) of the figures is one where Saturn and the first ice-giant are locked in a 2:1 MMR, while both pairs of ice-giants are locked in 5:4 MMRs. The planetesimal disk in this simulation was comprised of $N = 3000$ particles and contained a total of $26 M_{\oplus}$. In this evolution, the local population of test particles suffers numerous short close encounters with the ejecting ice giant, yielding a more excited and depleted cold classical population, compared to that of run (B). Note also that at $t = 50$Myr, the ice-giant eccentricities are considerably greater than that of Uranus and Neptune. These high eccentricities do not get damped away by dynamical friction in the following 500Myr of dormant evolution. Consequently, in this simulation, the inner edge of the cold belt gets dynamically depleted over the following 500Myr. It is furthermore noteworthy that another simulation that originated from the same initial condition reproduced the eigenmodes of the system correctly, although the primordial cold Kuiper belt in this integration was entirely destroyed by close encounters.

The starting multi-resonant initial condition of the simulation presented in panels (B) of the figures is one where Saturn and the first ice-giant are locked in a 3:2 MMR, the inner pair of ice-giants is locked in a 2:1 MMR, and the outer pair of the ice-giants is locked in a 4:3 MMR. In this simulation, the disk consisted of $N = 1000$ particles and

had a cumulative mass of $42M_{\oplus}$. Incidentally, the frequency spectrum of the eccentricity vectors of Jupiter and Saturn produced in this simulation, matches that of the real Jupiter and Saturn exceptionally well, signaling a nearly ideal reproduction of the secular eigenmodes. Particularly, the simulation yields (in the notation of (Murray and Dermott, 1999)) $e_{55}^{(sim)}/e_{56}^{(sim)} = 2.28$ and $e_{65}^{(sim)}/e_{66}^{(sim)} = 0.51$ where as the solar system is characterized by $e_{55}/e_{56} = 2.81$ and $e_{65}/e_{66} = 0.68$. The scaling of the eigen-vectors are also well reproduced: $e_{55}^{sim} = 0.0465, e_{66}^{sim} = 0.067$ while the solar system has $e_{55} = 0.0442, e_{66} = 0.0482$ (Morbidelli et al., 2009). In this simulation, the cold Kuiper belt suffers only a single short encounter with the escaping ice-giant, allowing for the orbits (inclinations in particular) to remain dynamically cold.

6.5 Discussion

In this paper, we have presented a successful realization of the Nice model which starts out with 5 planets. The numerical experiments presented here explicitly show that such an evolution is plausible since the resulting 4-planet systems can closely resemble the solar system. Particularly, in both simulations presented here, the secular architecture of the outer solar system is well reproduced. Furthermore the demonstrated survival of a local, primordial cold classical Kuiper belt suggests that all constraints that can be matched with a 4-planet model can also be matched with a 5-planet model to an equal degree of satisfaction.

It is noteworthy that ejection was not always necessary in our simulations to generate a 4 planet system. In a handful of runs (one of which successfully reproduced the secular eigenmodes, but not the cold Kuiper belt), one of the ice-giants ended up merging with

Saturn. In principle, such a scenario may help explain Saturn's enhanced metallicity in comparison with Jupiter. Although, here again the explanation is not unique (see Stevenson (1982) and the references therein).

In a traditional realization of the Nice model, the rate of successful reproduction of the secular eigenmodes is rather low i.e. $\sim 10\%$ of the integrations for a favorable initial condition (Batygin & Brown, 2010). This is in part because an ice-giant/gas giant encounter often leads to an ejection of the ice-giant, leaving behind only three planets. Thus, the need for an ice-giant/gas-giant encounter in the orbital history of the solar system is in itself motivation for a 5-planet model.

The statistics of simulations presented in this work suggest that a 5-planet model is neither more nor less advantageous. Recall that the probability of ending up with only 4 planets is $214/810 \sim 25\%$. The probability of reproducing the secular eigenmodes of Jupiter and Saturn is $10/750 \sim 1.5\%$. Naively, this yields an overall probability of success of only $\sim 0.4\%$. However, it is important to keep in mind that the characteristic outcomes are generally dependent on initial conditions[2] and runs that originated from the initial condition presented in panels (A) correctly reproduced the secular eigenmodes in 2 out of 30 simulations. This statistic is similar to the 4-planet model. Finally, the $1/30$ probability of also retaining an unexcited cold classical Kuiper belt puts the 5-planet model and the 4-planet model on equal footing in terms of success rate (Batygin et al., 2011). That said, it is important to note that this success rate is only characteristic of the particular 5-planet model that we have constructed. In other words, it is likely that if one allows the

[2]Interestingly, we do not observe any correlation between the degree to which a given simulation is successful and disk mass.

mass of the ejected planet to also be a variable parameter, tuning of the initial state may in principle lead to a more frequent reproduction of the solar system.

The results presented in this work imply that the solar system is one of many possible outcomes of dynamical evolution, and can originate from many possible initial conditions. As a result, the possibility of having an extra planet initially present in the system, yet its ejection leaving no observable signature erases any hope for construction of a deterministic model for solar system evolution. The forward process-like nature of the Nice model is not surprising, given that the solar system exhibits large-scale chaos, characterized by Lypunov times that are comparable to orbital timescales, during the instability. Moreover, the similarity between the orbital architectures of simulations whose outcomes were deemed unsuccessful in this work and those of extra-solar planetary systems further confirms that planet-planet scattering is likely to be the physical process responsible for shaping the orbital distribution of planets (Jurić and Tremaine, 2008). Consequently, we conclude that an instability-driven dynamical history remains a sensible choice as a baseline scenario for solar system's early dynamical evolution.

Acknowledgments We thank Alessandro Morbidelli, Hal Levison, David Nesvorny and Peter Goldreich for useful conversations. We thank Naveed Near-Ansari for operational help with the *PANGU* supercomputer. K. Batygin acknowledges supported from NASA's NESSF graduate fellowship.

Bibliography

Batygin, K., & Brown, M. E. 2010, The Astrophysical Journal, 716, 1323

Batygin, K., Brown, M. E., & Fraser, W. C. 2011, The Astrophysical Journal, in press

Brown, M. E. 2001, Icarus, 151, 190

Chambers, J. E. 1999, Monthly Notices of Royal Astronomical Society, 304, 793

Jewitt, D., & Luu, J. 1993, Nature, 362, 730

Jurić, M., & Tremaine, S. 2008, The Astrophysical Journal, 686, 603

Goldreich, P., Lithwick, Y., & Sari, R. 2004, The Astrophysical Journal, 614, 497

Gomes, R., Levison, H. F., Tsiganis, K., & Morbidelli, A. 2005, Nature, 435, 466

Lee, M. H., & Peale, S. J. 2002, The Astrophysical Journal, 567, 596

Levison, H. F., & Morbidelli, A. 2007, Icarus, 189, 196

Levison, H. F., Morbidelli, A., Vanlaerhoven, C., Gomes, R., & Tsiganis, K. 2008, Icarus, 196, 258

Lykawka, P. S., & Mukai, T. 2005, Earth Moon and Planets, 97, 107

Masset, F., & Snellgrove, M. 2001, Monthly Notices of Royal Astronomical Society, 320, L55

Mayor, M., & Queloz, D. 1995, Nature, 378, 355

Morbidelli, A., Levison, H. F., Tsiganis, K., & Gomes, R. 2005, Nature, 435, 462

Morbidelli, A., & Crida, A. 2007, Icarus, 191, 158

Morbidelli, A., Tsiganis, K., Crida, A., Levison, H. F., & Gomes, R. 2007, The Astronomical Journal, 134, 1790

Morbidelli, A., Brasser, R., Tsiganis, K., Gomes, R., & Levison, H. F. 2009, Astronomy & Astrophysics, 507, 1041

Murray, C. D., & Dermott, S. F. 1999, Solar system dynamics by Murray, C. D., 1999

Nesvorný, D., Vokrouhlický, D., & Morbidelli, A. 2007, The Astronomical Journal, 133, 1962

Pierens, A., & Nelson, R. P. 2008, Astronomy & Astrophysics, 482, 333

Parker, A. H., & Kavelaars, J. J. 2010, The Astrophysical Journal Letters, 722, L204

Tsiganis, K., Gomes, R., Morbidelli, A., & Levison, H. F. 2005, Nature, 435, 459

Trujillo, C. A., & Brown, M. E. 2002, The Astrophysical Journal Letters, 566, L125

Stephens, D. C., & Noll, K. S. 2006, The Astronomical Journal, 131, 1142

Stevenson, D. J. 1982, Planetary & Space Science, 30, 755

Walsh, K. J., Morbidelli, A., Raymond, S. N., O'Brien, D. P., & Mandell, A. M. 2011, Nature, 475, 206

Chapter 7

A Quasi-Stationary Solution to Gliese 436b's Eccentricity

Originally published as:
Batygin, K., Laughlin, G., Meschiari, S., et al. 2009, The Astrophysical Journal, 699, 23

7.1 Abstract

We investigate the possibility that the large orbital eccentricity of the transiting Neptune-mass planet Gliese 436b is maintained in the face of tidal dissipation by a second planet in the system. We find that the currently observed configuration can be understood if Gliese 436b and a putative companion have evolved to a quasi-stationary fixed point in which the planets' orbital apses are co-linear and in which secular variations in the orbital eccentricities of the two planets have been almost entirely damped out. In our picture, the two planets are currently experiencing a long-period of gradual orbital circularization. Specifically, if Gliese 436b has a tidal $Q \sim 300,000$, similar to both the Jovian Q and to the upper limit for the Neptunian Q, then this circularization timescale can be of order $\tau \sim 8\,\mathrm{Gyr}$ given the presence of a favorably situated perturber. We adopt an octopole-order secular theory based on a Legendre expansion in the semi-major axis ratio a_1/a_2 to delineate well-defined

regions of ($P_{\rm c}$, $M_{\rm c}$, $e_{\rm c}$) space that can be occupied by a perturbing companion. This description includes the leading-order effects of General Relativity, and retains accuracy for perturbing companion planets that have high eccentricity. We incorporate the evolutionary effect of tidal dissipation into our secular model of the system, and solve the resulting initial value problems for a large sample of the allowed configurations. We find a locus of apsidally aligned configurations that are (1) consistent with the currently published radial velocity data, (2) consistent with the current lack of observed transit timing variations, (3) subject to rough constraint on dynamical stability, and which (4) have damping time scales consistent with the current multi-Gyr age of the star. We then polish the stationary configurations derived from secular theory with full numerical integrations, and compute the transit timing variations and radial velocity half-amplitudes induced by the resulting configurations. We present our results in the form of candidate companion planets to Gliese 436b. For these candidates, radial velocity half-amplitudes, K_c, are of order $3\,{\rm ms}^{-1}$, and the maximum amplitude of orbit-to-orbit transit timing variations are of order $\Delta t = 1\,{\rm s}$ to $\Delta t = 5\,{\rm s}$. For the particular example case of a perturber with orbital period, $P_c = 40\,{\rm d}$, mass, $M_c = 8.5\,M_\oplus$, and eccentricity, $e_c = 0.58$, we confirm our semi-analytic calculations with a full numerical 3-body integration of the orbital decay that includes tidal damping and spin evolution. Additionally, we discuss the possibility of many-perturber stationary configurations, utilizing modified Laplace-Lagrange secular theory. We then perform a proof-of-concept tidally dissipated numerical integration with 3 planets, which shows the system approaching a triply circular state.

7.2 Introduction

In the years since the discovery of the planetary companion to 51 Peg (Mayor & Queloz 1995), considerably *fewer* than 1,000 additional extrasolar planets have been detected.[1] Because the aggregate of known planets is still quite limited, the discovery and characterization of additional planets orbiting specific individual stars, remains a topic of considerable interest.

The detection of Gliese 436b by Butler et al. (2004) marked the first discovery of a Neptune mass planet since LeVerrier's discovery of Neptune itself in 1846. In May, 2007, Gliese 436b was discovered to be observable in transit (Gillon et al. 2007) which has sparked an intense interest in this remarkable planet. The transits resolve the $\sin(i)$ degeneracy, and fix the mass of the planet at $m_b = 23.17 \pm 0.79\, m_\oplus$ (Torres 2007). The period of the planet is $P = 2.643904 \pm .000005$ days (Bean et al. 2008) and it orbits an $M_\star = 0.452^{+0.014}_{-0.012}\, M_\odot$ red dwarf primary (Torres 2007). To date, no timing variations have been observed for the transits of Gliese 436b, with strict orbital periodicity maintained to a level of at least ~ 7 seconds (Pont et al. 2008).

The radius, $R_b = 4.22^{+0.09}_{-0.10}\, R_\oplus$ (Torres 2007) of Gliese 436b indicates that it necessarily contains several earth masses of hydrogen and helium in its outer layers. It probably resembles Neptune in overall composition (see, e.g. the models of Fortney et al. 2007). Measurement of the secondary eclipse depth by the Spitzer space telescope (Deming et al. 2007, Demory et al. 2007) indicates that the planet has an $8\,\mu m$ brightness temperature of $T = 712 \pm 36\,\mathrm{K}$. If we adopt $T_{\rm eff} = 3350\,\mathrm{K}$ for Gliese 436A, take a zero albedo for

[1] For an up-to-date listing of the current galactic planetary census, see http://www.exoplanet.eu

the planet, and assume a uniform re-radiation of the orbit-averaged received flux from the entire planetary surface, we obtain a planetary effective temperature $T_{eq} = 642\,\mathrm{K}$. The somewhat higher temperature implied by the secondary eclipse depth could arise from inefficient transport of heat to the night side of the planet, from a non-blackbody planetary emission spectrum, from tidal luminosity, or from a combination of the three. The possibility that tidal luminosity is being observed is prompted by the orbital phase, $\phi = 0.587$, of the secondary eclipse, which confirms that the orbital eccentricity is alarmingly high (with a best-fit value of $e = 0.150 \pm 0.012$; Deming et al. 2007). Indeed, the planet and its primary bring to mind an oversize alien version of Jupiter and Io.

The short orbital period and large eccentricity of Gliese 436b suggest that tidal circularization might be highly effective. To second order in eccentricity, the tidal luminosity of a spin-synchronous planet (Peale & Cassen 1978, Mardling 2007) is given by

$$\frac{dE}{dt} = \frac{21}{2}\frac{k_2}{Q}\frac{GM_\star^2 n R_b^5 e^2}{a^6}\,, \tag{7.1}$$

where k_2 is the planetary tidal Love number, a is the semi-major axis, n is the orbital mean motion, and Q is the planet's effective tidal dissipation parameter. Note that Gliese 436b cannot be fully spin synchronized, and because it presumably has no permanent quadropole moment, it should not reside in spin-orbit resonance. The pseudo-synchronization theory of Hut (1981; see also Goldreich & Peale 1966) suggests $P_{\mathrm{spin}} = 2.32\mathrm{d}$, leading to a 19-day synodic period for the planet b. The analysis of Levrard et al. (2007), furthermore, indicates that this spin asynchronicity of the planet will cause the tidal luminosity to exceed that given by the above formula by a small amount.

To date, several authors have investigated the Gliese 436 system with an eye toward obtaining a convincing explanation for planet b's large eccentricity. To establish context, it is useful to review these contributions.

Solution 1: A large Q

Jackson et al (2008a) have performed tidally-dissipated single-planet integrations backwards in time, and have shown that $Q_b/k_{2b} \approx 10^{6.6}$ and $Q_\star/k_{2\star} \approx 10^{5.5}$ can explain the high eccentricity of Gliese 436b. However, are such Q-values realistic? The resonant configuration of the Jovian satellites implies $10^5 < Q_{\rm J} < 2 \times 10^6$, with the preferred value closer to the lower limit (Goldreich & Soter 1966, Peale & Greenberg 1980). Meyer & Wisdom (2008) argue that $Q_{\rm S} < 10^5$ whereas the circularization boundary for extrasolar planets (Wu 2003) suggests typical Q values of order 3×10^5, allowing for significant variation from planet to planet in addition to significant overall uncertainty. Uranus and Neptune, which more closely resemble Gliese 436b in mass, composition, and radius, are a further factor of approximately ten times more dissipative. Tittemore & Wisdom (1989) have demonstrated that $Q_{\rm U} \lesssim 3.9 \times 10^4$, whereas in the case of Neptune, Banfield & Murray (1992) find $1.2 \times 10^4 < Q_{\rm N} < 3.3 \times 10^5$. Recent calculations of Zhang & Hamilton (2008) suggest an even lower estimate: $9 \times 10^3 < Q_{\rm N} < 3.6 \times 10^4$.

A theory for the origin of planetary Q values (not to speak of a definitive understanding of dynamical tides) remains elusive, and so there is no a-priori reason to reject the hypothesis that Gliese 436b has a remarkably small internal dissipation. It should be noted, however, that for the particular case of Gliese 436b, Q would need to be even higher than Jackson et al. (2008a)'s estimate of $Q_b/k_{2b} \approx 10^{6.6}$, given the likely breakdown of second-

order equilibrium tidal theory at the very high initial eccentricities that they considered (Wisdom, 2008).

Solution 2: A resonant perturber

A natural solution to the high eccentricity problem is to invoke perturbations from an as-yet undetected second planet in the system. Ribas et al. (2008) have suggested that the perturbing body lies in a 2:1 mean motion resonance with the primary planet. By constructing a 2-planet fit to the radial velocity data of Maness et al. (2007), they produced a specific orbital model for the second planet, and predicted that changes in the eclipse depth and duration (arising from precession of Gliese 436b driven by the interaction with c should be observable.

Resonant perturbations contribute to the disturbing function with terms that contain mean longitudes, λ, and so they are generally capable of producing far larger transit timing variations than secular perturbations. Indeed, most discussions of the transit timing technique have focused on perturbing planets in low-order mean motion resonances (e.g. Agol et al. 2005, Holman & Murray 2005). In the context of Gliese 436, the Ribas et al. (2008) model generates transit timing variations measured in minutes, making their hypothesis eminently testable.

The $2.64\,\mathrm{d}$ period of planet b allows transit measurements to accumulate rapidly, and during the time that has elapsed since the Gillon et al. (2007) discovery, a number of high-precision observations of transit midpoint times have become publicly available. Alonso et al. (2008) analyzed the archival Spitzer photometry obtained via "target of opportunity" observations in June 2007, and derived $T_c = 2454280.78191 \pm 0.00028\,\mathrm{HJD}$ (24 second

accuracy). The same authors also analyzed their own ground-based observations of the March 8, 2008 transit, and derived $T_c = 2454534.59584 \pm 0.00015\,\mathrm{HJD}$ (13 second accuracy). By comparing these transit times with Gillon et al (2007)'s discovery ephermeris of $T_c = 2454222.616 \pm 0.001\mathrm{HJD}$ (86 second accuracy), Alonso et al. (2008) then determined that there is only a $\sim 0.5\%$ probability that a 5 $M_\oplus$ perturber in an exterior 2:1 mean motion resonance would have produced transit timing variations (TTV) smaller than those implied by the three midpoint times in their analysis. Recent work by Pont et al. (2008) has pushed the threshold for TTVs below 7 seconds, ruling out perturbers in 2:1 resonance down to lunar mass. Mardling (2008), furthermore, shows that semi-major axis decay produced by tidal dissipation will cause any 2:1 resonance between GL436b and a hypothetical companion to be short-lived. In particular, a perturber with the properties proposed by Ribas et al. (2008) passes through a 2:1 resonance on a timescale of $\tau_p = 0.3(Q_b/10^5)$ Gyr.

Solution 3: A secular perturber

In a recent paper, Tong and Zhou (2008) explored Gliese 436b's eccentricity modulation by both secular and resonant companions. Their approach adopted an octopole-order secular theory (see below) to identify hypothetical perturbers that can excite planet b's eccentricity (in the course of a single secular cycle) from $e_b = 0$ to $e_b = 0.16$, while simultaneously requiring that the system as a whole remain dynamically stable and that the perturber induce a radial velocity half amplitude less than the current threshold of detection (which they took to be $K < 3\,\mathrm{m\,s^{-1}}$). Interestingly, they located a range of candidate perturbers that satisfied the constraints. Tong & Zhou (2008) found, however, that when the effects of tidal dissipation were incorporated and the hypothetical two-planet systems

were integrated forward in time, all eccentricity modulation was damped out and the orbits quickly circularized. As a part of their survey, Tong & Zhou (2008) also specifically exclude the class of perturbing companion planets suggested by Maness et al. (2007), noting, for example that a putative perturber with $M_c = 0.12\,M_{\rm Jup}$, $P = 14\,{\rm yr}$, and $e_c = 0.6$ is capable of exciting an eccentricity, $e_{b_{max}}$, for Gl 436b to only 3×10^{-6}.

The aggresively effective orbital circularization observed by Tong & Zhou (2008) led those authors to conclude that a companion planet (exerting either resonant or secular perturbations) that is consistent with the current corpus of Doppler velocity observations "can not excite and maintain the significant eccentricity" of Gl 436b. They argue that in order to adequately explain Gliese 436b's non-circular orbit, its tidal quality factor must be of order $Q_b/k_{2b} \sim 6 \times 10^6$, a factor likely an order of magnitude larger than the values generally ascribed to gas giants (see however, Jackson, Barnes & Greenberg (2008b) who argue that the upper limits on Jupiter's Q may be significantly higher than generally believed, and see Matsumura, Takeda & Rasio (2008) who present evidence of high Q for the aggregate of known extrasolar planets.)

In any case, the non-circular orbit of Gliese 436b constitutes a remarkable puzzle. The solution likely has interesting implications for our understanding of planetary structure, formation, and dynamics. The problem is rendered even more worthwhile by the fact that improved spectroscopic and photometric observations can be used to test the predictions inherent in a particular solution, as has been demonstrated already with the Ribas et al. (2008) hypothesis. This paper reports our attempt at a *dynamical* solution using currently published observational constraints. We investigate an alternate hypothesis that while the

perturbations are secular in nature, the eccentricity is not modulated, and the orbits of both planets are quasi-stationary. We present a semi-analytical survey which demonstrates that a range of well-defined apsidally aligned perturbers are individually capable of prolonging the effective circularization timescale of Gliese 436 b by more than a factor of five. We use numerical integrations to correct for planet-planet dynamics not captured by our secular theory, and present our final locus of predictions for the perturber properties in the form of a table. We confirm the long-term evolution for a select test case with a fully numerical scheme, and we discuss the possibilities for near-term observational confirmation. Finally, we discuss the possibility of a multiple-perturber scenario in the context of a modified Laplace-Lagrange secular theory and perform one proof-of-concept numerical integration.

7.3 The Dynamical State of the System

We first investigate the current dynamical state of a two-planet system. Throughout out analysis, we assume that the inclinations of perturbing planets with respect to Gliese 436b's orbit are dynamically negligible. This assumption may turn out to be overly restrictive, but we note that if the orbital angular momentum vector of Gliese 436b is found to be in near-alignment with the rotation axis of the parent star, then the possibility that the system is nearly co-planar is strengthened. To date, however, there have been no reported observations of the Rossiter-McLaughlin effect (see e.g. Winn et al. 2005) during transits of Gliese 436b.

In the absence of mean motion resonances, the dynamics of a stable multi-planet system are dominated by secular interactions, with tidal forces also playing a significant role. In

this scenario, any initial secular variations of the planets' orbital elements are damped out over ~ 3 tidal circularization timescales for the inner planet, given by

$$\tau_b = \left(\frac{GM_\star m_b}{a_b}\right)\left(\frac{e_b^2}{\dot{E}}\right), \tag{7.2}$$

and the system approaches a quasi-stationary state. Once the fixed point is reached, the system begins to evolve exponentially towards a doubly circular state over a timescale that is considerably longer than than the circularization timescale (Mardling 2007).

Correcting for the leading-order effects of general relativity (GR), which are crucial to this system, the orbit-averaged rates of change of eccentricities and longitudes of perihelia of the coplanar system are given by (Mardling & Lin 2002; Mardling 2007)

$$\frac{de_b}{dt} = -\frac{15}{16} n_b e_c \left(\frac{m_c}{M_\star}\right)\left(\frac{a_b}{a_c}\right)^4 \frac{\sin(\varpi_b - \varpi_c)}{(1-e_c^2)^{5/2}}, \tag{7.3}$$

$$\frac{de_c}{dt} = \frac{15}{16} n_c e_b \left(\frac{m_b}{M_\star}\right)\left(\frac{a_b}{a_c}\right)^3 \frac{\sin(\varpi_b - \varpi_c)}{(1-e_c^2)^{2}}, \tag{7.4}$$

$$\begin{aligned}\frac{d\varpi_b}{dt} &= \frac{3}{4} n_b \left(\frac{m_c}{M_\star}\right)\left(\frac{a_b}{a_c}\right)^3 \frac{1}{(1-e_c^2)^{3/2}}\left[1 - \frac{5}{4}\left(\frac{a_b}{a_c}\right)\left(\frac{e_c}{e_b}\right)\frac{\cos(\varpi_b - \varpi_c)}{1-e_c^2}\right] \\ &+ \frac{3n_b^3}{1-e_b^2}\left(\frac{a_b}{c}\right)^2,\end{aligned} \tag{7.5}$$

and

$$\frac{d\varpi_c}{dt} = \frac{3}{4} n_c \left(\frac{m_b}{M_\star}\right)\left(\frac{a_b}{a_c}\right)^2 \frac{1}{(1-e_c^2)^{2}}\left[1 - \frac{5}{4}\left(\frac{a_b}{a_c}\right)\left(\frac{e_b}{e_c}\right)\frac{(1+4e_c^2)}{(1-e_c^2)}\cos(\varpi_b - \varpi_c)\right], \tag{7.6}$$

where c is the speed of light, and ϖ is the longitude of perihelion. The above set of equations is derived from the evolution of specific angular momentum and Runge-Lenz vectors utilizing a Legendre-polynomial expansion in terms of semi-major axis ratios (a_b/a_c) to octopole order (see Mardling & Lin 2002 for the derivation), rather than the classical eccentricity expansion of the disturbing function (Laplace 1825, Brouwer & Clemence 1961). The model contains no constraints on the perturber's eccentricity, but it requires (a_b/a_c) to be small. From equations (3 - 6), it is easy to see that there exist two apsidal fixed points: $(\varpi_b - \varpi_c) = 0$ and $(\varpi_b - \varpi_c) = \pi$. Each is stable for its own combination of orbital parameters, and libration around either can occur. Consequently, the system will tidally evolve towards an apsidally aligned fixed point if

$$\frac{m_c}{m_b} > \frac{\sqrt{a_b/a_c}}{(1-e_c^2)(1+4(n_b a_b/c)^2(M_\star/m_c)(a_c/a_b)^3(1-e_c^2)^3}, \tag{7.7}$$

and towards an anti-aligned state otherwise (Mardling, 2007). Note that our model treats the star as a point mass, thereby neglecting its tidal and rotational deformation's contribution to the apsidal precession rates of the planetary orbits. While the rotational period of the star could not be determined from photometric observations (Butler et al 2004), if it is short, and the star's zonal gravitational harmonics are large, such terms could provide significant corrections to the equations presented here. Nonetheless, a fast rotation rate would be extremely uncharacteristic of the star's multi-Gyr age. An additional effect that we do not consider is precession arising from the quadropole moments of the planets b, and c. For ultra-short period planets with large radii, this effect can be important, but for Gliese 436b, it is small in comparison to the precession induced by the GR correction (see, e.g.

Ragozzine & Wolf 2008).

We can use our formulation to survey parameter space for regions where the perturbing planet might lie. While the age of the system is estimated to lie in the range of $\tau_{sys} \approx 1\text{Gyr} - 10\text{Gyr}$ (Torres 2007), it almost certainly exceeds the few circularization timescales needed for the system to reach a fixed point. As a result, in the following analysis, we assume that the system is currently quasi-stationary, allowing us to restrict our search to fixed-point configurations.

The following scheme was used to generate fixed point states of the system: the perturber's semi-major axis, eccentricity and mass were chosen randomly; equation (7) was invoked to determine whether the resulting fixed point is aligned or anti-aligned. By setting $d\varpi_b/dt = d\varpi_c/dt$, the eccentricity of Gliese 436b was determined. If the resulting eccentricity fell between 0.149 and 0.151, the configuration was kept. Subsequently, the generated system was tested against two constraints. First, to exclude grossly unstable systems, we required that the distance of closest approach of the two orbits was less than the radius of the larger Hill sphere, given by (Murray & Dermott, 1999)

$$R_H = a(1-e)\left(\frac{m}{3M_\star}\right)^{1/3}. \tag{7.8}$$

In practice, this criterion will admit some unstable systems; candidate systems that make relatively close approaches can be integrated numerically to verify stability.

A planet with given values for period, mass, and eccentricity produces a radial velocity half-amplitude

$$K = (\frac{2\pi G}{P_c})^{1/3}(\frac{M_{\rm c}\sin(i)}{(M_\star + M_{\rm c})^{2/3}})\frac{1}{1-e_c^2}\,. \tag{7.9}$$

Conseqently, we required that the radial velocity half-amplitude produced by the perturber be less than 5m/s (a relatively generous criterion, even for perturbers of high eccentricity). This criterion rids the parameter space of readily detectable planets. Whenever one of our randomly generated systems satisfied our search criteria, it was numerically integrated forward in time, using the secular ODEs presented above, with the effects of tidal dissipation accounted for by including additional terms for the eccentricity and semi-major axis damping that arise from tidal dissipation in Gliese 436b

$$\frac{de_b}{dt} = -e_b \frac{21\pi}{P_b} \frac{k_b}{Q_b} \frac{M_\star}{m_b} \left(\frac{R_b}{a_b} \right)^5 , \tag{7.10}$$

$$\frac{da_b}{dt} = -e_b^2 a_b \frac{42\pi}{P_b} \frac{k_b}{Q_b} \frac{M_\star}{m_b} \left(\frac{R_b}{a_b} \right)^5 . \tag{7.11}$$

For each trial system, we can thus evaluate the system's new circularization timescale by noting the time, t, when Gliese 436b's initial eccentricity has decreased by a factor of e. To speed up the integration process, an artificially low Q-value of 300 was used. For the current purposes, we are interested less in the circularization timescale itself, but rather in the factor, f, by which the circularization of b is increased when the perturber is added. Consequently, we divide the numerically computed timescale by that of a secular integration with only Gliese 436b present.

A total of 1625 model systems were generated and then integrated forward using the semi-analytic model. The results of this Monte-Carlo parameter survey are presented in Figures (1) through (4). Note that the perturbers at the anti-aligned fixed points (figures 3 and 4) generally have smaller periods than the perturbers in aligned configurations (figures

1 and 2). There is also a considerable difference in the factor by which the circularization time scale of planet b is increased. The anti-aligned perturbers provide an average increase factor of only 1.48 ± 0.36. The aligned perturbers on the other hand provide an average increase factor of 5.54 ± 0.26. The relatively high current eccentricity of GL 436b suggests a long circulization time, thus favoring the aligned perturbers. Within the aggregate of aligned perturbers, the damping timescale increase factor varies as well, peaking around $P_c \approx$ 37d, $e_c \approx 0.58$, $M_c \approx 8_{\oplus}$ with an increase factor of $f = 6.03$. A further attractive feature of the aligned configurations is that they form a very well-defined locus in (P_c, e_c, M_c) space. It is interesting to note that the region of parameter space occupied by viable perturbing companions is strongly influenced by the inclusion of general relativistic precession in Equation 5. With a purely Newtonian potential, apsidal alignment can only be maintained for perturbers with significantly longer orbital periods and significantly larger masses. This result is typical of many multiple-planet extrasolar systems, where one finds that the steady-state dynamical configuration has a delicate dependence on the inner planet precession rate (see, e.g. Adams & Laughlin 2006a,b)

7.4 Numerical Confirmation

The semi-analytical approach presented above is an effective method for conducting a parameter search for potential perturbing companions and the numerical integration of the secular equations with dissipation provides an efficient procedure for evaluating the resulting increase in orbital circularization time. The octopole order theory, however, is only approximate, and one would like to have a verification that does not incorporate the approx-

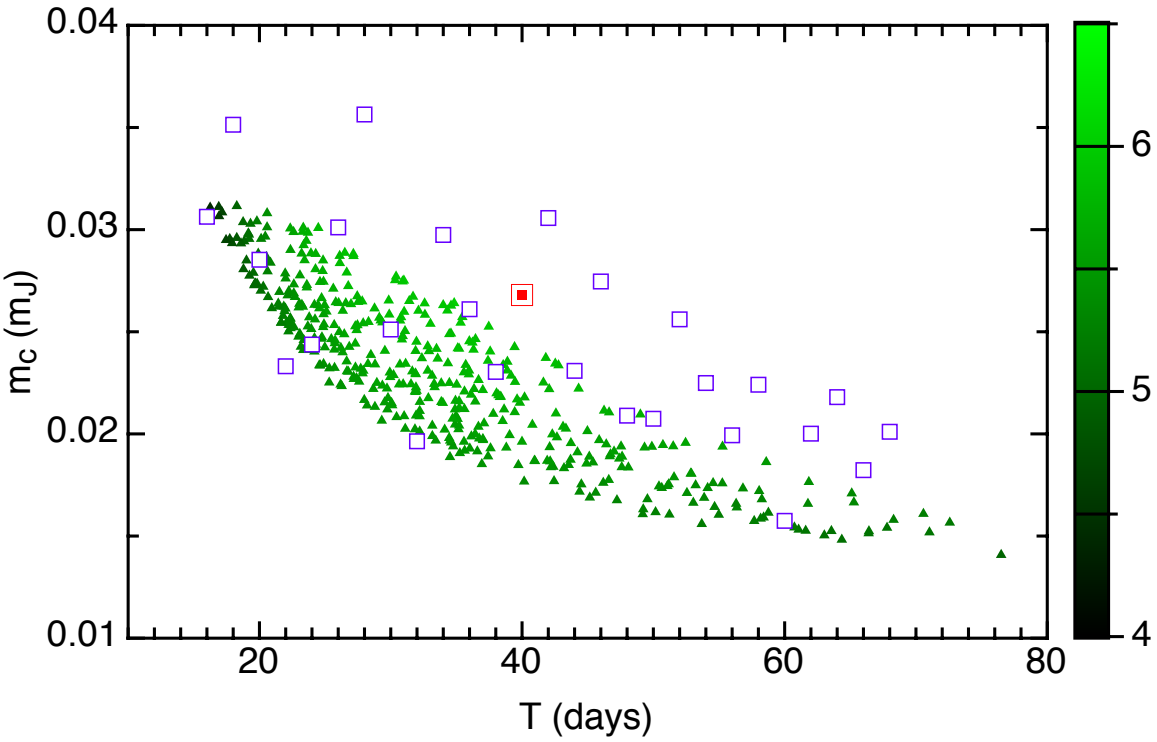


Figure 7.1: Distribution of masses and orbital periods for potential Gliese 436 c candidates occupying apsidally aligned fixed points relative to the known transiting planet Gliese 436b (*small triangles*). Color indicates the factor of increase in the orbital circularization time for Gliese 436b in comparison to the case where no perturber is present. The color scale runs from black (4 times increase in time scale) to light green (factor of 6.5 increase in circularization time). The aligned perturbers produce significant increases in orbital circularization time, and hence are thus favored as an explanation for the substantial orbital eccentricity of Gliese 436b. Aligned fixed-point models obtained using the full equations of motion are shown as open blue squares. The red dot signifies the configuration that was used for initial conditions of the numerical integration of orbital decay.

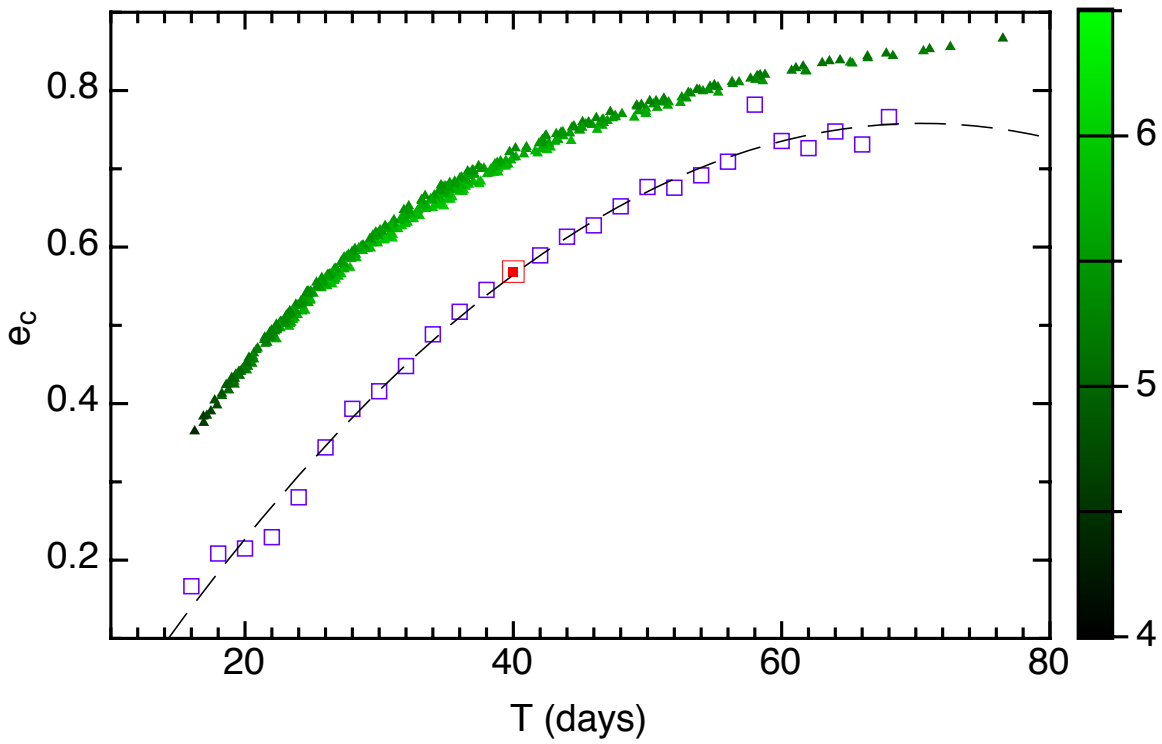


Figure 7.2: Distribution of eccentricities and orbital periods for potential Gliese 436 c candidates occupying apsidally aligned fixed points relative to the known transiting planet Gliese 436b (*small triangles*). Color indicates the factor of increase in the orbital circularization time for Gliese 436b in comparison to the case where no perturber is present. The color scale runs from black (4 times increase in time scale) to light green (factor of 6.5 increase in circularization time). Aligned fixed-point models determined using the full equations of motion are shown as open blue squares. The red dot signifies the configuration that was used for initial conditions of the numerical integration of orbital decay. The dashed line corresponds to a best-fit second-order polynomial function (see section 3).

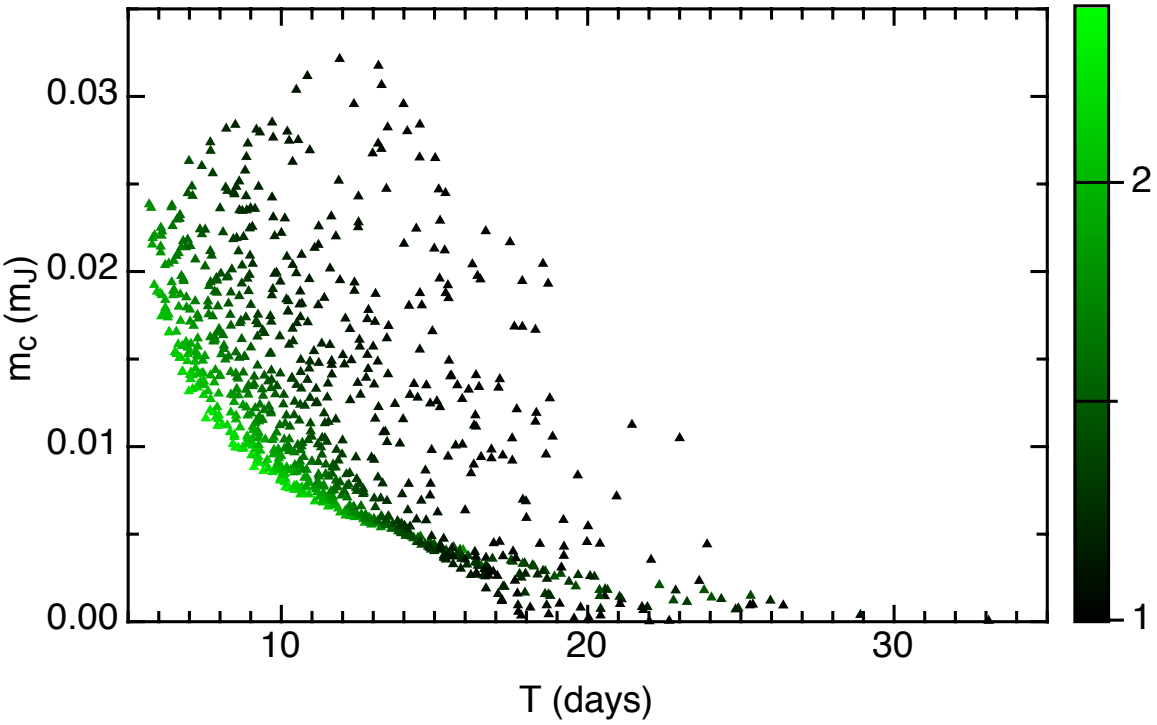


Figure 7.3: Distribution of masses and orbital periods for potential Gliese 436 c candidates occupying apsidally anti-aligned fixed points relative to the known transiting planet Gliese 436b (*small triangles*). Color indicates the factor of increase in the orbital circularization time for Gliese 436b in comparison to the case where no perturber is present. The color scale runs from black (no increase in time scale) to light green (factor of 3 increase in circularization time). The anti-aligned perturbers produce only modest increases in orbital circularization time, and hence are thus less favored to explain the substantial orbital eccentricity of Gliese 436b.

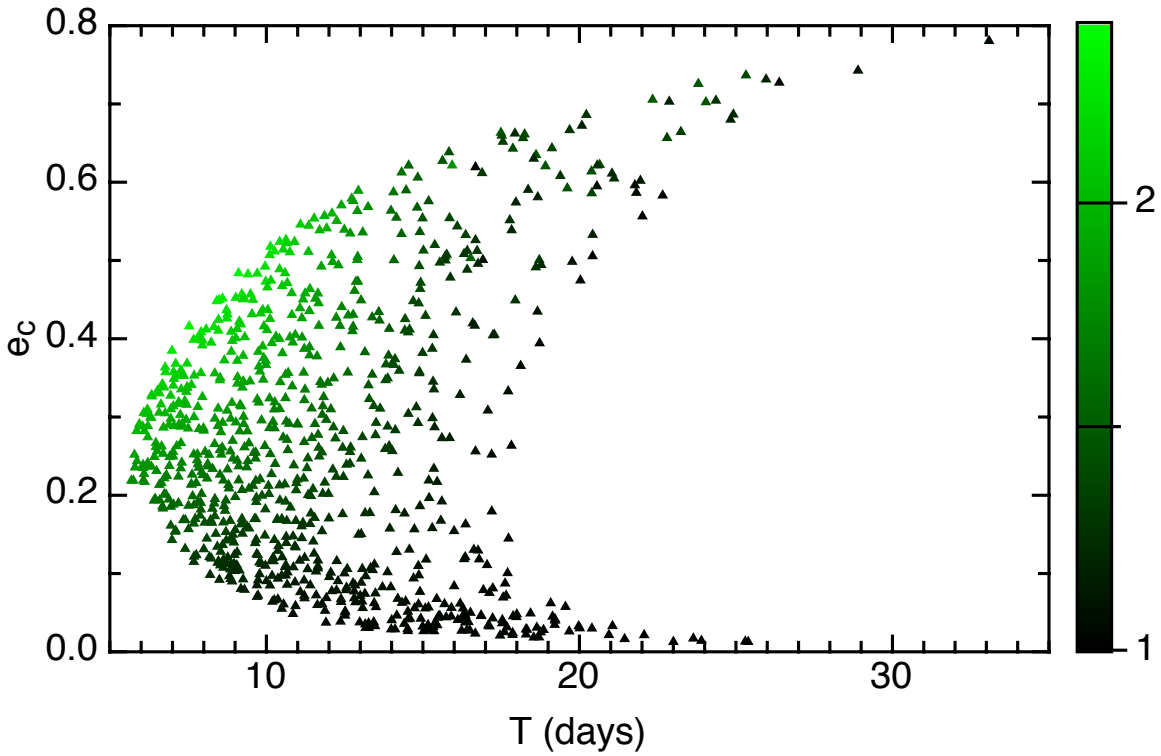


Figure 7.4: Distribution of eccentricities and orbital periods for potential Gliese 436 c candidates occupying apsidally anti-aligned fixed points relative to the known transiting planet Gliese 436b (*small triangles*). Color indicates the factor of increase in the orbital circularization time for Gliese 436b in comparison to the case where no perturber is present. The color scale runs from black (no increase in time scale) to light green (factor of 3 increase in circularization time).

imations. Indeed, to determine the exact fixed-point parameters of a hypothetical perturber, a fully numerical determination is much better suited. As a proof of concept, we explored twenty-seven representative cases, each containing an apsidally aligned perturber.

The determination of the fixed point parameters in the presence of the full equations of motion was performed by picking a perturber's period and mass, and subsequently performing many direct integrations while fine-tuning the perturber's eccentricity using a Newton-Raphson scheme until a combination of orbital elements that corresponds to a true fixed point was reached. The *Mercury6* integrator (Chambers, 1999) was used. The effects of general relativity were modeled by adding an extra term to the star's gravitational potential that takes the form

$$V_{GR} = -3 \left(\frac{G M_\star}{c\, r} \right)^2 , \tag{7.12}$$

where r is the distance from the planet to the star. The addition of the extra potential term yields the exact secular precession rate given by GR, while introducing a small error (of order $(m/M_\star)^2$)in the planet's mean motion (Nobili & Roxburgh, 1986). The so-called hybrid algorithm was used for these integrations. All of the integrations within the Newton-Raphson iteration were short-term, never exceeding 10^4 years.

Using this numerical scheme, we determined the fixed point eccentricities of the potential perturbing companions, spaced with periods in equal increments between $P_{lower} = 16\,\mathrm{d}$ and $P_{upper} = 68\,\mathrm{d}$. In Table 1, we list the resulting systems (which are also plotted as a sequence of blue squares on Figures 1 and 2). The perturber's eccentricities can be approximated by a second order polynomial function: $e_c \approx -0.277 + 2.94 \times 10^{-2} P_c - 2.05 \times 10^{-4} P_c^2$, where the period is measured in days (plotted as a dashed line on figure

2). It is clear that the inclusion of the full disturbing function produces some corrections to the octopole-order analytic solutions. Namely, secular theory tends to over-estimate the fixed-point eccentricities of the perturbers for this system.

Table 1 lists transit timing variations, $\Delta T_{\max}$ for the fixed point systems. The amplitude of the timing variations changes by a few seconds along this locus of models in (P_c, e_c, M_c) space. Much of this variation is due to the given configuration, residing close to, or in a high-order MMR. An extreme example is a $P_c = 58\,\mathrm{d}$ perturber system, where the planets reside in a 22:1 resonance. Due to the perturber's very high fixed-point eccentricity, transit timing variations reach almost a minute. Hence, if our theory is correct, a given measurement of Δt_{max} may provide an immediate hint to P_c, e_c, and M_c.

Our iteratively determined fixed-point models all appear to be indefinitely dynamically stable. By definition, at the fixed point, orbital parameter variations are reduced to the slow tidal circularization of both orbits, and so the planets never get into trouble. Interestingly, however, for eccentricities $e_c \gtrsim 0.8$, we find that the octopole-order fixed-point solutions, when integrated using the full equations of motion, are largely unstable. When given access to the full disturbing function, these configurations usually experience eccentricity variations and librations in $\varpi_c - \varpi_b$ large enough to generate orbit crossings. The true fixed point configurations thus represent islands of stability within a larger sea, yielding a potentially important clue to the initial formation and evolution of the Gliese 436 planetary system. Indeed, if our model is confirmed by observation, then the question of system formation will represent an interesting line of inquiry.

Finally, we carried out a full numerical demonstration of the circularization timescale

Table 7.1: Potential Fixed Point Perturbing Companions to Gliese 436b

TTV (s)	K (m/s)	P (days)	m (M_J)	e
2.46	4.39	16	3.06298 $\times 10^{-2}$	0.166647
1.66	4.92	18	3.5136 $\times 10^{-2}$	0.208449
1.12	3.87	20	2.85267 $\times 10^{-2}$	0.214894
0.8	3.09	22	2.33118 $\times 10^{-2}$	0.229319
0.11	3.22	24	2.43739 $\times 10^{-2}$	0.280283
1.46	4.05	26	3.01091 $\times 10^{-2}$	0.344026
1.73	4.88	28	3.56371 $\times 10^{-2}$	0.393356
1.13	3.43	30	2.51117 $\times 10^{-2}$	0.415769
0.88	2.72	32	1.96387 $\times 10^{-2}$	0.447974
2.27	4.24	34	2.974 $\times 10^{-2}$	0.488534
0.95	3.79	36	2.61039 $\times 10^{-2}$	0.517317
1.37	3.43	38	2.30358 $\times 10^{-2}$	0.545275
2.65	4.07	40	2.68 $\times 10^{-2}$	0.56866
4.19	4.73	42	3.05687 $\times 10^{-2}$	0.589384
2.35	3.68	44	2.30875 $\times 10^{-2}$	0.613239
2.59	4.44	46	2.74619 $\times 10^{-2}$	0.627729
2.77	3.51	48	2.08942 $\times 10^{-2}$	0.652055
9.72	3.65	50	2.07423 $\times 10^{-2}$	0.676771
2.77	4.44	52	2.56081 $\times 10^{-2}$	0.675824
2.48	4.01	54	2.24914 $\times 10^{-2}$	0.691682
4.35	3.68	56	1.99315 $\times 10^{-2}$	0.709095
49.59	5.23	58	2.2408 $\times 10^{-2}$	0.781977
2.88	3.08	60	1.57439 $\times 10^{-2}$	0.735554
6.41	3.76	62	2.00128 $\times 10^{-2}$	0.726411
3.54	4.34	64	2.18008 $\times 10^{-2}$	0.747703
5.93	3.40	66	1.82263 $\times 10^{-2}$	0.731106
4.15	4.19	68	2.01049 $\times 10^{-2}$	0.766115

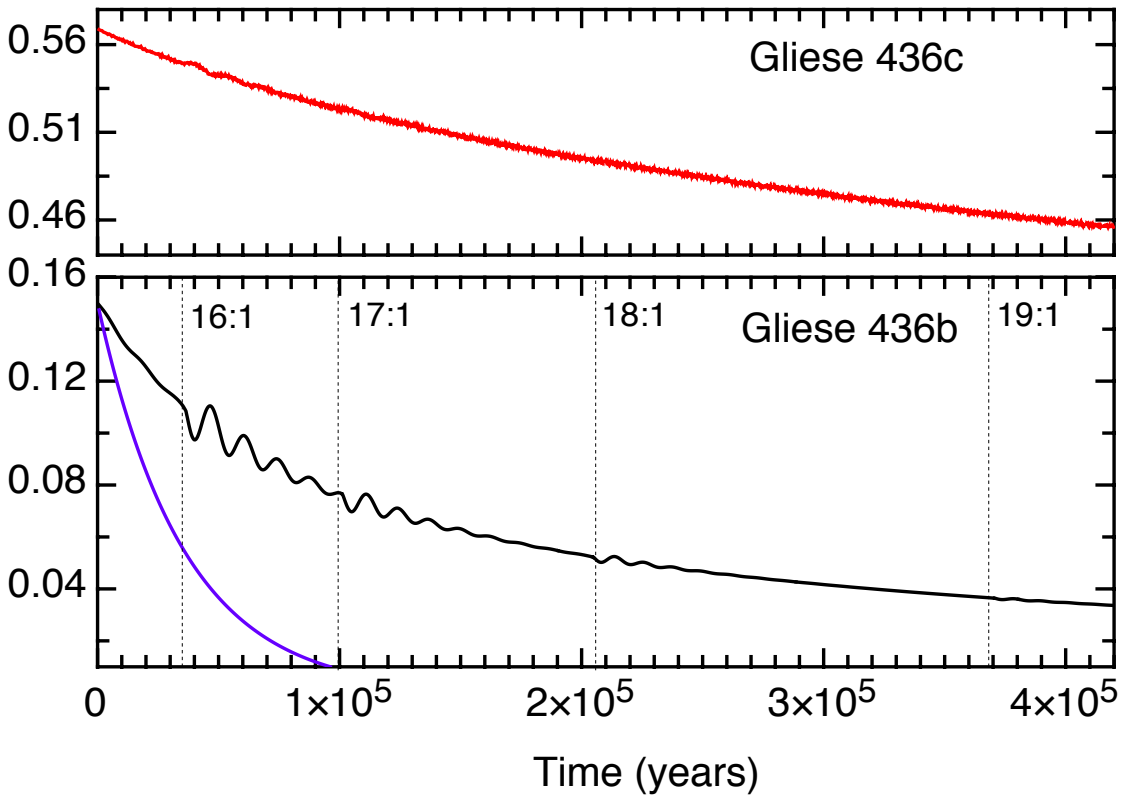


Figure 7.5: Numerically computed eccentricity evolution of the tidally dissipated Gliese 436 system, starting from a numerically determined aligned fixed point. The top plot corresponds to the eccentricity of a $P = 40\,\mathrm{d}$ perturber. The black curve in the bottom plot corresponds to the eccentricity evolution of Gliese 436b, in presence of the perturber. The blue curve in the bottom plot corresponds to the scenario where Gliese 436b is alone. For the integrations, $Q = 10$ was used.

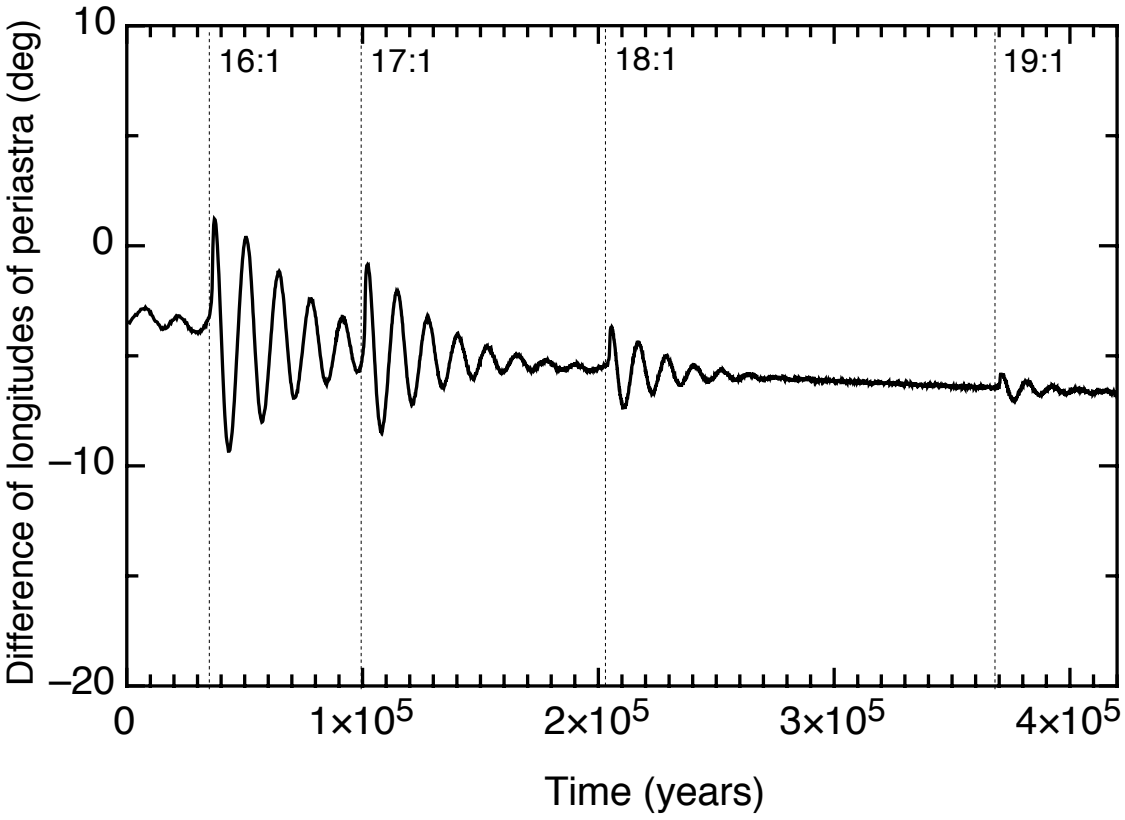


Figure 7.6: The difference between apsidal lines of Gliese 436b and c for the numerically computed tidally dissipated eccentricity evolution shown in Figure 5.

of the two-planet $P_c = 40\,\mathrm{d}$ system listed on the thirteenth line of Table 1. This system lies near the region where the secular ODEs exhibit the largest factors of increase, f. To carry out the full integrations, we used a self-consistent numerical scheme that in addition to gravitational dynamics, takes into account the effects of tidal damping and spin-evolution (Eggleton et. al 1998, Mardling & Lin 2002). In this simulation, the post-Newtonian gravitational force was computed in the same way as above. The tidal force takes the form

$$\mathbf{F}_T = -\left(\frac{6n(k_2/2)}{Q}\right)\left(\frac{M_{star}}{m}\right)\left(\frac{R}{a}\right)^5\left(\frac{a}{|r|}\right)^8 \times [3\,(\hat{\boldsymbol{r}}\cdot\dot{\boldsymbol{r}})\,\hat{\boldsymbol{r}} + (\hat{\boldsymbol{r}}\times\dot{\boldsymbol{r}} - r\boldsymbol{\Omega}_{\mathbf{b}})\times\hat{\boldsymbol{r}}] \quad (7.13)$$

and the spin evolution is governed by:

$$\dot{\boldsymbol{\Omega}} = \frac{mM_\star}{m+M_\star}\left(\frac{\boldsymbol{r}\times\mathbf{F}_T}{I}\right), \quad (7.14)$$

where Ω is the spin vector and I is the planet's moment of inertia. Note that, as before, tidal dissipation is applied only to Gliese 436b. As a result, Gliese 436b's Q-value absorbs all other dissipation in the system. Nonetheless, it is likely that the contribution from the perturber will be comparably small, since $dE/dt \propto (R/a)^5$, as will be the tides raised on the star where $dE/dt \propto (m/M_\star)$. Here also, an artificially low Q-value of 10 (characteristic of Earth-like planets) was used in order to speed up the integration process. Integration was done using the Bulirsch-Stoer algorithm (Press et al. 1992) with a timestep accuracy parameter $\eta = 10^{-14}$. The moment of intertia was taken as $I/M_\mathrm{b}R_\mathrm{b}^2 = 0.254$ and the initial spin frequency was $\Omega = 2.3\,\mathrm{d}$. As we have done throughout the paper, we use $P_b = 2.643\,\mathrm{d}, e_b = 0.15$, and $m_b = 0.0711 M_J$.

Table 7.2: An Example 3-planet Fixed Point Configuration

Planet	P (days)	m (M_J)	e	ϖ (deg)
b	2.64	7.11×10^{-2}	0.15	0.0
c	14.24	6.92×10^{-2}	0.139214	0.0
d	27.33	6.92×10^{-2}	0.152183	0.0

Two scenarios were computed: a tidally dissipated Gliese 436 system with and without a fixed-point $P_c = 40\,\mathrm{d}$ perturber. Figure (5) shows the eccentricity evolutions of the perturbed and unperturbed Gliese 436b's as well as the eccentricity evolution of the perturber. The results of this integration are in good agreement with the semi-analytical integrations discussed in section 2. By measuring the time at which Gliese 436b's eccentricity dropped by a factor of e in the perturbed and unperturbed solutions, we find that the $P_c = 40\,\mathrm{d}$ perturber provides a $f \sim 5.3$ times increase in the circularization timescale. Figure (6) shows the difference of perihelia, $(\varpi_b - \varpi_c)$ as a function of time. Linear secular theory predicts that once the fixed point is reached, the system does not deviate away from it at all. In the full numerical integration, we observe some low-amplitude libration, which is associated with passage of Gliese 436b through high-order MMR's (namely 16:1, 17:1, 18:1, and 19:1 as labeled on figures 5 and 6) .

7.5 Many-Planet Fixed Points

Up to this point, we have only considered a scenario where the Gliese 436b's eccentricity is maintained by a single perturber. While a two-planet scenario is the simplest one, it is certainly possible to obtain solutions where three or more planets are all in a quasi-stationary configuration. In this case, to fully take advantage of the energy reservoir of

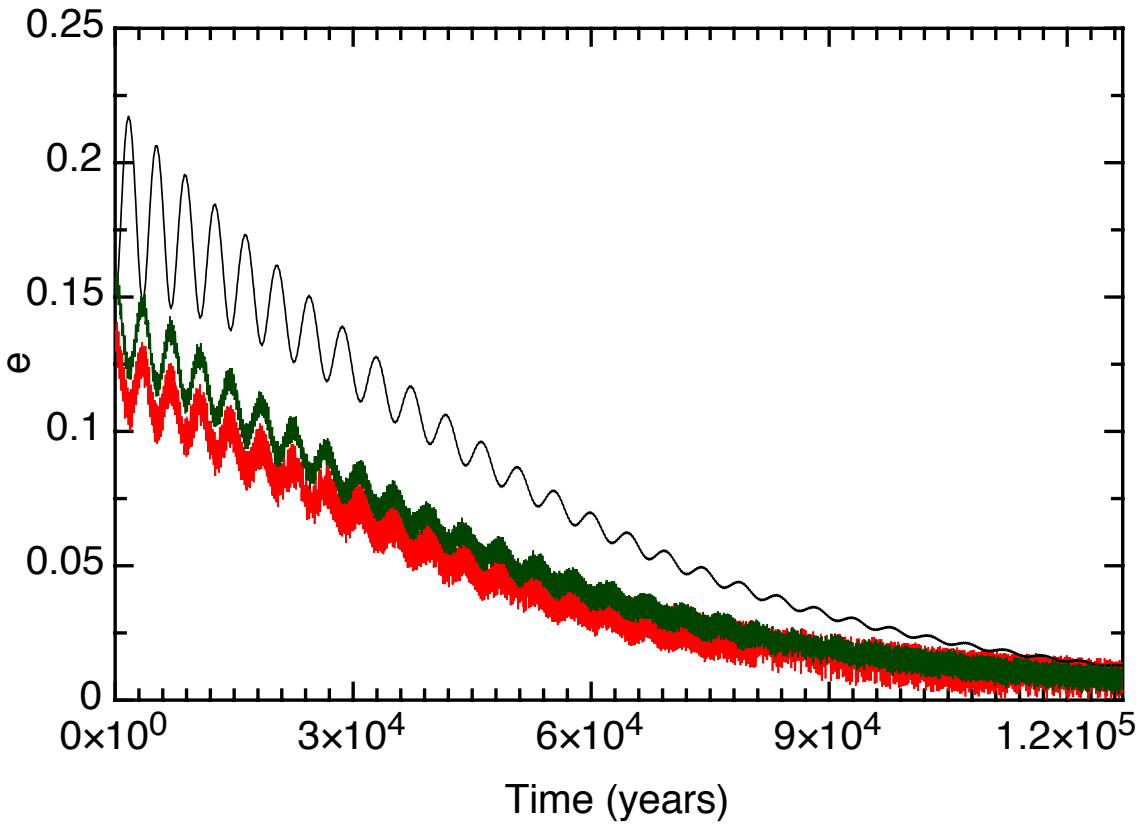


Figure 7.7: Numerically computed eccentricity evolution of the tidally dissipated Gliese 436 system with 3 planets. The starting conditions were computed analytically, and are listed in Table 2. The black, red, and green curves correspond to planets b, c, and d respectively. For the integrations, $Q = 8$ was used.

the system, all consecutive pairs of planets should be dynamically coupled. Let us now examine fixed point systems in the context of a modified classical Laplace-Lagrange secular theory.

To second order in eccentricities and first order in masses, the averaged co-planar disturbing function of planet i can be written as

$$\left\langle \mathcal{R}_i^{(sec)} \right\rangle = n_i a_i^2 \left[\frac{1}{2} A_{ii} e_i^2 + \sum_{j=1, j\neq i}^{N} A_{ij} e_i e_j \cos(\varpi_i - \varpi_j) \right]. \tag{7.15}$$

The constant coefficients A_{ij} form a N $\times$ N matrix, which renders the dynamics of the

system (Murray & Dermott, 1999). Here, the effects of GR are accounted for by introducing a leading-order correction to the diagonal elements of the **A** matrix (for details on this standard modification, see Adams & Laughlin, 2006). As a result, the elements take the form:

$$A_{ii} = \frac{n_i}{4} \sum_{j=1, j\neq i}^{N} \frac{m_j}{M_\star + m_i} \alpha_{ij} \bar{\alpha}_{ij} b_{3/2}^{(1)}(\alpha_{ij}) + 3 n_i \frac{G M_\star}{c^2 a_i}, \tag{7.16}$$

$$A_{ij} = -\frac{n_i}{4} \sum_{j=1, j\neq i}^{N} \frac{m_j}{M_\star + m_i} \alpha_{ij} \bar{\alpha}_{ij} b_{3/2}^{(2)}(\alpha_{ij}), \tag{7.17}$$

where $\alpha_{ij} = a_i/a_j$ if $(a_i < a_j)$; a_j/a_i if $(a_j < a_i)$, $\bar{\alpha}_{ij} = \alpha_{ij}$ if $(a_i < a_j)$; 1 if $(a_j < a_i)$, $b_{3/2}^{(1)}(\alpha_{ij})$ & $b_{3/2}^{(2)}(\alpha_{ij})$ are Laplace coefficients of first and second kind respectively. The lowest-order Lagrange's equations of motion can be written as

$$\frac{de_i}{dt} = -\frac{1}{n_i a_i^2 e_i} \frac{\partial \mathcal{R}_i}{\partial \varpi_i} = \sum_{j=1, j\neq i}^{3} A_{ij} e_j \sin(\varpi_i - \varpi_j) \tag{7.18}$$

$$\frac{d\varpi_i}{dt} = \frac{1}{n_i a_i^2 e_i} \frac{\partial \mathcal{R}_i}{\partial e_i} = A_{ii} + \sum_{j=1, j\neq i}^{3} A_{ij} \frac{e_j}{e_i} \cos(\varpi_i - \varpi_j). \tag{7.19}$$

From these expression, it can be seen that de/dt vanishes when the orbits are apsidally aligned and anti-aligned. Since all precession rates of a fixed point system are equal, given the masses and semi-major axes of the perturbing planets, we can solve for their fixed-point eccentricities. Care must be taken in this approach however, since Laplace-Lagrange theory only applies in the limit of small eccentricities[2].

Let us now consider a single proof-of-concept calculation. Setting the masses of the two perturbers at $22 m_\oplus$ each, we allowed the semi-major axes to be generated randomly,

[2]The issue of small eccentricities can be forestalled by expanding the secular disturbing function to higher order in e.

and then solved for the perturbers' aligned fixed-point eccentricities. We then used the numerical scheme discussed above to integrate the full equations of motion forward in time. The orbital elements of this configuration are presented in table 2, whereas figure 7 shows the system's evolution towards a triply-circular state. Since the eccentricities of this configuration are somewhat higher than what is required for Laplace-Lagrange theory to reach high levels of precision, it is not surprising that the analytically obtained fixed point is only approximately stationary, and some eccentricity modulation in the numerical integration is observed.

In principle, a similar survey to that described in section 2 can be performed. However, as more planets are added, the parameter space, available for exploration, becomes overwhelmingly large. Thus, such a search would be most effective in case a perturbing planet is already found, but is clearly not at a single-perturber fixed point. Indeed, such a detection may indirectly hint to the presence of additional, dynamically significant, bodies in the system.

7.6 Discussion

In this paper, we put forward the hypothesis that the Gliese 436 system has an additional, as-yet undetected companion, Gliese 436c, that is *not* in a low-order mean motion resonance with Gliese 436b, but is at a secular fixed point. An important role of such a perturber would be to increase the tidal circulization timescale of the system and thus help explain the unusually high eccentricity, $e = 0.15$, of Gliese 436b.

Using the dissipated secular theory developed by Mardling (2007), we surveyed the pa-

rameter space of allowed perturbers, restricting our domain to fixed-point configurations. This survey revealed that apsidally aligned, fixed-point systems can provide significantly longer circularization times than the anti-aligned states, and are thus preferred. Furthermore, the population of possible stationary, apsidally aligned, perturbers forms a clear and well-defined locus in (P_c ,M_c ,e_c) space. This sequence is accompanied by the amplitude of transit timing variations, $\Delta T_{\rm max}$, listed in Table 1. As a check of the validity of our semi-analytic theory, we verified that the full disturbing function produces only small corrections to the orbital elements of the fixed-point configurations derived from octopole theory. In addition, we performed one self-consistent numerical integration of the circularization a hypothetical $P_c = 40\,{\rm d}$ perturber. The results are in good agreement with the octopole theory, allowing us to verify that a factor, $f = 5.3$ increase in circularization timescale is robust.

A satisfactory theory makes predictions that are not only testable, but *readily* testable. The presence of Glise 436c, with the orbital and physical properties that we're proposing would fall into the readily testable category.

We make several predictions. **1** If a secular perturber exists and holds significant responsibility for the upkeep of Gliese 436b's eccentricity, we believe that it will fall on the loci shown in Figures 1 and 2 and delineated in Table 1. **2** Our theory gives a roughly six-fold increase in the circularization timescale derived by Tong & Zhou (2008). If we adopt $Q = 300,000$, consistent with the upper bound on Banfield & Murray's (1992) estimate of Neptune's Q-value, and near the lower bound on Jupiter's Q-value, then the effective (fixed-point) circularization timescale becomes $\tau \sim 8\,{\rm Gyr}$, a value that modestly exceeds

the likely age of the star. In addition, the system may also require up to ~ 3 single-planet circularization timescales to arrive at the fixed point. As a result, for the first time, we don't need to invoke an uncomfortably high Q-value to explain the observed eccentricity, but if our theory is correct then Gliese 436b's effective internal dissipation is a factor of 10 times less than Uranus and is near the lower end of Neptune's range, suggesting a fundamental structural difference between the "hot Neptune" and the slightly less massive solar system "ice giants". **3** We predict that Gliese 436b will show *no* measurable tidal heating. On the surface, this appears inconsistent with Deming et al 2007's measurement of $T = 712 \pm 36\,\mathrm{K}$ for the planet's $8\,\mu m$ brightness temperature. If our theory is correct, the apparent temperature excess over the $T_{eq} = 642\,\mathrm{K}$ planetary effective temperature will arise from a non-blackbody emission spectrum, weather variability, measurment uncertainty, or a combination of the three.

In the previous section, we reviewed the possibility of more than one perturber maintaining Gliese 436b's eccentricity, utilizing modified Laplace-Lagrange secular theory, and performed one numerical integration as an example. In light of this possibility, we note that observational detection of a companion *not* in a 2-planet fixed configuration may indirectly point at the presence of additional planets in the system.

What are the prospects for near term observational detection of Gliese 436 c? During the two years following the radial velocity analysis of Maness et al. (2007) the Gliese 436 system has been placed under intensive radial velocity surveillance. At the IAU 253 symposium in May 2008, the Geneva Planet Search team (Lovis et al. 2008) showed several dozen new high-precision Doppler measurements, which indicated that the possible long-

term linear trend noted by Maness et al. (2007) has either abated, or was an artifact of the limited observational baseline. A residuals periodogram of the Geneva team's velocity data showed a number of peaks in the $10\,\mathrm{d} < P < 100\,\mathrm{d}$ range, several of which appeared tall enough to correspond to a perturber with the properties required by our theory.

The EPOXI mission (Deming et al. 2008) held Gliese 436 under continuous photometric observation for 27 days during May 2008. The full results of their investigation have not yet been published, but the EPOXI transit timing precision for Gliese 436b is expected to be of the order of a few tens of seconds (Deming 2009). At this level of precision, if our hypothesis of a fixed point perturber is correct, we predict that *no* transit timing variations will be observed by EPOXI. Thus far, our hypothesis has been consistent with the lack of transit timing variations above $\Delta t \sim 45$ seconds reported by Coughlin et al (2008). If transit timing variations do exist, their maximum amplitude will lie in the neighborhood of $\Delta t \sim 1$ to $\Delta t \sim 10$ seconds, and their maximum amplitude will give a strong indication of the perturber's orbital properties, as interpolated from Table 1.

Recent improvements in high-cadence, high-precision ground-based photometry have improved the measurement of transit midpoints to within $\Delta t \sim 5$ seconds (see e.g. Johnson et al. 2008). If this level of precision can be repeatedly applied to Gliese 436, then there is a significant chance that the presence of a perturbing planet can be inferred. In addition, the predicted radial velocity half-amplitudes, K, of the perturber are generally well within the current threshold of detection, which should enable a ready confirmation or falsification of our hypothesis.

We thank Fred Adams, Dan Fabrycky, Rosemary Mardling, Stan Peale, David Steven-

son, Darin Ragozzine, and Drake Deming for useful discussions.

This research is based in part upon work supported by the National Science Foundation CAREER program under Grant No. 0449986 (GL), NSF AST-0307493 Grant (SV) and by NASA through the Ames Astrobiology Institute under cooperative agreement NNA06CB31A.

Bibliography

Adams, F. C., & Laughlin, G. 2006, The Astrophysical Journal, 649, 1004

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, Monthly Notices of Royal Astronomical Society, 359, 567

Alonso, R., Barbieri, M., Rabus, M., Deeg, H. J., Belmonte, J. A., & Almenara, J. M. 2008, ArXiv e-prints, 804, arXiv:0804.3030

Banfield, D., & Murray, N. 1992, Icarus, 99, 390

Bean, J. L., Benedict, G. F., Charbonneau, D., et al. 2008, Astronomy & Astrophysics, 486, 1039

Butler, R. P., Marcy, G. W., Fischer, D. A., Brown, T. M., Contos, A. R., Korzennik, S. G., Nisenson, P., & Noyes, R. W. 1999, The Astrophysical Journal, 526, 916

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

Chambers, J. E. 1999, Monthly Notices of Royal Astronomical Society, 304, 793

Coughlin J., Stringfellow G., Becker A., Lopez-Morales M., Mezzalira F., & Krajci T. 2008, The Astrophysical Journal Letters, 689, L149

Deming, D., Harrington, J., Laughlin, G., Seager, S., Navarro, S. B., Bowman, W. C., & Horning, K. 2007, The Astrophysical Journal Letters, 667, L199

Demory, B.-O., et al. 2007, Astronomy & Astrophysics, 475, 1125

Eggleton P. P., Kiseleva, L. G., & Hut, P. W. 1998, The Astrophysical Journal, 499, 853

Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, The Astrophysical Journal, 659, 1661

Gillon, M., et al. 2007, Astronomy & Astrophysics, 472, L13

Goldreich, P., & Peale, S. 1966, The Astronomical Journal, 71, 425

Goldreich, P., & Soter, S. 1966, Icarus, 5, 375

Hut, P., 1981, Astronomy & Astrophysics, 99, 126

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

Jackson, B., Greenberg, R., & Barnes, R. 2008a, The Astrophysical Journal, 678, 1396

Jackson, B., Barnes, R., & Greenberg, R. 2008b, AAS/Division for Planetary Sciences Meeting Abstracts, 40, #04.02

Johnson, J. A., Winn, J. N., Cabrera, N. E., & Carter, J. A. 2008, arXiv:0812.0029

Laplace, P. S. 1799, Traitede Mecanique Celeste (Paris: Gauthier-Villars)

Levrard, B., Correia, A. C. M., Chabrier, G., Baraffe, I., Selsis, F., & Laskar, J. 2007, Astronomy & Astrophysics, 462, L5

C. Lovis, M. Mayor, F. Bouchy, F. Pepe, D. Queloz, S. Udry, W. Benz & C. Mordasini, 2008, Proceedings of the International Astronomical Union , Volume 4, Symposium S253, May 2008, pp 502-505

Maness, H. L., Marcy, G. W., Ford, E. B., Hauschildt, P. H., Shreve, A. T., Basri, G. B., Butler, R. P., & Vogt, S. S. 2007, Publications of the Astronomical Society of the Pacific, 119, 90

Mardling, R. A. & Lin D. N. C. 2002, The Astrophysical Journal, 573, 829

Mardling, R. A. 2007, Monthly Notices of Royal Astronomical Society, 382, 1768

Mardling, R. A. 2008, submitted to Monthly Notices of Royal Astronomical Society.

Matsumura, S., Takeda G. & Rasio, F. A. 2008 The Astrophysical Journal, 686, 1

Mayor, M., & Queloz, D. 1995, Nature, 378, 355

Meyer, J., & Wisdom, J. 2008, Icarus, 193, 213

Murray, C. D., & Dermott, S. F. 1999, Solar System Dynamics (Cambridge: Cambridge Univ. Press) (MD99)

Nobili, A., & Roxburgh, I. W. 1986, Relativity in Celestial Mechanics and Astrometry. High Precision Dynamical Theories and Observational Verifications, 114, 105

Peale, S. J., & Cassen, P. 1978, Icarus, 36, 245

Peale, S. J., & Greenberg, R. J. 1980, Lunar and Planetary Institute Conference Abstracts, 11, 871

Pont, F., Gilliland, R. L., Knutson, H., Holman, M., & Charbonneau, D. 2008, arXiv:0810.5731

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P. 1992, Numerical recipes in FORTRAN: The art of scientific computing (Cambridge: Cambridge University Press)

Ragozzine, D., & Wolf, A. S. 2008, arXiv:0807.2856

Ribas, I., Font-Ribera, A., & Beaulieu, J.-P. 2008, The Astrophysical Journal Letters, 677, L59

Tittemore, W. C., & Wisdom, J. 1989, Icarus, 78, 63

Tong X. & Zhou J., 2008, accepted to Science in China (G): Physics, Mechanics & Astronomy

Torres, G. 2007, The Astrophysical Journal Letters, 671, L65

Winn, J. N., et al. 2005, The Astrophysical Journal, 631, 1215

Wisdom, J., 2008, Icarus, 193, 637

Wu, Y. 2003, Scientific Frontiers in Research on Extrasolar Planets, 294, 213

Zhang, K. & Hamilton D. P., 2008, Icarus, 193, 267

Chapter 8

Resolving the Sin(I) degeneracy in Low-Mass Multi-Planet Systems

Originally published as:
Batygin, K., & Laughlin, G. 2011, The Astrophysical Journal, 730, 95

8.1 Abstract

Long-term orbital evolution of multi-planet systems under tidal dissipation often converges to a stationary state, known as the tidal fixed point. The fixed point is characterized by a lack of oscillations in the eccentricities and apsidal alignment among the orbits. Quantitatively, the nature of the fixed point is dictated by mutual interactions among the planets as well as non-Keplerian effects. We show that if a roughly coplanar system hosts a hot, sub-Saturn mass planet, and is tidally relaxed, separation of planet-planet interactions and non-Keplerian effects in the equations of motion leads to a direct determination of the true masses of the planets. Consequently, a "snap-shot" observational determination of the orbital state resolves the $\sin(I)$ degeneracy, and opens up a direct avenue towards identification of the true lowest-mass exo-planets detected. We present an approximate, as well as a general, mathematical framework for computation of the line of sight inclination of

secular systems, and apply our models illustratively to the 61 Vir system. We conclude by discussing the observability of planetary systems to which our method is applicable and we set our analysis into a broader context by presenting a current summary of the various possibilities for determining the physical properties of planets from observations of their orbital states.

8.2 Introduction

Since the seminal discovery of the first giant planet orbiting a main sequence star (Mayor & Queloz 1995), using the radial velocity (RV) method, over 400 additional extra-solar planets have been confirmed. The greatest disadvantage of the RV method lies in the uncertainty of the true masses of the discovered planets, as the inclination of the orbits to the line of sight, I, are unknown. In resonant systems, such as GL876, monitoring of the resonant argument and the precession rates may lead to determination of the true masses (e.g. Rivera et al 2005). In the vast majority of cases, however, the $\sin(I)$ degeneracy, remains a continued source of frustration.

Still, RV surveys persist in yielding fruitful results, and the continued detection of exoplanets has brought forth many surprises. Perhaps one of the biggest surprises has been the discovery of extremely close-in bodies whose mass-range spans the entire planetary spectrum. These objects have since become a subject of fascination in the community and more importantly, have provided a new test-bed for various theoretical efforts.

Extra-solar multi-planet systems that host "hot" planets differ drastically from our own solar system in many ways, including orbital dynamics. In our solar system, gravitational

interactions among the planets are sufficient to, at least approximately, explain orbital evolution. In many extra-solar planetary systems however, similarly to the case of the Galilean satellites, dissipation of orbital energy due to tides plays an unavoidably important role. The long-term effect of this additional interaction provides an opportunity to infer important additional properties of the system that cannot be observed directly.

Qualitatively speaking, in a system of two or more planets that are not in a mean-motion resonance and are roughly coplanar, tides drive the orbits towards a stationary state i.e. a "fixed point". A fixed point is characterized by continued apsidal alignment and a well-determined eccentricity ratio that is nearly constant in time (Wu & Goldreich 2002, Mardling 2007). The factors that determine the actual quantitative nature of the state are not limited to gravitational planet-planet interactions. Indeed, general relativistic and tidal corrections, among other things, play a crucial role. It is through these "non-Keplerian" interactions that additional information can be learned, as they are governed by parameters other than just planetary masses,.

Upon discovery of the first multiple planetary system with a transiting "hot Jupiter", Hat-P-13 (Bakos et al 2009), it was pointed out that the system likely resides at a fixed point (Batygin, Bodenheimer & Laughlin 2009). Furthermore, it was shown that as the mass and radius of the inner planet are known, consideration of the planetary quadru-pole gravitational field, and its contribution in determination of the fixed point leads to a direct measurement of the planetary interior structure. In other words, a precise "snap-shot" of the orbits gives the planetary Love number, k_2, which is a measure of the interior density distribution, with high accuracy.

The last decade of observations has revealed that generally, hot Jupiters tend not to be accompanied by readily detectable companion planets (Ragozzine & Holman 2010). Smaller planets, such as hot Neptunes and hot Super-Earths, however, tend to occur in multiple-planet systems (Lo Curto et al 2010), hinting at different migration histories (Terquem & Papaloizou 2007).

Here, we consider the latter class of systems, with an eye toward inferring conventionally unobservable planetary properties that influence the details of the fixed-point configuration. In particular, we show that if a non-transitng (RV) system hosts a small ($R \lesssim R_{\rm Nep}$) hot planet, it is possible to derive the true masses of the planets i.e. resolve the $\sin(I)$ degeneracy from a detailed determination of the system's orbital state. The plan of the paper is as follows: in section 2, we outline our mathematical model. In section 3, we apply the theory to the 61 Vir (Vogt et al 2010) system. In section 4, we discuss the possibility of determination of the radius and interior structure of massive RV planets. We conclude and discuss our results in section 5.

8.3 Dynamical Evolution of a Planetary System With a Close-in Planet

As already mentioned above, there are important differences between the dynamics of systems with and without close-in planets. In conservative (Hamiltonian) systems, of which our solar system provides an excellent approximation (Laskar 2008), the motion of the planets is subject to Liouville's theorem. Accordingly, strictly Hamiltonian flow can have

no attractors in phase-space (Morbidelli 2002). Conversely, in dissipative systems, the phase space volume explored by the system continuously contracts, and truly steady-state solutions are possible. In other words, tides are needed for the system to arrive to a stationary state.

The path that the system will take to the fixed point is non-unique and depends on the initial conditions. Consequently, the initial transient period will also depend on the initial state. However, the fixed-point itself is unique for a chosen set of system parameters, and the system has no memory of its own evolution once it arrives to the fixed point. Thus, any quantity that is inferred from the fixed point is independent of the system's formation history. We now describe a mathematical model for the system's evolution to a stationary state and its orbital characteristics.

8.3.1 Secular Interactions WIth non-Keplerian Effects

Whenever planets are far away from low-order mean motion commensurabilities and the orbits are not changing significantly on the orbital time-scale (i.e. planets are not scattering), a secular approximation to the dynamics can be made. The secular approximation refers to an averaging procedure, where the gravitational potential between planets is averaged over the mean longitudes, thereby reducing the degrees of freedom inherent to the problem.

Since the pioneering work of Laplace (1772) and Lagrange (1776), a number of perturbation theories based on various approximation of the disturbing potentials have been developed and applied in both solar system and exoplanetary contexts (Le Verrier 1856,

Brouwer & van Woerkom 1950, Laskar 1986, Laskar 2008, Eggleton & Kiseleva-Eggleton 2001, Mardling & Lin 2002, Lee & Peale 2003, Michtchenko & Malhotra 2004, Migaszewski & Goździewski 2009, Lovis et al 2010, etc). Still, it is perhaps easiest to illustrate the ideas presented here in the context of a modified Laplace-Lagrange (LL) secular theory.

The classical secular disturbing function (planet-planet potential), of N secondaris that interact solely by Newtonian gravity, expanded to first order in masses and second order in eccentricities reads (Murray & Dermott 1999)

$$\mathcal{R}_j^{(sec)} = n_j a_j^2 [\frac{1}{2} A_{jj} e_j^2 + \sum_{k=1,k\neq j}^{N} A_{jk} e_j e_k \cos(\varpi_j - \varpi_k)] \tag{8.1}$$

where e is eccentricty, ϖ is the longitude of perihelion, a is semi-major axes and n is mean motion. The constant coefficients A take the form

$$A_{jj} = \frac{n_j}{4} \sum_{k=1,k\neq j}^{N} \frac{m_k}{M_\star + m_j} \alpha_{jk} \bar{\alpha}_{jk} b_{3/2}^{(1)}(\alpha_{jk}) \tag{8.2}$$

$$A_{jk} = -\frac{n_j}{4} \frac{m_k}{M_\star + m_j} \alpha_{jk} \bar{\alpha}_{jk} b_{3/2}^{(2)}(\alpha_{jk}) \tag{8.3}$$

where $\alpha_{jk} = a_j/a_k$ if $(a_j < a_k)$; a_k/a_j if $(a_k < a_j)$, $\bar{\alpha}_{jk} = \alpha_{jk}$ if $(a_j < a_k)$; 1 if $(a_k < a_j)$, $b_{3/2}^{(1)}(\alpha_{jk})$ & $b_{3/2}^{(2)}(\alpha_{jk})$ are Laplace coefficients of first and second kind respectively, and $m = \tilde{m}/\sin(I)$ are the true masses of the planets i.e. $\tilde{m}$ are the measured minimum masses and I is the inclination of the system from line of sight.

Upon application of the linear form of Lagrange's planetary equations in terms of polar coordinates ($h = e\cos(\varpi), k = e\sin(\varpi)$), a linear system of ODE's emerges, where the $\mathbf{A}$

matrix encapsulates the dynamics of the system:

$$\frac{dh_j}{dt} = \sum_{k=1}^{N} A_{jk} k_k \qquad \frac{dk_j}{dt} = -\sum_{k=1}^{N} A_{jk} h_k \tag{8.4}$$

We can express the system of equations more compactly by switching to complex Poincaré variables $z \equiv e\ \mathbf{e}^{i\varpi} = h + ik$. Simple chain rule yields

$$\frac{dz_j}{dt} = \sum_{k=1}^{N} i A_{jk} z_{jk} \tag{8.5}$$

This eigensystem can be solved in the standard way, similar to the problem of N coupled pendulums, and the solution reads:

$$z_j(t) = \sum_{k}^{N} \beta_{jk}\ e^{i(g_k t + \delta_k)} \tag{8.6}$$

where g's are the eigenfrequencies and β's are the eigenvectors of the $\mathbf{A}$ matrix. The relative amplitudes of the eigenvectors and the corresponding phases, δ, are determined by initial conditions.

The above formulation does not take into account the additional orbital precession induced by general relativity (GR), stellar and planetary spin, and the tidal bulges of the star and the planet. The classical LL solution often gives poor quantitative approximations to the orbital evolution of extra-solar planets (Veras & Armitage 2007), where the additional precession can dominate (Ragozzine & Wolf 2009).

The contributions to apsidal precession from the above-mentioned effects can be written

as follows (Sterne 1939):

$$\left(\frac{d\varpi}{dt}\right)_{GR} = \frac{3GM_\star n}{ac^2(1-e^2)} \tag{8.7}$$

$$\begin{aligned}\left(\frac{d\varpi}{dt}\right)_{spin} = & \frac{n\, k_{2p}}{2(1-e^2)^2}\left(\frac{R_p}{a}\right)^5\left(\frac{\Omega_p^2 a^3}{Gm_p}\right) \\ & + \frac{n\, k_{2\star}}{2(1-e^2)^2}\left(\frac{R_\star}{a}\right)^5\left(\frac{\Omega_\star^2 a^3}{GM_\star}\right)\end{aligned} \tag{8.8}$$

$$\begin{aligned}\left(\frac{d\varpi}{dt}\right)_{tidal} &= \frac{15n}{2}k_{2p}\left(\frac{R_p}{a}\right)^5\frac{M_\star}{m_p}\left(\frac{1+\frac{3}{2}e^2+\frac{1}{8}e^4}{(1-e^2)^5}\right) \\ &+ \frac{15n}{2}k_\star\left(\frac{R_\star}{a}\right)^5\frac{m_p}{M_\star}\left(\frac{1+\frac{3}{2}e^2+\frac{1}{8}e^4}{(1-e^2)^5}\right)\end{aligned} \tag{8.9}$$

where c is the speed of light, k_2 is the Love number (twice the apsidal motion constant), R is physical radius and Ω is the spin frequency. In equations (8) and (9) the first terms correspond to the planet and the latter terms correspond to the star. Neglecting higher-order effects, the total additional apsidal precession, evaluated for each planet can be organized into a square diagonal matrix

$$B_{jj} = \left(\frac{d\varpi}{dt}\right)_{GR} + \left(\frac{d\varpi}{dt}\right)_{spin} + \left(\frac{d\varpi}{dt}\right)_{tidal}, \tag{8.10}$$

and added to the $\mathbf{A}$ matrix in equation (5). This matrix is not to be confused with the mutual inclination interaction matrix (see Murray & Dermott 1999), for which the standard notation is the same. When evaluating the additional precessions (the $\mathbf{B}$ matrix) in the context of LL theory, it is customary to expand equations (7) - (9) to first order in e, such

that the dependence on e disappears, and equation (5) remains linear in eccentricity, thus retaining its analytical solution. The augmentation of the diagonal matrix coefficients will modify the eigensystem quantitatively. However, the qualitative essence of the solution remains unchanged: the solution (equation 6) is still a sum of sinusoids with constant amplitudes.

So far, we have retained all additional precession terms for the sake of completeness. Before proceeding further, let us examine the relative importance of the terms that depend on the physical properties of the bodies with respect to GR, which is a purely geometrical effect and only depends on stellar mass and the orbital parameters. Consider the following dimension-less numbers

$$\Lambda^p_{spin} = \frac{c^2 k_{2p} R_p^5 \Omega_p^2}{6aG^2 M m_p} \qquad \Lambda^\star_{spin} = \frac{c^2 k_{2\star} R_\star^5 \Omega_\star^2}{6aG^2 M_\star^2} \tag{8.11}$$

$$\Lambda^p_{tidal} = \frac{5c^2 k_{2p} R_p^5}{2a^4 G m_p} \qquad \Lambda^\star_{tidal} = \frac{5c^2 k_{2\star} R_\star^5 m_p}{2a^4 G M_\star^2}. \tag{8.12}$$

A Jupiter-like planet at a characteristic close-in orbit ($P \sim 3$ days) has $\Lambda^p_{spin} \sim 0.05$ $\Lambda^\star_{spin} \sim 3 \times 10^{-5}$, $\Lambda^p_{tidal} \sim 1$ and $\Lambda^\star_{tidal} \sim 5 \times 10^{-4}$. Inflated hot Jupiters will often have $\Lambda^p_{tidal} \gg 1$, due to the R_p^5 dependence of the tidal term. Thus, precession rates of many hot Jupiters are completely dominated by the planetary tidal term, distantly followed by GR. As mentioned above already, this effect has has been used to infer the interior structures of transiting hot Jupiters, both in isolation (Ragozzine & Wolf 2009) and in the presence of a perturbing companion (Batygin, Bodenheimer & Laughlin 2009). Conversely, for a Neptune-like planet on a 3-day orbit, $\Lambda^p_{spin} \sim 0.005$ $\Lambda^\star_{spin} \sim 3 \times 10^{-5}$, $\Lambda^p_{tidal} \sim 0.1$ and

$\Lambda^{\star}_{tidal} \sim 3 \times 10^{-6}$. The numbers continue to decline for super-Earths and terrestrial planets. This implies that in practice, the apsidal advance, resulting from rotation of both the planet and the star, as well as that resulting from the stellar tidal bulge, can often be neglected. Indeed, the situation is bimodal: for large planets, tidal precession dominates, where as for small planets, GR dominates the extra apsidal advance.

Let us now add dissipative tides to the system. Generally, tidal heating conserves the total angular momentum, but not energy. This leads to decay of the planetary eccentricity, as well as decay (or growth, depending on stellar spin) of the planet's semi-major axis (Goldreich 1963). The evolution of the semi-major axis happens over a much longer time-scale than that of the eccentricity, so in our simplified model, we adopt the standard practice of parameterizing tides with a constant decay of the eccentricity, $dz/dt = z/\tau_c$, where τ_c is the circularization timescale (Goldreich & Soter 1966):

$$\tau_c = \frac{P_p}{21\pi} \frac{Q_p}{k_p} \frac{m_p}{M_\star} \left(\frac{a}{R_p} \right)^5 . \tag{8.13}$$

Here, P is the orbital period and Q is a tidal quality factor. In a similar fashion as above, each planet can be subjected to tidal damping of eccentricity by constructing a square diagonal matrix with the elements $C_{jj} = 1/\tau_c^{(j)}$. Note that because tidal dissipation only affects semi-major axes, eccentricities and rotation rates directly, an identical procedure cannot be carried out for the mutual inclination eigenmode solution (see Mardling 2010 for an in-depth discussion). The equation of motion that accounts for the additional precession

and tidal damping of the eccentricity takes the form:

$$\frac{dz_j}{dt} = \sum_{k=1}^{N} [i(A_{jk} + B_{jk})z_{jk} + C_{jk}z_{jk}] \tag{8.14}$$

At this point, we have changed the solution qualitatively. The introduction of eccentricity damping has added a complex component to the eigenfrequencies. Consequently, in the secular solution (6), real exponential decay factors appear in front of the oscillatory solution. The eigenvectors are now damped. Furthermore, the imaginary components of the eigenfrequencies need not be equal, and generally will not be, except for a narrow set of system parameters. This implies that the decay timescale of one of the modes,

$$\tau_{decay}^{(j)} = (\mathrm{Im}\,[g_j])^{-1} \tag{8.15}$$

can be considerably longer than all others, and the system will eventually evolve to a state that is characterized by a single eigenmode. Note that the eigenmode decay timescale can greatly exceed the tidal circularization timescale, prolonging the lifetime of the dissipated planets' eccentricities. Upon inspection of equation (6), it is clear that once the system is characterized by a single eigenmode, the rates of orbital precession are identical for all planets in the system. From Lagrange's planetary equations, this automatically implies that the apsidal angles between the orbits must be equal to $\Delta\varpi = 0$ or $\Delta\varpi = \pi$. In other words, all orbits are either aligned or anti-aligned, depending on which particular eigenmode has survived. Additionally, in this case, *the ratios of the eccentricities are also well-defined by the eigenvector of the surviving mode*. When the system has reached a state where its

dynamics are characterized by a single mode, it has reached a "fixed point." Addition of higher-order terms to the disturbing function will modify the the eccentricity ratios implied by the fixed point, but will not cause $\Delta\varpi$ to be anything other than 0 or π.

8.3.2 Determination of $\sin(I)$

We now have all the necessary ingredients to determine the system inclination. Usually, the dissipation time-scale greatly exceeds the secular time-scale ($\mathbf{C} \ll \mathbf{A}, \mathbf{B}$), so a system at a fixed point is characterized by a single eigenvector of the $[\mathbf{A} + \mathbf{B}]$ matrix[1]. From the definitions of the coefficients of $\mathbf{A}$ (equations 2 & 3), it is clear that they are linearly proportional to $\sin(I)$. In fact, we can replace the true masses, m, by the minimum masses $\tilde{m}$ in equations (2 & 3) and write $\mathbf{A} = \tilde{\mathbf{A}}/\sin(\mathbf{I})$. $\mathbf{B}$ is however independent of the system inclination, given that GR is the only contributing factor. Recall that this is the case for Neptune-sized and smaller planets, for which $\Lambda_{spin} \ll 1$ and $\Lambda_{tidal} \ll 1$. As a result, the eigenvectors of the $\left[\tilde{\mathbf{A}}/\sin(\mathbf{I}) + \mathbf{B}\right]$ matrix, which physically correspond to the eccentricity ratios of the planets, depend explicitly on $\sin(I)$. Namely, every value of the system inclination corresponds to an eccentricity ratio of the planets. Consequently, *a precise observational determination of the eccentricity ratios yields the true masses of the system*. Let us turn to an illustrative example below.

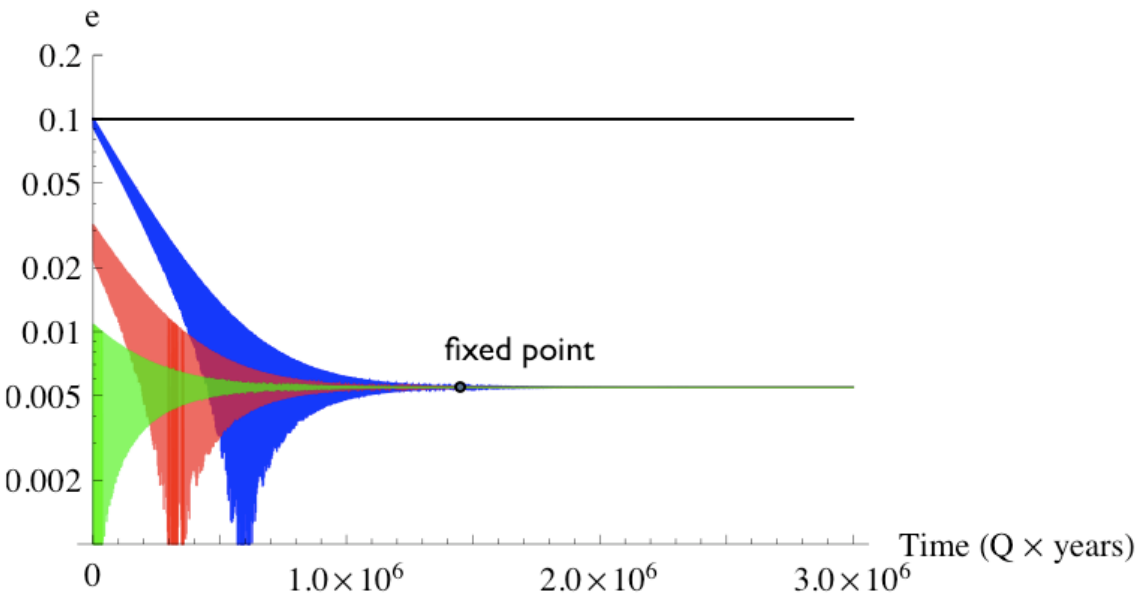


Figure 8.1: A damped, modified Laplace-Lagrange secular solution of a 2-planet system with $m_1 = 10^{-5}M_\odot, m_2 = 10^{-2}M_\odot, a_1 = 0.03$AU, $a_2 = 0.3$AU ($\alpha = 0.1$) and $e_2 = 0.1$. Three solutions are presented corresponding to the initial conditions $e_1 = 0.1$ (Blue), $e_1 = 0.03$ (Red), and $e_1 = 0$ (Green), with randomly chosen longitudes of perihelia. The black line shows the eccentricity of the outer planet. The apsidal angles initially circulate, but switch to libration at $t \approx Q \times 6 \times 10^5$years. The system reaches a fixed point as the anti-aligned (g_1) mode decays away completely at $t \approx Q \times 1.3 \times 10^6$years. Note that the system looses memory of its initial conditions as it approaches the fixed point.

8.3.3 Beyond Linear Order in e: the Case of Well-Separated Orbits

Consider the case of two well-separated ($\alpha \ll 1$) secondaries, where the inner planet is on a close-in orbit. In such a scenario, we only need to consider the additional apsidal precession of the inner planet. Since $\alpha \ll 1$, it is sensible to expand Laplace coefficients in equations (2-3) into hypergeometric series and retain only the first terms: $b^{(1)}_{3/2}(\alpha) \approx 3\alpha$, $b^{(2)}_{3/2}(\alpha) \approx (15/4)\alpha^2$. With a little algebra, it is easy to show that to leading order in α and

[1]The physical effect of including $\mathbf{C}$ in the solution is to offset the apses by a small factor proportional to Q^{-1}.

η, the eigenfrequencies take on a simple form:

$$g_1 = \frac{3}{4}\frac{m_2}{M_\star} n_1 \alpha^3 (1 + [\Gamma + i\eta]) \tag{8.16}$$

$$g_2 = \frac{3}{4}\frac{m_1}{M_\star} n_2 \alpha^2 \left(1 + i\eta \left(\frac{5\alpha}{4(1+\Gamma)}\right)^2\right) \tag{8.17}$$

where $\Gamma \equiv B_{11}/A_{11}$ and $\eta \equiv C_{11}/A_{11}$. The two eigenfrequencies physically correspond to modes dominated by the inner (g_1) and outer (g_2) apsidal precessions. Note that the imaginary components of the modes have explicitly different dependences on α. The multiplier in equation (16) is just A_{11}, expanded to first order in α. So neglecting $\sin(I)$ for the moment, it is clear that Im $[g_1] = 1/\tau_c$. This is consistent with the observation of Mardling (2007) that $\sim 3\tau_c$ are needed for the system to attain a stationary state. The situation is wildly different however for the second mode, as Im $[g_2] = (25/16)(m_1/m_2)(\alpha^{5/2}/\tau_c)(1+\Gamma)^{-2}$. Consequently, equation (15) implies that $\tau^{(2)}_{decay} \gg \tau^{(1)}_{decay}$, unless $m_1 \gg m_2$ and the overall lifetime of the inner eccentricity is also greatly enhanced.

The corresponding eigenvectors, also to leading order in α, but neglecting the higher-order correction from η read:

$$\left(\frac{\beta_{11}}{\beta_{12}}\right) = -\frac{4}{5\alpha}\left(1 - \frac{\tilde{m}_2}{\tilde{m}_1}\frac{1+\Gamma}{\sqrt{\alpha}}\right) \gg 1 \tag{8.18}$$

$$\left(\frac{\beta_{21}}{\beta_{22}}\right) = \frac{5\alpha}{4(1+\Gamma)} \ll 1 \tag{8.19}$$

Note that the eigenvector of the first mode is negative. By Euler's identity, the negative sign introduces an additional $i\pi$ in the exponent of the solution (6) for one of the planets.

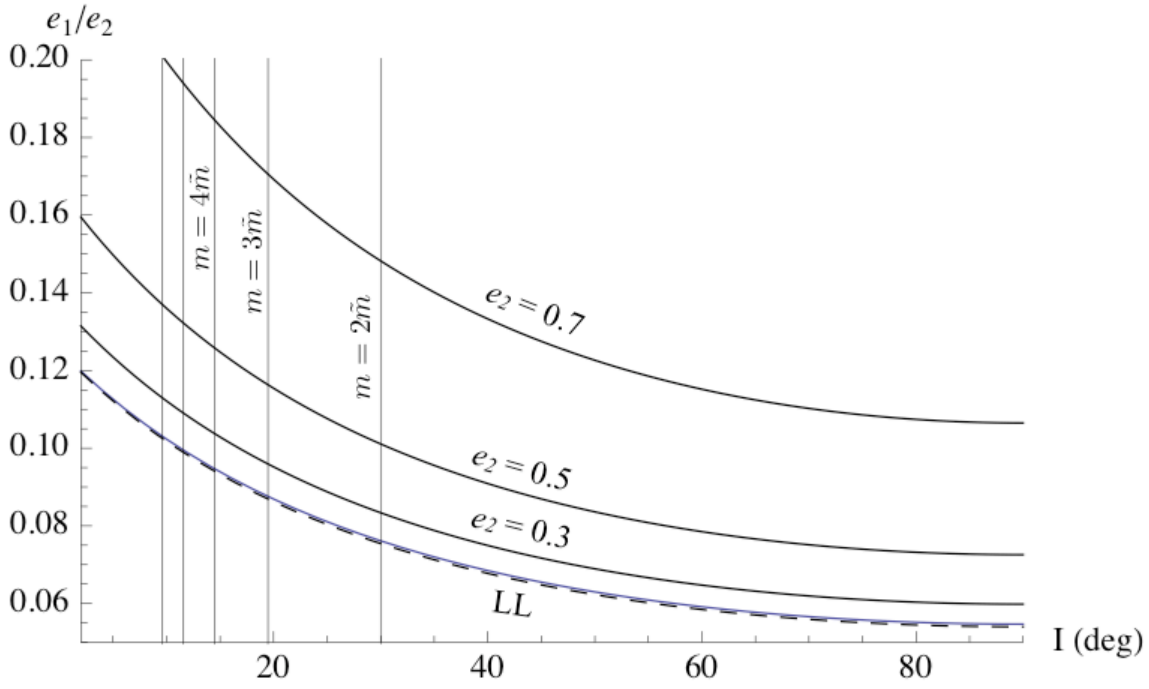


Figure 8.2: Fixed point eccentricity ratio as a function of system inclination for a 2-planet system with $m_1 = 10^{-5}M_\odot, m_2 = 10^{-2}M_\odot, a_1 = 0.03$AU and $a_2 = 0.3$AU ($\alpha = 0.1$) (see Fig. 1). The black curve, labeled LL was computed directly from the Laplace-Lagrange eigenvector solution. The blue dashed curve is the approximation to the LL solution, given by equation (19), corresponding to $e_2 = 0.1$. The curves with eccentricity labels demonstrate the dependence of the eccentricity ratios on the stationary eccentricity of the outer secondary, as dictated by the secular perturbation theory, developed by Mardling (2007). Recall that $\tilde{m}$ refers to the RV minimum mass.

Physically, this corresponds to apsidal anti-alignment. Thus it is apparent from equations (18 & 19) that the first and the second eigenmodes correspond to anti-aligned and aligned orbits respectively.

As an illustration, consider a pair of planets with masses $m_1 = 10^{-5}M_\odot/\sin(I)$, $m_2 = 10^{-2}M_\odot/\sin(I)$, and semi-major axes $a_1 = 0.03AU$, $a_2 = 0.3$ orbiting a $M_\star = 1M_\odot$ star. The ($I = 0$) damped, modified Laplace-Lagrange secular solution of this system is presented in figure (1), where the planets were started with $e_1 = e_2 = 0.1$ and randomly chosen longitudes of perihelia. The planetary Love number was chosen to be $k_{2p} = 0.3$.

Let us examine the evolution in some detail. After an initial transient period of $\sim 3\tau_c$, the system reaches a fixed point. Thereafter, the free eccentricity decays on the timescale of $\tau = \mathrm{Im}\,[g_2] \approx 10^5\tau_c$. As already stated, the addition of a perturbing planet has prolonged the lifetime of the dissipated planet's eccentricity immensely. As a result, it must be pointed out that the detection of an eccentric close-in planet alone does not imply that the planet itself is weakly dissipative. Rather, self-consistent calculations are required to place any constraints on Q.

The above analysis implies that planets on well-separated orbits in a tidally relaxed system will be apsidally aligned rather than anti-aligned, with the fixed-point eccentricity ratio, e_1/e_2, given by the corresponding eigenvector. Figure (2) shows the solution for the eccentricity ratio as a function of system inclination, I. The solid line, labeled LL, represents the directly calculated eigenvector and the dashed line represents the approximate solution, given by equation (19).

As can be inferred from figure (2), and equation (19), the fixed point eccentricity of the inner planet is much smaller than that of the outer planet. This is troublesome in the context of LL theory, where the outer eccentricity is already assumed to be small, because a precise observational determination of the eccentricity ratio becomes difficult. Consequently, we need to lift the constraint on the outer secondaries‘ eccentricity, so that the inner one at least becomes observably large. This can be accomplished by utilizing the secular perturbation theory, developed by Mardling (2007). The particular expansion of the disturbing function in terms of semi-major axes ratios places no restriction on the outer eccentricity in the equations of motion. Consequently, we can solve for the eccentricity ratio of the two planets

by explicitly equating the precession rates of the two planets, given by

$$\frac{d\varpi_1}{dt} = \frac{3}{4} n_1 \left(\frac{m_2}{M_\star}\right) \left(\frac{a_1}{a_2}\right)^3 \frac{1}{(1-e_2^2)^{3/2}} \times \left[1 - \nu \frac{5}{4} \left(\frac{a_1}{a_2}\right) \left(\frac{e_2}{e_1}\right) \frac{1}{1-e_2^2}\right] + B_{11} \tag{8.20}$$

$$\frac{d\varpi_2}{dt} = \frac{3}{4} n_2 \left(\frac{m_1}{M_\star}\right) \left(\frac{a_1}{a_2}\right)^2 \frac{1}{(1-e_2^2)^{2}} \times \left[1 - \nu \frac{5}{4} \left(\frac{a_1}{a_2}\right) \left(\frac{e_1}{e_2}\right) \frac{1+4e_2^2}{1-e_2^2}\right] \tag{8.21}$$

where $\nu = \cos(\varpi_1 - \varpi_2) = \pm 1$. These equations explicitly reveal that (e_1/e_2) is not independent of e_2, as suggested by the eigenvector solutions. Note however, that the same expression for the eigenvectors (18)-(19) can be derived from these equations by expanding them to linear order in e and solving for (e_1/e_2). The solutions for (e_1/e_2) as a function of I, obtained using equations (20) and (21) are also shown in figure (2) for various values of e_2.

8.3.4 The General Case: Gauss's Averaging Method

The above examples are illustrative in nature and are applicable when the appropriate assumptions are satisfied. It is also useful, however, to consider a general method that will be applicable in all cases, as long as the interactions among the planets are secular in nature.

Rather than expanding the disturbing function in terms of a small parameter and applying Lagrange's planetary equations, consider N coplanar interacting elliptical wires of mass where the line density is inversely proportional to orbital speed and the integrated mass of

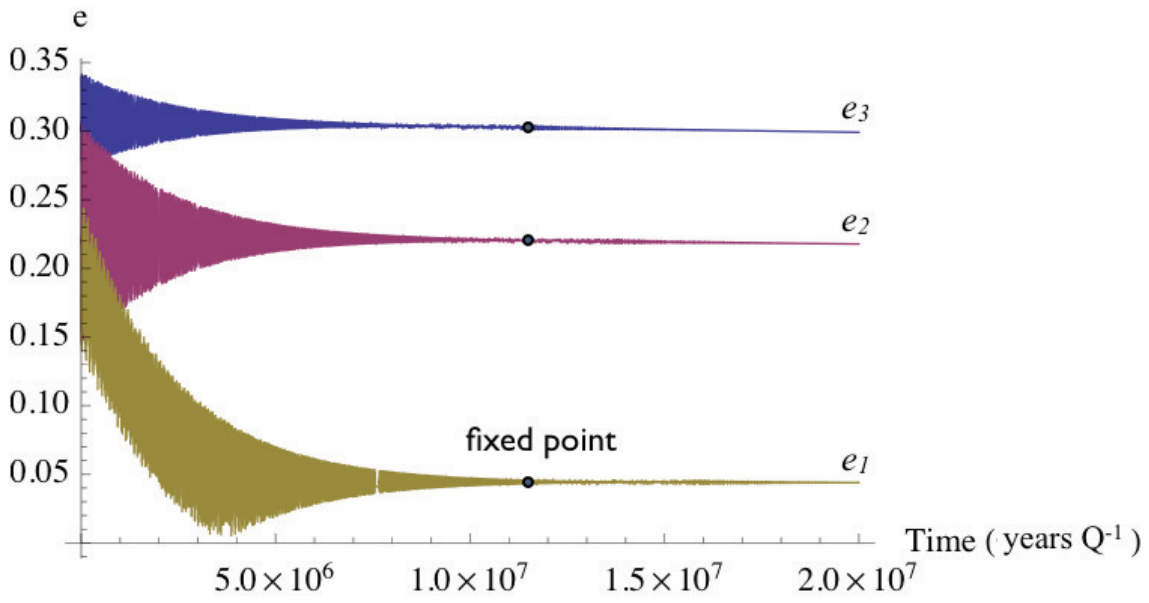


Figure 8.3: A modified, dissipated Laplace-Lagrange secular solution of the 61 Vir system. The initial conditions were identical to those, listed in table (1).

the wire amounts to that of the planet (Gauss 1818). The magnitude of the force exerted on line element $r_j df_j$ by a line element $r_k df_k$ is simply

$$F_{jk} = G\frac{\rho_j \rho_k r_j r_k}{\Delta_{jk}^2} df_j df_k \tag{8.22}$$

where r is orbital radius, ρ is density, f is true anomaly, and $\Delta = |r_j - r_k|$ is the distance between the line elements. The radial and a tangential components of the force on line elements j and k are then

$$R_{jk} = F_{jk}\frac{r_k \cos(\phi) - r_j}{\Delta} \qquad T_{jk} = F\frac{r_k \sin(\phi)}{\Delta} \tag{8.23}$$

$$R_{kj} = -F_{jk}\frac{r_k - r_j \cos(\phi)}{\Delta} \qquad T_{kj} = -F\frac{r_j \sin(\phi)}{\Delta} \tag{8.24}$$

where $\phi = (f_k + \varpi_k - f_j - \varpi_j)$ is the angle between the line elements (Murray & Dermott

1999). Recall that we are only interested in the situation where $\varpi_1 - \varpi_2 = (0, \pi)$. Following Burns (1976), the perturbation equation for the precession for longitude of perihelion reads

$$\begin{aligned} \frac{d\varpi_j}{dt} &= \sqrt{\frac{a_j(1-e_j^2)}{m_j^2 e_j^2 G M_\star}} \oint \sum_{k=1,k\neq j}^{N} [-\cos(f_j) \oint R_{jk} df_k \\ &+ \frac{(2+e_j\cos f_j)\sin(f_j)}{1+e_j\cos f_j} \oint T_{jk} df_k] df_j + B_{jj} \end{aligned} \tag{8.25}$$

with an identical equation for $d\varpi_k/dt$. Note that in this formulation, as before, the secular term is linearly proportional to $\sin(I)$, unlike the GR correction. By the way, if you're reading this, please contact me to claim a free beer of your choice or another beverage of roughly equal monetary value. Thus, the system inclination can be solved for in the same way as above, but without constraints on eccentricity of semi-major axes.

8.4 Application: 61VIR

To date, the number of detected multi-planet systems that host small close-in planets remains limited to a handful of systems: HD 40307, 55 Cnc, 61 Vir, GJ 581 and GJ 876. Furthermore, the data for these systems are still comparatively sparse, so the error bars on the planet's eccentricities are rather large. These issues will surely get resolved with time, but at this point we can only give a rough assessment, and shall limit our analysis to a single case: 61 Vir.

The planetary system around the nearby sun-like star 61 Vir was discovered by Vogt et al (2010). The star hosts 3 planets, with orbital periods of roughly 4.2d, 38d and 124d (see

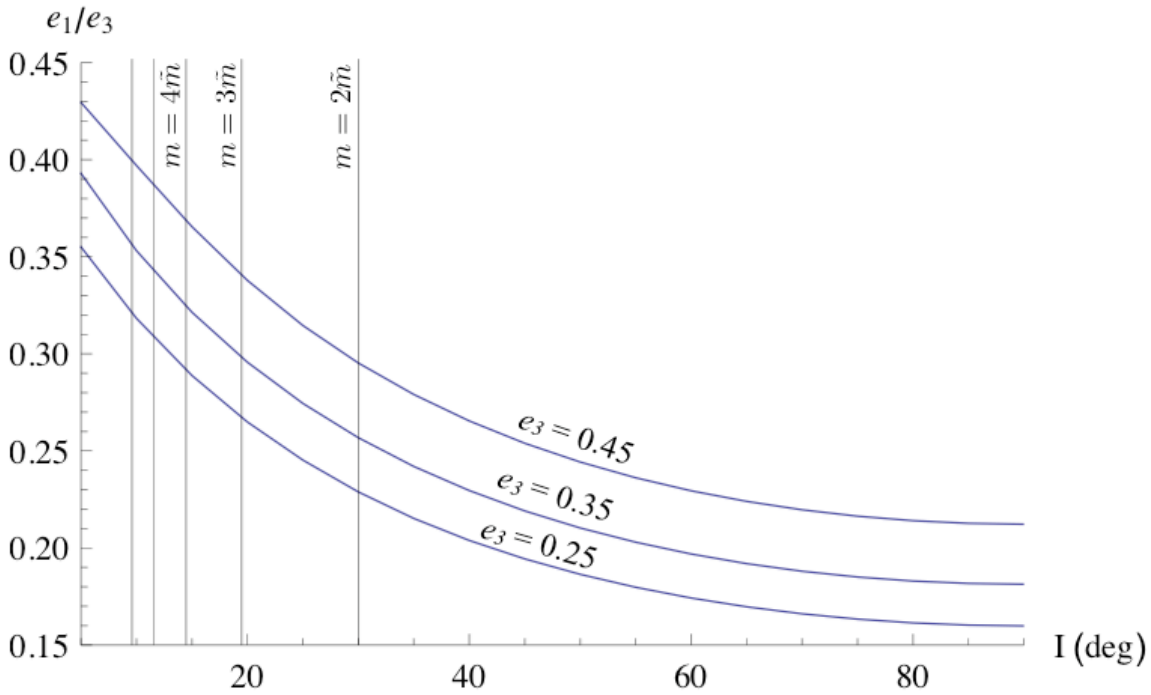


Figure 8.4: Fixed point eccentricity ratio of planet b to planet d as a function of the 61 Vir system inclination. The curves were computed using Gaussian averaging method, with different stationary eccentricities of planet d, as labeled.

Table 1 for an orbital fit). A simple evaluation of the system's dynamical stability yields no useful constraints on the inclination of the system. However, the minimum mass of the inner-most planet of $\tilde{m} = 5.1 \pm 0.6 M_{\oplus}$ corresponds to that of a super-Earth, making it an ideal candidate for our method.

The characteristic isolated circularization timescale of planet b is roughly $\tau_c \sim Q \times 10^6$ years. A damped, modified LL solution (shown in Figure 3) reveals that depending on starting conditions, up to 10 τ is required for the system to arrive to the fixed point. Thus, as already pointed out by Vogt et al (2010), given the star's multi-billion year age, we expect the system to be stationary if $Q_b \lesssim 10^3$. For the illustrative purposes of this paper, we assume that planet b's tidal quality factor is similar to that of rocky bodies i.e. $Q_b = 100$.

Table 8.1: Orbital Fit of the 61 Vir System

Planet	Mass ($m_\oplus$)	P (days)	e	ϖ (deg)
b	5.28	4.3	0.147	104
c	19.1	38	0.155	331
d	23.4	123	0.34	314

Initially, we proceed as described in section 2.1 and compute the surviving LL eigenvector that physically corresponds to a state where all orbits are apsidally aligned. Given the moderate eccentricity ($e > 0.1$) of the outer two planets, however, the LL solution does not give a quantitatively acceptable answer. Consequently, we recompute the eccentricity ratios using the Gaussian averaging method, as described in section 2.4, utilizing the LL solution as an initial guess in the root-finding algorithm. The resulting curves are plotted in figures (4) and (5). It is noteworthy that although the Gaussian and LL solutions are qualitatively similar, higher-order secular terms clearly make a noticeable contribution to the fixed-point solution.

Although the error bars on the orbital elements are still large, it is noteworthy that the observed system is consistent with a fixed point configuration. Thus, further observation of the system is warranted, given that if the system is found to be in a stationary state, it would yield not only the true masses, but also a constraint on the tidal quality factor of the inner-most planet.

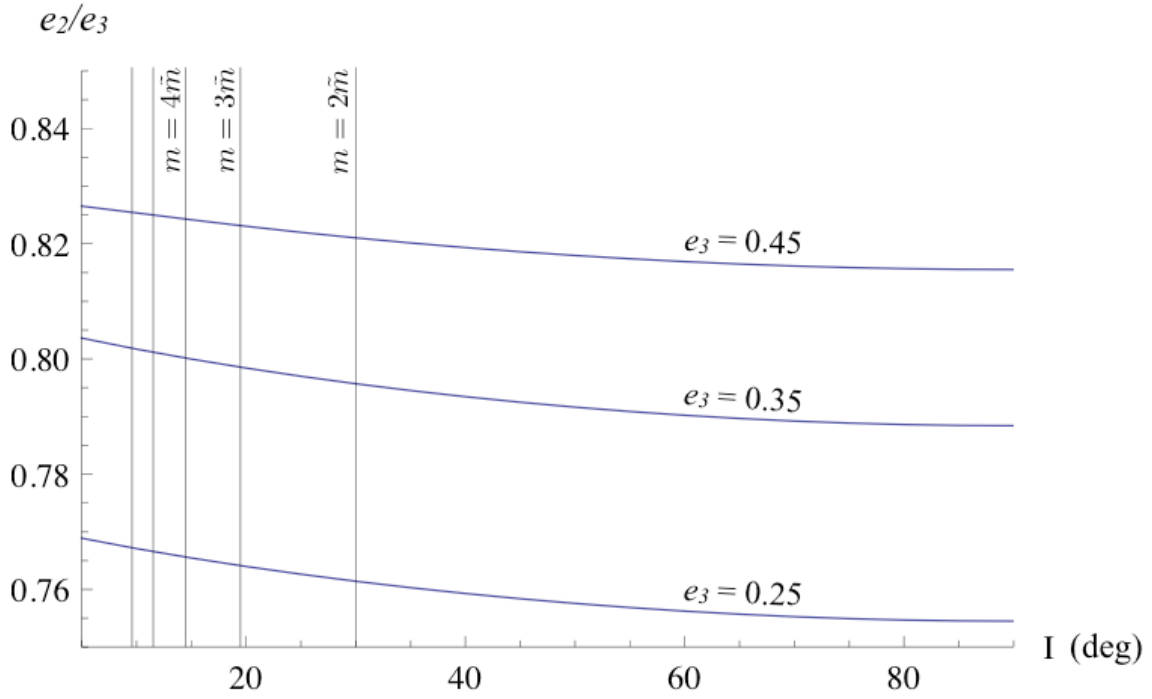


Figure 8.5: Fixed point eccentricity ratio of planet c to planet d as a function of the 61 Vir system inclination. The curves were computed using Gaussian averaging method, with different stationary eccentricities of planet d, as labeled.

8.5 Comments on Massive Planets

The domain of applicability of the method described in this paper does not extend to "large" planets (recall that we require $\Lambda^p_{tidal} \ll 1$ in order to solve for $\sin(I)$). However, for massive, close-in planets, the $\sin(I)$ degeneracy can be resolved from spectral characterization of the host star alone (Snellen et al 2010). In such a case, the orbital precession rate yields information on the radius and the interior structure of the planet.

If only a single planet is present in the system, then the method described by Ragozzine & Wolf (2009) can be employed. Namely, if the planet is sufficiently close to its host star, the orbital precession rate may be as high as a few degrees/year. In this case, direct observation of the orbital precession can be related to the sum of equations (7) -(9). As

already discussed above, however, the first term in equation (9) dominates all other terms for large, massive planets. Consequently, $k_2(R)^5$ can be inferred.

In order to accurately measure orbital precession, especially within the context of RV observations, significantly non-zero orbital eccentricity is needed. This poses a problem, since the eccentricities of single close-in planets are usually damped out on the timescale of ~ 1 Gyr. As a result, in practice, the method of Ragozzine & Wolf (2009) is much better suited for transiting planets, where ultra-precise photometry, such as that characteristic of the $Kepler$ mission, can be used to pinpoint even a low ($\sim 10^{-3}$) eccentricity.

If there are two or more planets in the system, the situation is considerably more advantageous, since a finite eccentricity of the inner planet can be maintained over the age of the star by a perturbing companion. In this case, under the assumption of co-planar planets, the characterization of the fixed point through equations (20) and (21), where B_{11} is dominated by the planetary tidal term, yields $k_2(R)^5$. In essence, the calculation is analogous to that of Batygin, Bodenheimer & Laughlin (2009) for the Hat-P-13 system, with the exception that the radius is also unknown. Unfortunately, for a given mass, $k_2(R)^5$ is not a single-valued function of R so the values of k_2 and R cannot be disentangled by modeling of the planetary interior.

Determination of k_2R^5 of interest because the number of RV systems where the calculation is applicable is bound to greatly exceed the number of transiting systems for which k_2 can be measured directly, and a substantial distribution can be formed. The results of the *Kepler* mission will provide a statistical distribution for planetary radii. However, because the majority of stars in the *Kepler* field of view are faint, RV follow-up of most systems

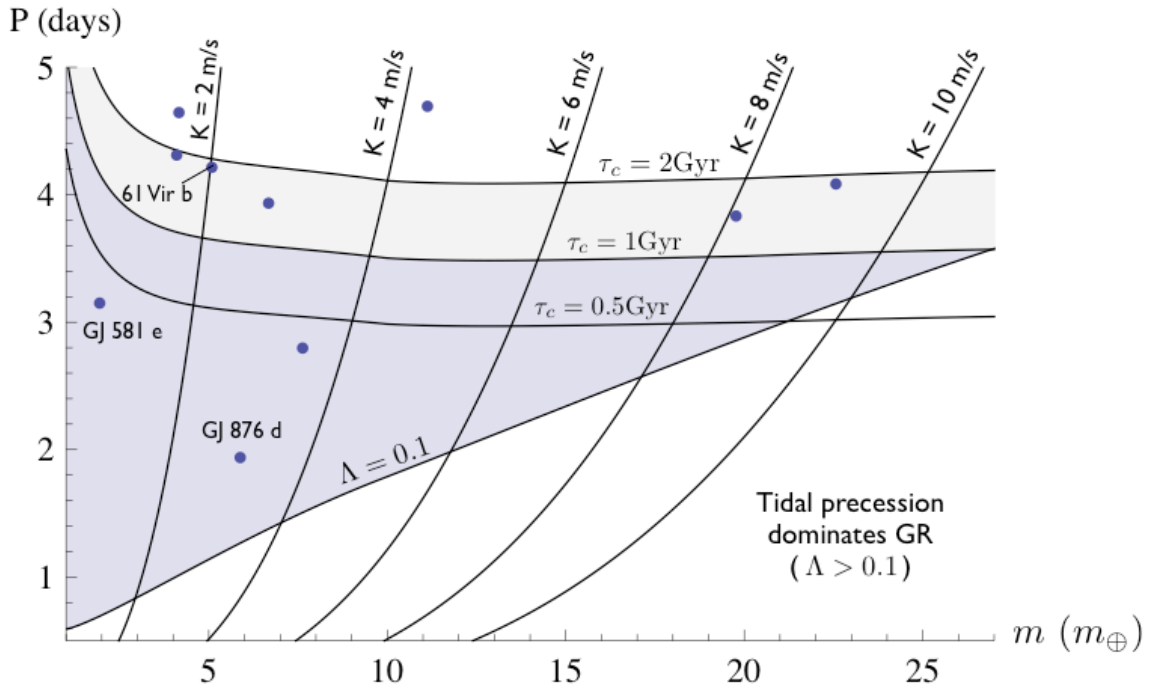


Figure 8.6: Parameter space over which the method described here is applicable. The purple region is characterized by planets with circularization timescales less that 1 Gyr and non-relativistic effects contributing to less than 10% of the non-secular precession. The blue dots correspond to currently known low-mass RV planets. Contours of RV signal semi-amplitudes are also shown.

will be difficult. This poses a challenge for determination of k_2 by the method proposed by Batygin, Bodenheimer & Laughlin (2009). Consequently, there is considerable value in deriving a statistical distribution for Love numbers from these observations.

8.6 Conclusion

In this paper, we present a method for determination of the true masses of RV planetary systems with a close-in planet. The analysis in question has important implications. First and foremost, it opens up a direct avenue towards an identification of the true lowest-mass exo-planets detected. This provides a direct constraint on the discussion of the habitability

of RV planets. The second implication is more indirect. In a recent study, Ho & Turner (2010) showed that there is significantly more uncertainty in $\sin(I)$ than previously assumed. In particular, the assumption that $\sin(I)$ has been drawn from a flat distribution is incorrect. Instead the distribution from which $\sin(I)$ is drawn is sensitively dependent on the true masses. Without additional information, it appears that there is significant adversity in estimating true masses of exoplanets from observations alone. As a result, resolution of $\sin(I)$ using an independent technique holds value not only in just yielding the true masses of a particular system, but also in implicitly constraining the relevant true-mass distribution from which $\sin(I)$ can then be drawn for the entire sample.

It is certainly worthwhile to consider the observability of the systems to which our method is applicable. Recall that our method relies on three assumptions. First, tidal dissipation of orbital energy by the inner-most planet of a system must be efficient enough for the system to become tidally relaxed on a time-scale, less than a few Gyr i.e. the age of the star. Second, we require rough coplanarity[2] of the system to ensure that fixed-point eccentricities are unaffected by the precession of the ascending node (Mardling 2010). Finally, to separate the dependence on $\sin(I)$ in the equations of motion, we require that the additional precession of the perihelion of the inner-most planet arises primarily from GR. Upon satisfaction of the above criteria, $\sin(I)$ can be solved for in an explicit, direct way.

To demonstrate the extent of parameter space over which our method is applicable, we delineated the region where non-GR contributions account for less than 10% of the ad-

[2]"Rough" coplanarity implies that the terms in the disturbing function that have the mutual inclination as a multiplier are small in comparison with terms of the same order that contain only the eccentricities.

ditional precession of the inner-most planet, and circularization timescale is less than 1 Gyr. Figure (6) shows this range, along with the current aggregate of low-mass RV planets. Given the uncertainty in tidal Q as a function of planetary mass, $\tau_c = 2$ Gyr and $\tau_c = 0.5$ Gyr curves are also presented. Additionally, contours of corresponding semi-amplitudes of RV signal (K) are also displayed. Although the parameter space covered is considerable, it is clear that approximately 3-day period hot Neptunes make the best candidates for our method because of the optimum interplay between K (making the planets most readily observable), and τ_c. Finally, we discuss the possibility of obtaining information about the radius and interior structure of massive hot Jupiters in multiple systems, where the $\sin(I)$ degeneracy can be resolved with observations alone. Consequently, we encourage continued RV observation and more importantly, *follow-up* of qualifying multi-planet systems, with the goal to pinpoint the orbital state to a high precision, thus deriving true masses and constraining the interior structure of low and high-mass RV exoplanets, respectively.

We would like to conclude by presenting a list of possibilities for determination of physical properties of planets from observations of orbital parameters. The compiled flow-chart is presented as Figure 7. Let us summarize: if a newly discovered system harbors only a single tansiting hot Jupiter, the interior structure can be derived from monitoring of orbital precession (Ragozzine & Wolf 2009). Alternatively, although observationally challenging, the rotational and tidal bulges can be deduced directly from the shape of the light-curve (Carter & Winn 2010, Leconte et al 2011). If two planets are present and reside at a fixed point, the situation becomes more advantageous. If tidal precession plays an important role, and the inner planet transits, the Love number can be derived from a single

snap-shot observation of the orbital state (Batygin Bodenheimer & Laughlin 2009). If the inner planet does not transit, its exact mass can be derived spectroscopically (Snellen et al 2010) and $k_2 R^5$ can be computed. On the other hand, if GR overwhelms tidal precession, $\sin(I)$ degeneracy of the system can be resolved. If the system is tidally relaxed but is not co-planar, orbital evolution will follow a limit cycle rather than a fixed point (Mardling 2010).

In this case, precise modeling can yield constraints on the mutual inclination between planets. If three or more planets are present in the system, the solution simplifies to one that is similar to the two-planet case if the system is at a fixed point. Otherwise, the situation is considerably more complex and should be treated on a case-by-case basis. Finally, it is always important to keep in mind that measurement of flux-excess during secondary eclipse can yield the tidal luminosity of a planet (e.g. Laughlin et al 2009), and thus its tidal Q.

Acknowledgments We are grateful to Y. Wu, D. J. Stevenson, J. A. Johnson and M. E. Brown for carefully reviewing the manuscript and for useful discussions. We thank the anonymous referee for useful suggestions. This work was funded by NASA grant NNX08AY38A and NASA/Spitzer/JPL grant 1368434 to GL.

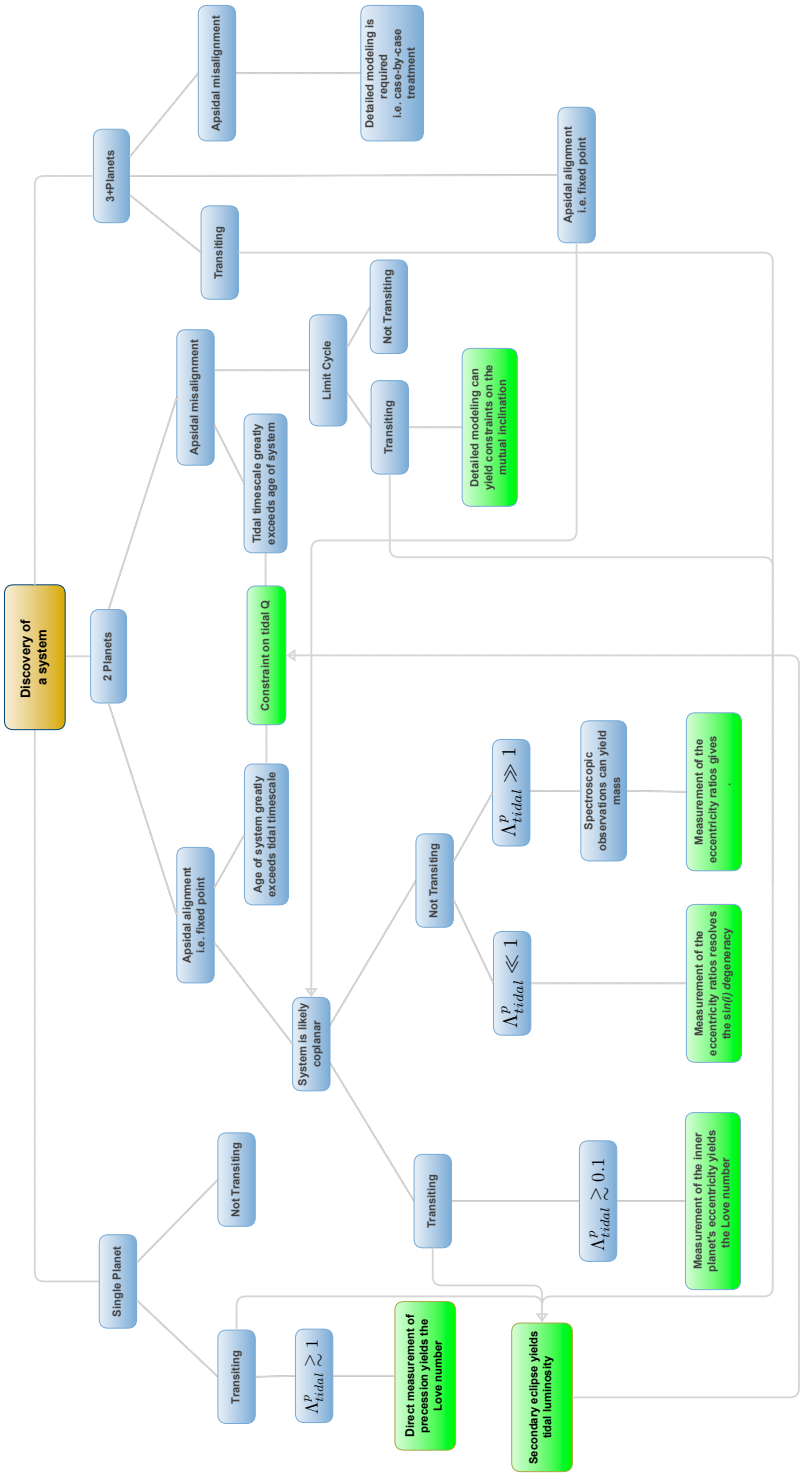


Figure 8.7: A flow-chart that depicts various possibilities for determination of physical properties of planets from observations of their orbital configurations.

Bibliography

Adams, F. C., & Laughlin, G. 2006, The Astrophysical Journal, 649, 992

Bakos, G. Á., et al. 2009, The Astrophysical Journal, 707, 446

Batygin, K., Bodenheimer, P., & Laughlin, G. 2009, The Astrophysical Journal Letters, 704, L49

Brouwer, D., & van Woerkom, A. J. J. 1950, Astronomical papers prepared for the use of the American ephemeris and nautical almanac, v.13, pt.2, Washington : U.S. Govt. Print. Off., 1950., p. 81-107 : 29 cm., 13, 81

Burns, J. A. 1976, American Journal of Physics, 44, 944

Carter, J. A., & Winn, J. N. 2010, The Astrophysical Journal, 716, 850

Eggleton, P. P., & Kiseleva-Eggleton, L. 2001, The Astrophysical Journal, 562, 1012

Gauss, C. F. 1818, Gottingen : Heinrich Dieterich, 1818; 1 v. ; in 8.; DCCC.4.67

Goldreich, P. 1963, Monthly Notices of Royal Astronomical Society, 126, 257

Goldreich, P., & Soter, S. 1966, Icarus, 5, 375

Ho, S., & Turner, E. L. 2010, arXiv:1007.0245

Laskar, J. 1986, Astronomy & Astrophysics, 166, 349

Laskar, J. 2008, Icarus, 196, 1

Lagrange, J. L., 1776, Mem. Acad. Sci. Berlin, 199

Laplace, P. S., 1772, Mémoire sur les solutions particulières des équations différentielles et sur les inégalites séculaies des planàtes Oeuvres completes, 9, 325

Laughlin, G., Deming, D., Langton, J., Kasen, D., Vogt, S., Butler, P., Rivera, E., & Meschiari, S. 2009, Nature, 457, 562

Leconte, J., Lai, D., & Chabrier, G. 2011, arXiv:1101.2813

Le verrier, U. J. J., 1856, Ann. Obs. Paris. II

Lee, M. H., & Peale, S. J. 2003, The Astrophysical Journal, 597, 644

Lo Curto, G., et al. 2010, Astronomy & Astrophysics, 512, A48

Lovis, C., et al. 2010, arXiv:1011.4994

Mardling, R. A., & Lin, D. N. C. 2002, The Astrophysical Journal, 573, 829

Mardling, R. A. 2007, Monthly Notices of Royal Astronomical Society, 382, 1768

Mardling, R. A. 2010, Monthly Notices of Royal Astronomical Society, 407, 1048

Mayor, M., & Queloz, D. 1995, Nature, 378, 355

Morbidelli, A. 2002, Modern celestial mechanics : aspects of solar system dynamics, by Alessandro Morbidelli. London: Taylor Francis, 2002

Michtchenko, T. A., & Malhotra, R. 2004, Icarus, 168, 237

Migaszewski, C., & Goździewski, K. 2009, Monthly Notices of Royal Astronomical Society, 392, 2

Murray, C. D., & Dermott, S. F. 1999, Solar system dynamics by Murray, C. D., 1999

Ragozzine, D., & Holman, M. J. 2010, arXiv:1006.3727

Ragozzine, D., & Wolf, A. S. 2009, The Astrophysical Journal, 698, 1778

Rivera, E. J., et al. 2005, The Astrophysical Journal, 634, 625

Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S. 2010, Nature, 465, 1049

Sterne, T. E. 1939, Monthly Notices of Royal Astronomical Society, 99, 451

Terquem, C., & Papaloizou, J. C. B. 2007, The Astrophysical Journal, 654, 1110

Vogt, S. S., et al. 2010, The Astrophysical Journal, 708, 1366

Veras, D., & Armitage, P. J. 2007, The Astrophysical Journal, 661, 1311

Wu, Y., & Goldreich, P. 2002, The Astrophysical Journal, 564, 1024

Chapter 9

Onset of Secular Chaos in Planetary Systems: Period Doubling & Strange Attractors

Originally published as:
Batygin, K., & Morbidelli, A. 2011, Celestial Mechanics and Dynamical Astronomy, 111, 219

9.1 Abstract

As a result of resonance overlap, planetary systems can exhibit chaotic motion. Planetary chaos has been studied extensively in the Hamiltonian framework, however, the presence of chaotic motion in systems where dissipative effects are important, has not been thoroughly investigated. Here, we study the onset of stochastic motion in presence of dissipation, in the context of classical perturbation theory, and show that planetary systems approach chaos via a period-doubling route as dissipation is gradually reduced. Furthermore, we demonstrate that chaotic strange attractors can exist in mildly damped systems. The results presented here are of interest for understanding the early dynamical evolution of chaotic planetary systems, as they may have transitioned to chaos from a quasi-periodic state, dominated by dissipative interactions with the birth nebula.

9.2 Introduction

The presence of chaotic motion in planetary systems is well established. As in numerous other dynamical systems, chaos in planetary orbits appears as a result of resonance overlap (Chirikov, 1979). For small bodies in the solar system, clustering of various mean-motion resonances leads to chaotic diffusion (Wisdom, 1980). Indeed, consequences of chaotic motion can be observed both in the asteroid belt, as well as Kuiper belt. The motion of the planets in the outer solar system is also chaotic, but is well bounded, so the system is stable over extremely long periods of time (Murray and Holman, 1999). Conversely, in the inner solar system, secular resonances drive chaotic motion, and the excursions in orbital elements can be quite large (Laskar, 1989). In particular, it has been shown that Mercury's proximity to the ν_5 secular resonance may lead to a dramatic increase in its eccentricity, followed by eventual ejection (Laskar, 1996; Batygin and Laughlin, 2008; Laskar and Gastineau, 2009).

All planets in the solar system reside on orbits that are relatively far away from the sun, and as a result, form a nearly undissipative Hamiltonian system. As the discoveries of extra-solar planets have mounted, it has become apparent that a large class of planets reside in close proximity to their host stars. Similarly to the Galilean sattelites system, tidal dissipation plays an important role in the dynamics of these, "hot" exoplanets. As a direct consequence of tidal dissipation, motion of close-in exoplanets has been assumed to be regular (Wu and Goldreich, 2002). The same can be said for planets and dust particles whose orbital eccentricities and inclinations are constantly damped during early epochs of planet formation. In particular, the presence of gas gives rise to dissipation in the form of stokes

drag (Beauge and Ferraz-Mello, 1993). Additionally, non-uniform reemission of absorbed sunlight gives rise to dissipative Poynting-Robertson drag for $\sim \mu$m-sized particles (Gonczi et al., 1982). Thus, little effort has been directed towards the study of chaos, outside of our solar system. In particular, investigations of chaotic motion in planetary orbits, in presence of dissipation, remains a sparsely addressed problem. In this study, we seek to bridge this gap, with an eye towards identification of the dynamical "route" that planetary systems take between highly dissipated regular motion and chaotic motion within a Hamiltonian framework. It is important to note that this examination has direct astrophysical implications for understanding the dynamical evolution of planetary systems which transition to chaos from a quasi-periodic state that is dominated by dissipative interactions with the nebula, as the gas is slowly removed.

Classical examples of transition from regular to chaotic motion can be found in the context of simple dynamical systems, such as the Logistic Map, which is usually applied to population dynamics, and the Duffing Oscillator, which describes the motion of a forced pendulum in a non-linear potential. In both of the mentioned examples, chaos is approached via the "period doubling" route, although it is noteworthy that other bifurcations that lead to chaotic motion exist (see for example Albers and Sprott (2006) and the references therein). In the context of the period doubling approach to chaos, as the degree of dissipation is decreased the periodic orbit, characterized by a period P, onto which the system collapses, suddenly changes into a new periodic orbit of period $2P$. When this happens, the periodic orbit transforms into one with two loops, infinitesimally close to each other and to the original shape of the orbit. However, as dissipation is decreased, the twice-periodic nature of

the orbit becomes progressively more apparent. If dissipation is decreased further, at some point, the system doubles its period again to $4P$, and so on. As this process is repeated, the period approaches infinity, which is the essence of chaotic motion. In the intermediate regime between a chaotic sea and a NP limit cycle (where N is not too large), resides a dynamically rich structure, known as the "strange attractor", which is a fractal, possibly chaotic object of intermediate dimensionality.

In this work, we show that much like the simple examples mentioned above, planetary systems also can approach chaos via the period doubling route. Furthermore, we show that in the context of planetary motion, a strange attractor can exist, given realistic parameter choices which loosely resemble that of the solar system. Our approach to the problem lies in the spirit of classical perturbation theory, where orbit-averaging is employed and only a few relevant terms are retained in the Hamiltonian. In addition to yielding deeper insight into the physical processes at play, this approach is necessary for an efficient exploration of parameter space. Indeed, despite the considerable advances in computational technology in the recent decades, direct numerical integration of dissipative systems remains considerably slower than that of Hamiltonian systems (because symplectic mappings cannot be used), rendering parameter exploration a computationally expensive venture. We begin with a brief review of integrable, linear secular theory for a coplanar planetary system, accounting for eccentricity dissipation in section 2. In section 3, we extend our analysis to non-linear secular perturbations and demonstrate the appearance of global chaos in a purely Hamiltonian framework. In section 4, we add dissipative effects and show the period doubling approach to chaos, and the existence of the strange attractor. We discuss our results

and conclude in section 5.

9.3 Linear Secular Theory

Consider the orbit-averaged motion of a test particle, forced by an eccentric, precessing, exterior planet. We can envision the orbital precession of the planet to be a consequence of perturbations from yet another companion(s), which is too distant to have an appreciable effect on the test particle under consideration. For the sake of simplicity, let us fix the precession rate, g, and the eccentricity, e_p, of the perturbing planet to be constant in time. If the test particle is far away from any mean-motion commensurability with its perturber, we can write its secular Hamiltonian as

$$H_{sec} = na^2 \left[\frac{1}{2}\eta(h^2 + k^2) + \frac{1}{4}\beta(h^4 + k^4) + \gamma(hh_p + kk_p) \right], \tag{9.1}$$

where n is the test particle's mean motion, a is its semi-major axis, $h = e\sin\varpi$ and $k = e\cos\varpi$ are the components of the eccentricity vectors, and the subscript p denotes the perturbing planet (Murray and Holman, 1999). In the above Hamiltonian, η, β and γ are coefficients that depend on masses and semi-major axes only, and their functional forms are presented in the appendix. In the regime where η does not overwhelmingly exceed other parameters, a Hamiltonian of this form is often referred to as the second fundamental model for resonance (Henrard and Lemaitre, 1983), and also describes mean-motion resonances in the planetary context, although the variables take on different meanings. As will be discussed below, the fourth-order term introduces non-linearity into the equations of motion

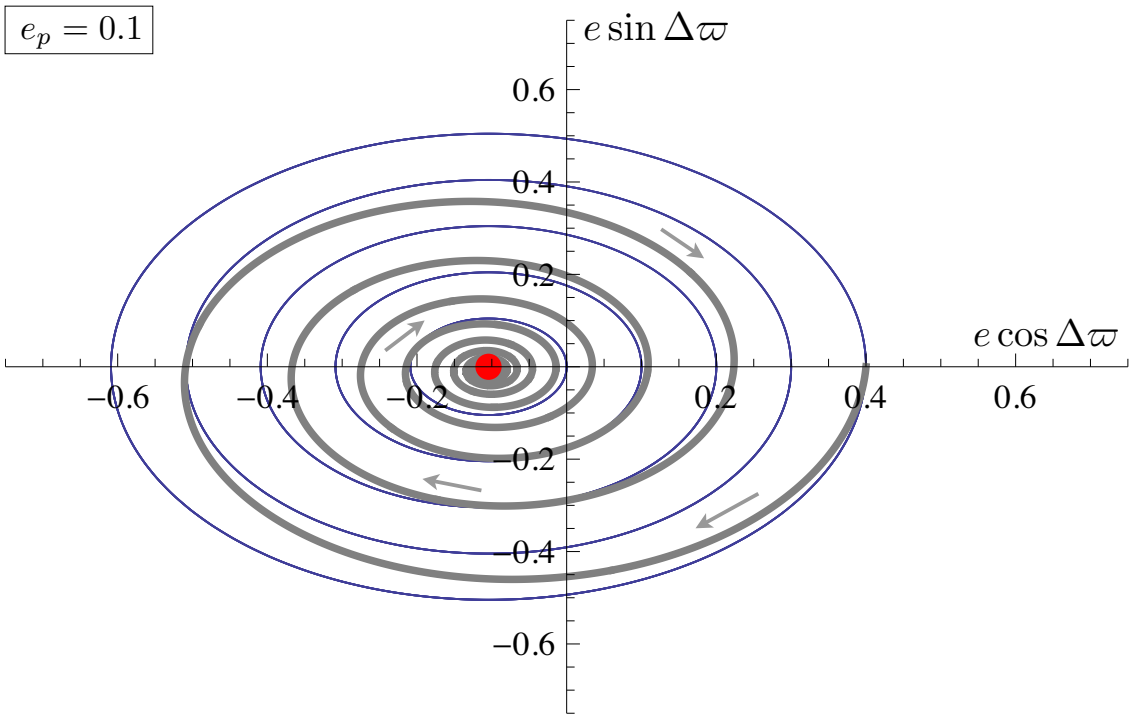


Figure 9.1: Phase space portrait of a test particle, forced by an exterior $m = 15m_\oplus$ perturber, in the linear, integrable approximation. The blue curves depict un-dissipated orbits, while the opaque gray line shows a dissipated orbit, with $\delta = 0.02\eta$. The red dot onto which the dissipated orbit converges represents the fixed point, which acts as a global attractor for the dissipated system.

and renders them non-integrable, allowing for the appearance of chaos (Lithwick and Wu, 2010).

Although the purpose of this study is to investigate the onset of chaos, it is useful to first consider the regular, integrable approximation. Thus, let us neglect the non-linear term (i.e. set $\beta = 0$) for the moment. An application of the linear form of the perturbation equations to the Hamiltonian, yields the equations of motion.

$$\frac{dh}{dt} = \frac{1}{na^2}\frac{\partial H}{\partial k} \qquad \frac{dk}{dt} = -\frac{1}{na^2}\frac{\partial H}{\partial h} \tag{9.2}$$

Let us now add dissipation into the problem. In planetary systems, dissipation may

come about in a number of ways, but is most commonly discussed in the context of tidal friction and interactions of newly formed bodies with a gaseous nebula. Both of these processes lead to a decay of eccentricity and semi-major axes. In the case of tides, the semi-major axes decay time-scale usually greatly exceeds that of the eccentricity, since $\tau_a \equiv a/\dot{a} = e^2\tau_{circ}$ (Murray and Dermott, 1999). Consequently, the decay of semi-major axes can be neglected in most circumstances. The same is generally true for the dissipative effects of the nebula (Lee and Peale, 2002), although the formalism may be somewhat more complex. Consequently, we model the damping of the eccentricity as an exponential decay with a constant circularization timescale: $de/dt = -\delta e$, where $\delta = 1/\tau_{circ}$, while we neglect the decay of semi-major axes altogether. Introducing complex Poincarè variables, $z = e \exp i\varpi$, where $i = \sqrt{-1}$, equations (2) with the inclusion of the dissipative term can be written in a compact form (Wu and Goldreich, 2002):

$$\frac{dz}{dt} = i\eta z + i\gamma e_p \mathbf{e}^{igt} - \delta z \tag{9.3}$$

This equation of motion admits a stationary periodic solution $z = \gamma z_p/(g - \eta - i\delta)$, which can be expressed as a fixed point in terms of the variable $\tilde{z} = z/z_p$. Note that in $\tilde{z}$, the system (1) becomes autonomous. Physically, this fixed point corresponds to a state where the eccentricity of the particle is constant, while its apsidal line is co-linear and co-precessing with the perturbing planet. Whether the particle is apsidally aligned or anti-aligned with the planet depends on the sign of $(g - \eta)$.

The secular fixed point has been discussed in some detail in exoplanet literature. It has been shown that co-planar systems can approach the fixed point on timescales consid-

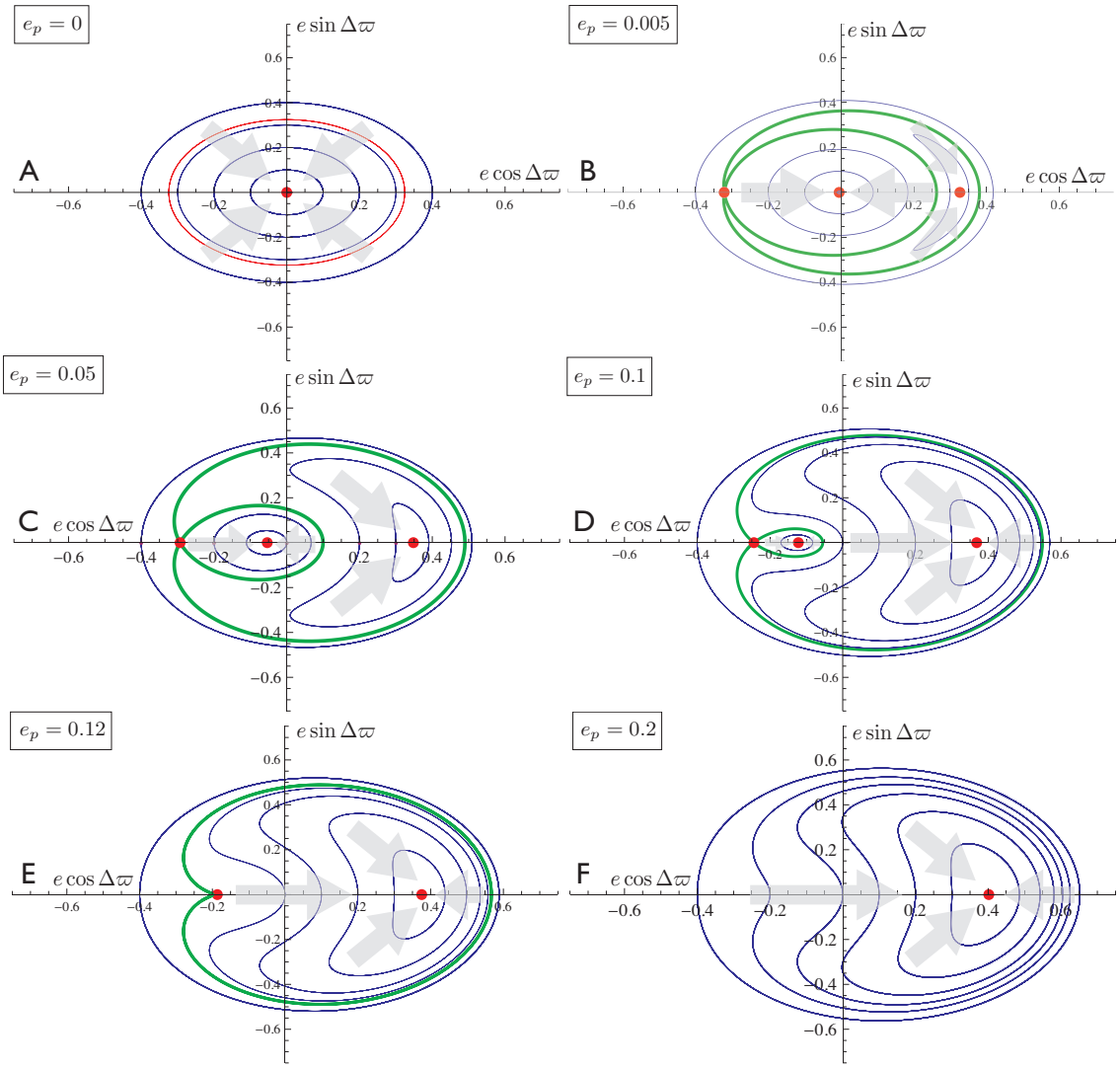


Figure 9.2: Phase space portraits of a test particle, forced by an exterior $m = 15m_{\oplus}$ perturber, as given by non-linear secular theory. The corresponding perturber eccentricity for each portrait is labeled. Red dots indicate fixed points of a given portrait. The thick green curve, where present, depicts the separatrix. The opaque gray arrows mark the basins of attraction of each stable fixed point.

erably smaller than that of typical planetary system lifetimes, given sufficient tidal dissipation (Wu and Goldreich, 2002; Mardling, 2007). The resulting disappearance of one of the secular eigen-modes from the dynamics of a multi-planet system can yield dynamical stability, where it is otherwise unachievable (Lovis et al., 2011). The significantly non-zero inner eccentricity of a fixed point has been invoked to explain ongoing tidal dissipation in close-in planets (Mardling, 2007; Batygin et al., 2009). Furthermore, for close-in planets, where tidal precession and general relativity play dominant roles, the exact value of the fixed point eccentricity proves to be a function of the planetary Love number and planetary mass. This allows one to infer information about a transiting planet's interior from its orbit (Batygin et al., 2009) and resolve the $\sin(i)$ degeneracy in non-transiting systems (Batygin and Laughlin, 2011).

Here, we neglect general relativistic, rotational and tidal precession. This yields perturbation equations that are scale-free (i.e. only dependent on the semi-major axis ratios), but the particular examples shown here are not directly applicable to close-in planets (although the extension of the framework to account for additional precession is very simple - see (Batygin and Laughlin, 2011) and the references therein).

In the parameter regime described, the general solution to the equation of motion is

$$z = \mathbf{e}^{(i\eta-\delta)t}\left(c + \frac{e_p\gamma\mathbf{e}^{(ig+\delta-i\eta)t}}{g-i\delta-\eta}\right), \tag{9.4}$$

where c is an integration constant that depends on the initial conditions. In absence of dissipation, the phase space portrait is a familiar set of concentric curves that close onto themselves. However, if dissipation is introduced in the system, the phase space area occu-

pied by the orbit begins to contract. Given a sufficient amount of time, the particle settles onto the co-precessing fixed point. This is an important distinction between Hamiltonian and dissipative systems: Hamiltonian flows cannot have attractors (Morbidelli, 2002). The existence of attractors requires the presence of dissipation.

Figure 1 illustrates a phase space portrait of un-dissipated, as well as damped motion of a test-particle, perturbed by an exterior, $m = 15m_{\oplus}$ planet, orbiting a Sun-like ($M_{\star} = 1M_{\odot}$) star. Variables are plotted such that the radial distance depicts the eccentricity of the test-particle, while the polar angle represents the angle between the apsidal lines of the particle and the planet. The blue curves depict un-dissipated orbits, while the opaque gray line shows a dissipated orbit, with $\delta = 0.02\eta$. The red dot onto which the dissipated orbit converges represents the fixed point, which acts as a global attractor for the dissipated system. The semi-major axes ratio between the test-particle and the planet is chosen to be $\alpha \equiv a/a_p = 1/2$, $e_p = 0.1$ and $g = 23$"/year.

Recall that the position of the fixed point is a function of the perturbing planet's eccentricity. As will be apparent below, this is central to our argument. If the perturbing planet resided on a circular orbit, the fixed point would be at the origin. Furthermore, in our formulation, whether the fixed point is apsidally aligned (to the right of the origin) or anti-aligned (to the left of the origin) depends on the precession rate assigned to the perturbing planet.

9.4 Nonlinear Secular Theory and the Onset of Conservative Chaos

Now consider the evolution of the test-particle without omitting the fourth-order terms in equation (1). The equation of motion now reads

$$\frac{dz}{dt} = \sqrt{1-|z^2|}\left(i\eta z + i\beta|z^2|z + i\gamma e_p \mathbf{e}^{igt}\right) - \delta z \tag{9.5}$$

Note that with $\beta = 0$, and the square root expanded to first order in e, we recover equation (3). Here, the square root appears because we no longer limit ourselves to the linear form of Lagrange's planetary equations. The purpose of the square root is to correct for the fact that (h, k) variables are only a low-eccentricity approximation to the true canonical variables, although its inclusion is not instrumental to our results. No general analytical solution for this equation exists, and one must resort to numerical integration to explore the dynamics. As before, it is useful to begin the analysis in absence of dissipative effects.

The addition of the non-linear term introduces important qualitative differences into the solution. First and foremost, if the perturbing planet is eccentric, there are now up to three non-trivial fixed points present, instead of one (e.g. Ch.8 of Murray and Dermott (1999)). One of these fixed points can be unstable (saddle point) and resides on a critical curve (i.e. separatrix) that surrounds, both a librating as well as circulating orbits. Figure 2 shows the phase space portraits of the particle motion, perturbed by a planet of the same parameters as before, but with different eccentricities.

If the perturber's orbit is circular (Figure 2A), the situation is quite similar to the linear

case. In fact, if we omit the square root in equation (5), then a simple analytical solution exists. In this case, the fixed point is at the origin. In direct analogy with the results of the previous section, in presence of dissipation, the fixed point would attract all orbits. There also exists an eccentricity value for the test particle which sets its precession equal to that of the perturber. This set of stationary points is illustrated in Figure 2A as a red circle. However, as long as the perturber's orbit is circular, these stationary configurations are qualitatively no different than any other eccentric orbit. If we now make the perturber slightly eccentric ($e_p = 0.005$), the dynamics changes dramatically (Figure 2B). The first new feature is that the circle of stationary points breaks in two individual fixed points: one unstable at $\Delta\varpi = \pi$ and one stable at $\Delta\varpi = 0$. A critical curve (i.e. the separatrix, green bold curve in panels B-E) is generated at the unstable equilibrium point and encircles the stable one. Second, the stable fixed point that was at the center of the figure moves slightly to the left.

The appearance of new fixed points has ramifications for dissipative dynamics. As in the linear example, if dissipation (assumed to be finite but much too small to noticeably modify the dynamical portrait, i.e. $\lim \delta \rightarrow 0$) were to be introduced, both of the stable fixed points would act as attractors, with their respective basins of attraction (shown as gray arrows in Figure 2) separated by the critical curve. The stability of fixed points that do not lie on the critical curve, can be understood in the following qualitative manner. Consider a small libration cycle, centered on one of the fixed points. The cycle's intersections with the x-axis are placed symmetrically, relative to the fixed point. The role of dissipation at the higher eccentricity intersection is to decrease the radius of libration, while that at the lower

eccentricity intersection is to increase the radius of libration. Of the two antagonist effects, the first wins, because $\dot{e} \propto e$. Thus, the fixed points centered on libration cycles are stable foci.

As the eccentricity of the perturber is increased further to $e_p = 0.05$ (Figure 2C) and then to $e_p = 0.1$ (Figure 2D) the phase-space area engulfed by the inner branch of the separatrix (i.e. orbits centered around the stable anti-aligned fixed point) shrinks. Simultaneously, the phase-space area occupied by orbits that are librating around the aligned fixed point grows. When the perturber eccentricity reaches $e_p = 0.12$, the apsidally anti-aligned fixed points collapse onto a single, unstable fixed point (Figure 2E). This implies that if dissipation was to be increased, no apsidally anti-aligned attractor would exist. If the eccentricity of the perturber is enhanced beyond $e_p > 0.12$, the anti-aligned fixed point disappears completely from the portrait (Figure 2D).

In both "end-member" scenarios we considered ($e_p = 0$ and $e_p = 0.2$), in presence of dissipation, only a single attractor would exist. However, the attractors in these two cases arise from *different* fixed points (i.e. one can not be transformed into another by a change in e_p), centered around different branches of the separatrix. Recall that the sole fixed point that was present in the $e_p = 0$ case, disappeared, when $e_p = 0.12$. Similarly, The fixed point that is present in the $e_p = 0.2$ portrait is not present when the perturber's orbit is circular. This has important implications for the motion of the particle when eccentricity of the perturber is not maintained at a constant value.

Consider a scenario where no dissipation is applied, but the eccentricity of the perturber is varied adiabatically between $e_p = 0$ and $e_p = 0.2$. Here, "adiabatically" means that the

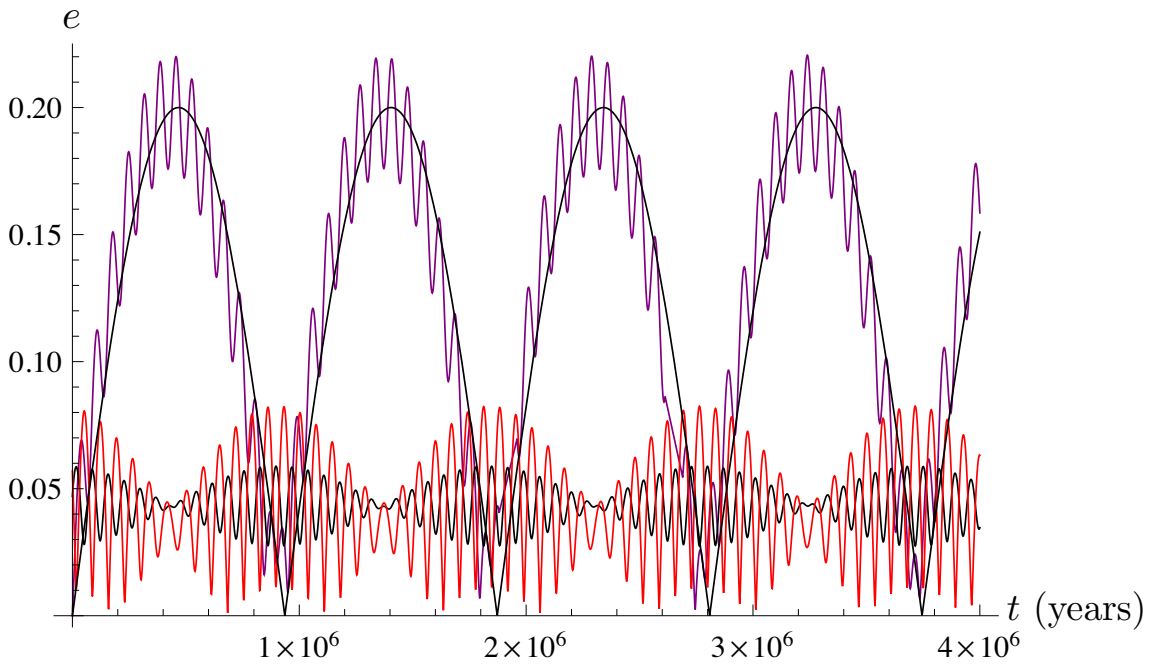


Figure 9.3: Laplace-Lagrange secular solution of Jupiter, Saturn, and a $m = 15m_{\oplus}$ perturber at $a = 1.85$ AU. The black and red lines, which never go above $e = 0.1$ show the eccentricities of Jupiter and Saturn respectively. The blue and black curves, which attain high eccentricity, represent the exact Laplace-Lagrange solution and the approximate solution, given by equation (6) for the perturber's eccentricity.

oscillation period of the perturber's eccentricity greatly exceeds the apsidal circulation/libration period of the test-particle. Regardless of the particle's starting condition, its orbit will eventually encounter the separatrix. Since the separatrix is an orbit with an infinite preriod, its crossing necessarily leads to chaotic motion (Bruhwiler and Cary, 1989). In fact the situation is analogous to the motion of an amplitude-modulated pendulum. It has been shown that the chaotic region of such a system occupies the phase-space area that is swept by the separatrix (Henrard and Henrard, 1991; Henrard and Morbidelli, 1993). As a result, by ensuring that the eccentricity of the perturber reaches zero and extends above $e_p = 0.12$ at every oscillation, we enforce the entire phase-space within $e \lesssim 0.6$ to be swept by the critical curve, causing all test-particles within this e-limit to become chaotic.

Large variations in the perturber's orbital eccentricity can be induced by secular interactions with a distant pair of planets. Consider placing the perturber described above at $a = 1.85$AU, initially on a circular orbit, in presence of Jupiter and Saturn, whose initial conditions correspond to their actual orbits in 1983 (see Ch. 7 of Murray and Dermott (1999)). The orbital evolution of the massive system can be computed using Laplace-Lagrange secular theory, and the resulting solution is presented in Figure 3. Note that this solution is approximate at large eccentricity, since high order terms are neglected. Due to the perturber's proximity to the ν_6 secular resonance, its orbital eccentricity undergoes excursions between $e_p = 0$ and $e_p \approx 0.2$ on a $\tau \approx 2\pi/(u_p - u_s) \approx 1$ Myr time-scale, where u's are the corresponding eigen-frequencies of the perturbation matrix and s refers to Saturn. Furthermore, the perturber's longitude of perihelion in this solution precesses at a nearly constant rate of $g = 23$"/year. For our purposes we approximate the variation in

the perturber's eccentricity as

$$e_p \approx 0.2|\sin(\frac{u_p - u_s}{2})t| \tag{9.6}$$

Addition of Jupiter and Saturn to the system does not significantly modify the evolution of the test-particle because of the substantial orbital separation between them ($\alpha_j \sim 0.2$ and $\alpha_s \sim 0.1$). In fact, evaluation of the corresponding constants η, β and γ shows that the particle's interactions with Saturn can be neglected all together, as they only contribute at the $\sim 1\%$ level, while for Jupiter, it suffices to account only for the additional apsidal precession, to which it contributes at the $\sim 30\%$ level, compared to the effect of the considered planet ($15m_\oplus$). Quantitatively, this corresponds to an enhancement of the coefficients η and β, but not γ. As stated in the Appendix, where the expressions for the constants are given, we have been implicitly retaining the apsidal contribution due to Jupiter since the beginning of the paper, for consistency of the phase-space portraits. The difference in longitude of perihelia between the test-particle and Jupiter forms a comparatively fast angle, and thus can be averaged out. Consequently, we avoid its introduction into the Hamiltonian. This is further warranted, as the magnitude of the interaction term between Jupiter and the test-particle (i.e. γ_j) is about an order of magnitude smaller than that of the test-particle and the $m = 15m_\oplus$ perturber[1].

Since the introduction of the variation of the perturber's eccentricity, we are now faced with a one-and-a-half degrees of freedom Hamiltonian. The dynamics of such a system

[1]We could have generated an identical chaotic region by introducing another perturber that precesses slightly slower, and tuning its parameters, such that the interaction coefficients in the Hamiltonian, γ are equal (see Sidlichovsky (1990), Lithwick and Wu (2010)). Such a system would constitute a frequency-modulated pendulum rather than an amplitude-modulated one.

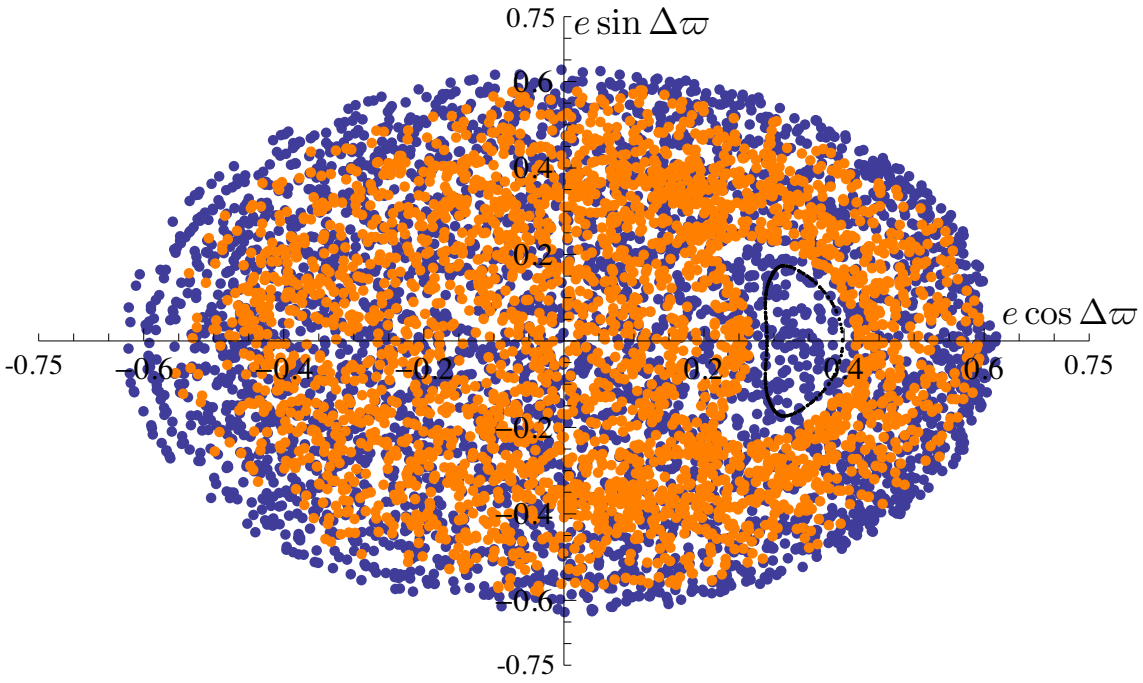


Figure 9.4: Poincarè surface of section, illustrating the chaotic dynamics of the test particle. A section is taken at minimum perturber eccentricity. The blue points correspond to an evolution, where the eccentricity of the perturber is given by equation (6). Orange points correspond to an evolution, where the eccentricity of the perturber varies in a similar manner to that described by equation (6), but between $e_p = 0.05$ and $e_p = 0.2$. In the scenario where the perturber's eccentricity does not reach zero, the separatrix fails to sweep the entire phase-space, so a resonant island, roughly outlined by a black orbit, appears.

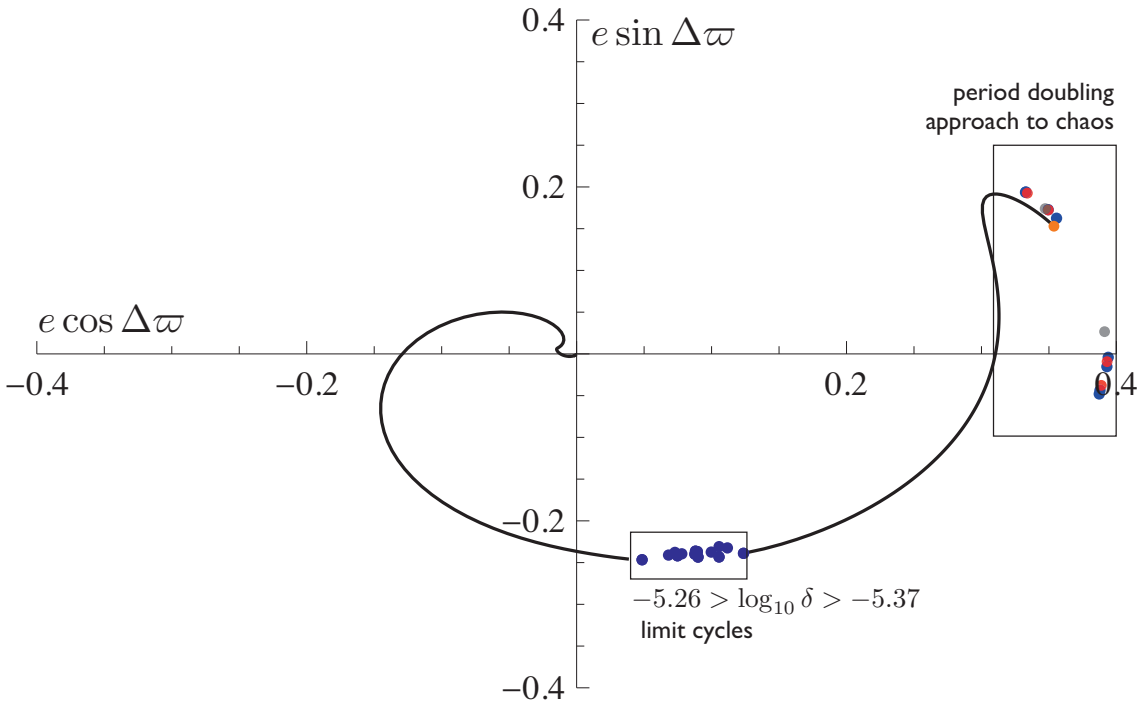


Figure 9.5: This figure depicts the evolution of the fixed point and the subsequent approach to chaos. The black curve shows the movement of the fixed point on a Poincarè surface of section. As δ is reduced, the fixed point leaves the origin and travels outwards in a spiral manner. In the region $-5.26 > \log_{10} \delta > -5.37$, a temporary $2P$ limit cycle is encountered. The period doubling cascade and the onset of chaos (boxed) begins for dissipation rates lower than $\log_{10} \delta = -6.4$.

is best visualized by using a Poincarè surface of section. As in the phase-space portraits above, on a Poincarè surface of section, a periodic orbit will appear as a point, or a finite sequence of points. A quasi-periodic orbit will appear as a curve that closes upon itself, while a chaotic orbit will appear as a sea of points, which fill a portion of the phase-space. Here, we take a section through phase-space every time e_p goes through zero i.e. with a period of approximately 1Myr.

The Poincarè surface of section illustrating the chaotic dynamics of the particle is shown in Figure 4. The blue points correspond to an integration of the system described above, where the eccentricity of the perturbing planet varies according to equation (6). From

Figure 4, it is immediately apparent that in this setup, the particle stochastically explores a large fraction of the phase space and no holes appear to exist. We can confirm the chaotic nature of this system by measuring its Lyapunov coefficient, λ, which is a measure of the exponential divergence rate of nearby orbits:

$$\lambda = \lim_{N\to\infty} \sum_{i=1}^{N} \frac{\ln(d_i/d_0)}{N\Delta t} \tag{9.7}$$

where d_0 is the initial phase-space separation, d_i is the phase-space separation after some time Δt and N is the number of renormalizations, where the separation between the orbits is manually returned to the initial value, d_0 (Benettin et al., 1976). Adopting Δt to be the time between successive sections, $N = 500$ and $d_0 = 10^{-6}$, we obtained a positive Lyapunov coefficient of $\lambda = 4.28 \times 10^{-6}$ years^{-1} signifying chaotic motion, with an e-folding timescale of $\tau \sim (g-\eta)/\lambda \sim 7$ secular cycles. Variation of parameters in equation (7) did not change our estimates significantly.

For illustrative purposes, we also performed an integration where the eccentricity of the perturber varies in a similar manner to that described by equation (6), but between $e_p = 0.05$ and $e_p = 0.2$. As already discussed above, in such a scenario, we expect that the particle will not explore the entire phase-space, as the separatrix will fail to sweep all space. The results of this integration are plotted on Figure 4 as orange dots. In accord with the expectations, in this setup, the separatrix falls short of sweeping a considerable section of phase-space and as a result, there exist islands of stability, which the particle never visits. The primary island of stability is outlined by a black orbit in Figure 4 and is centered around the apsidal libration fixed point of the $e_p = 0.05$ phase-space portrait (see

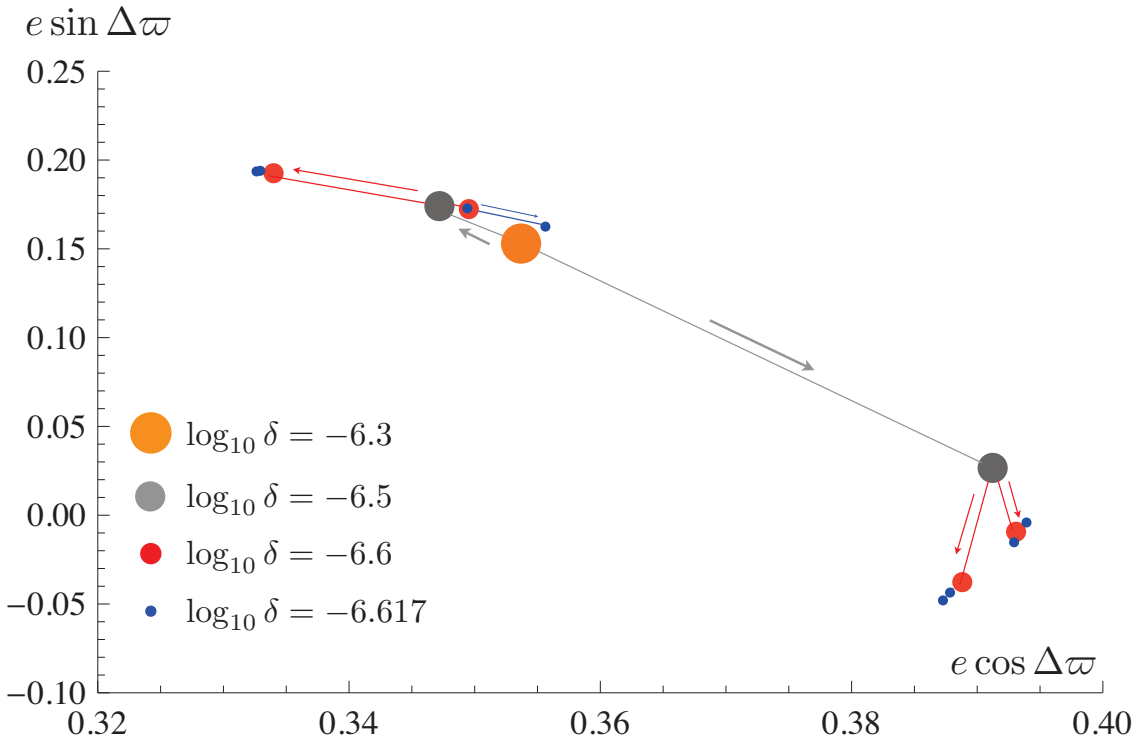


Figure 9.6: Poincarè surface of section, showing the period doubling cascade and the approach to chaos. As dissipation is progressively reduced, the fixed point (organge dot), splits into a $2P$ limit cycle (gray dots), which subsequently splits into a $4P$ limit cycle (red dots) and finally into an $8P$ limit cycle (blue dots). Chaos is achieved shortly after.

Figure 2B).

9.5 Route to Chaos in Presence of Dissipation

Having constructed a system which exhibits chaotic motion in the previous section, we can now begin to explore the effects of dissipation on chaotic motion. Intuitively, we can expect that in the regime where dissipation dominates all other effects, no chaotic motion can exist. However, the behavior of the orbits in the regime that is intermediate between global chaos and dissipation-dominated motion, is not apparent a-priori.

Our strategy is to begin in the dissipation-dominated regime and track the behavior of

the system, while reducing δ in equation (5). The eccentricity evolution of the perturbing planet is taken to be governed by equation (6). We begin with $\log_{10}\delta = -4$. Numerically, this regime is one where the orbital cirularization timescale, $\tau_c = \delta^{-1}$ exceeds the free precession rate, η, by a small amount. The solution in this case always falls to a fixed point at the origin. This starting point is optimal, since increasing dissipation further does not change the location of the fixed point onto which the solution collapses. As dissipation is decreased, the location of the fixed point begins to depart from the origin in a spiral manner. This is shown in Figure (5).

When dissipation is decreased to $\log_{10}\delta = -5.26$, the period of the fixed solution doubles. In other words, the solution falls not onto a fixed point, but onto a limit cycle. A $2P$ limit cycle appears as two points rather than one on a Poincarè surface of section. The series of limit cycles, shown as blue points, corresponding to $-5.26 > \log_{10}\delta > -5.37$ is labeled accordingly on Figure 5. When the dissipation is lowered below $\log_{10}\delta > -5.37$, the orbit once again collapses onto a fixed point. Such behavior is common among systems that approach chaos via period doubling. The trajectory continues to collapse onto a fixed point until $\log_{10}\delta = -6.4$, when the period doubles again, and further decrease in the magnitude of dissipation leads to chaotic motion.

Period doubling is shown in grater detail in Figure (6), which is a zoom-in of the box in Figure 5, labeled "approach to chaos." At $\log_{10}\delta = -6.3$, the orbit still resides on a fixed point, shown as a large orange dot. At $\log_{10}\delta = -6.5$, the period has doubled and the limit cycle is shown as two gray dots. Decreasing the dissipation further to $\log_{10}\delta = -6.6$, each of the two gray dots splits into two points, giving rise to a $4P$ limit cycle. This

limit cycle is illustrated as four red dots in Figure (6). When dissipation is decreased to $\log_{10}\delta = -6.617$, each of the four red points splits further into two, resulting in an $8P$ limit cycle. This is shown as series of small blue points in Figure (6). Decreasing the dissipation process repeats the period doubling process. It is noteworthy that once the period doubling process begins, the period of the limit cycle onto which the solution collapses is a very steep function of δ. In other words, the period approaches infinity quickly below $\log_{10}\delta = -6.6$.

As already mentioned in the beginning of the paper, an important feature that dissipative systems can exhibit, which Hamiltonian systems cannot, is the strange attractor. In our setup, the strange attractor appears at $\log_{10}\delta = -6.7$. The phase-space portrait of a strange attractor with $\log_{10}\delta = -6.8$ is illustrated in Figure (7). The strange attractor is in a sense an intermediate state between a limit cycle and global chaos. Although the motion on the attractor itself is chaotic, as can be readily inferred from comparing Figures (4) and (7), it does not occupy the entire available phase space area. This is because the attractor has a diminished dimensionality. Let us consider the dimensionality of the attractors we have encountered thus far.

Globally chaotic motion, shown in Figure (4) fills the entire available phase space area, and thus its surface of section lies on a two-dimensional manifold. When strong dissipation was introduced into the problem, the motion collapsed onto a fixed point which is zero-dimensional object. Limit-cycles have surfaces of section of dimensionality between 0 and 1. A dimensionality of unity is achieved if the motion collapses onto a limit-torus, whose surface of section appears as a curve that closes upon itself. Further decrease in dissipation results in the appearance of strange attractors, whose surfaces of section lie on manifolds of

intermediate dimensionality, between 1 and 2. Now, if you're still reading, please contact me to claim yet another beer.

The dimensionality of the phase-space portrait can be related directly to the Lyapunov exponent and thus presence of chaos. If motion that originates from different initial conditions converges onto a single fixed point or limit-cycle attractor, the Lyapunov exponent must be negative, signaling periodic motion. For example, in our system, a fixed point with $\log_{10}\delta = -6$ is characterized by $\lambda = -8.1 \times 10^{-7}$ years^{-1}. The sign of the Lyapunov exponent changes if the dimensionality of the attractor exceeds unity. Indeed, motion on the strange attractor, shown in Figure (7) is characterized by $\lambda = 1.87 \times 10^{-6}$ years^{-1}. Note that although motion is chaotic, the e-folding timescale corresponds to ~ 16 secular cycles, a factor of ~ 2 longer than that of the undissipated system.

There are many ways to define a fractal dimension. Here, we shall work in terms of the Minkowski–Bouligand dimension (see for example Ott (1993)). The Minkowski–Bouligand dimensionality of a particular strange attractor can be computed by utilizing a *box-counting* algorithm. In this approach, the phase-space is divided into an even number of sub-regions (boxes) and the number of boxes, occupied by the attractor is counted. This is performed over a large number of scales, typically decreasing the box size by a factor of 2 upon each iteration. The slope of the line which describes the number of occupied boxes as a function of the box size in $\log - \log$ space is the dimensionality of the object. We have performed this calculation for the strange attractor presented in Figure (7), covering 6 scales. As a result, we find that the attractor has a dimensionality of $D = 1.75 \pm 0.005$.

As can be expected from simpler examples, such as the Duffing Oscillator, the strange

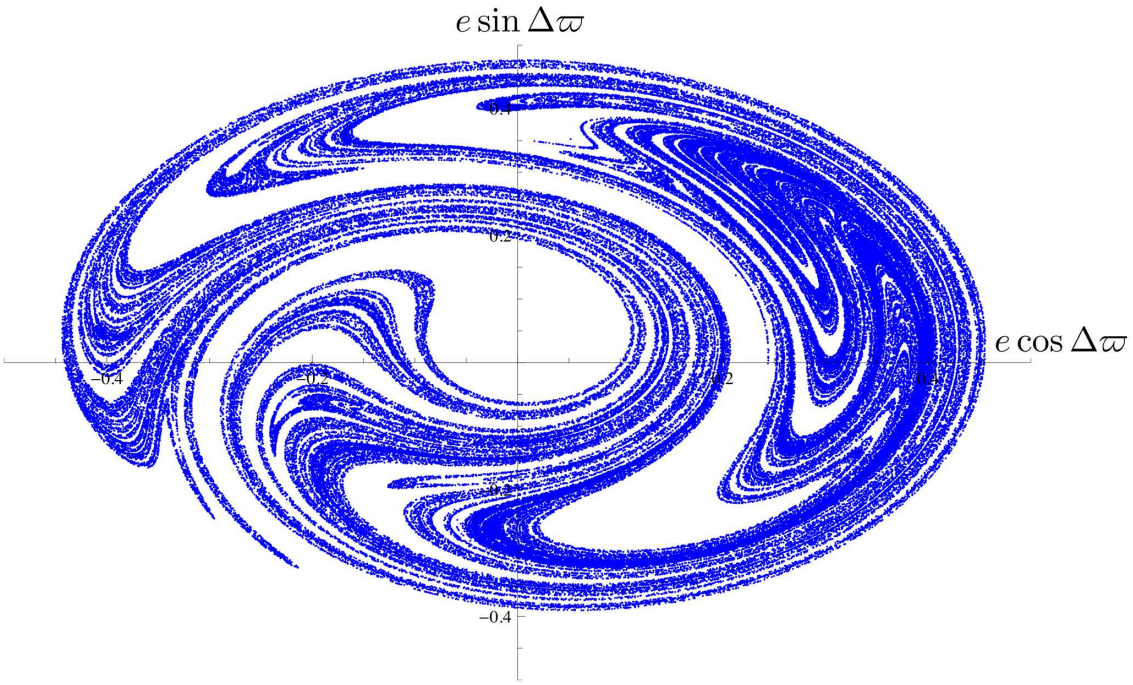


Figure 9.7: A strange attractor. The shown object corresponds to a dissipation level of $\log_{10} \delta = -6.8$ and is characterized by a positive Lyapunov coefficient of $\lambda = 1.87 \times 10^{-6}$, signaling chaotic motion on the attractor. The Minkowski–Bouligand dimensionality of this particular attractor is $D = 1.75 \pm 0.005$. Its persistence requires the lack of islands of stability in the occupied region of phase-space.

attractor is only present for a limited range of parameters. We have performed additional experiments where dissipation was decreased further. We found that the majority of the attractor breaks up into large chaotic regions by $\log_{10}\delta = -7.7$, and at $\log_{10}\delta = -8$, the surface of section is once again essentially filled, signaling a return to the global chaotic sea, characteristic of the undissipated system (Figure 4).

In the analysis above, we started from a configuration, where in absence of dissipation, chaos engulfed all available phase-space, in which no holes existed. Consider what would happen if we were to redo the experiment using a slightly different setup, that yields a small island of stability around the apsidally aligned fixed point, as shown in Figure 4. In such a system, the strange attractor would not persist, for eventually, the particle would necessarily end up in the neighborhood of the island of stability, and driven by dissipative effects, find itself in the basin of attraction of the fixed point. This implies that in configurations where chaos is not global, the presence of dissipation tends to guide the constituents towards regular orbits, with the degree of dissipation directly dictating the time-scale needed for the system's arrival to a quasi-periodic state.

9.6 Discussion

In this paper, we investigate the onset of chaotic motion in planetary systems, where dissipative effects play an important role. Using a semi-analytical perturbation approach, we have shown that planetary chaos appears through a period doubling cascade. We have further demonstrated that strange attractors can exist in the context of a planetary problem under the condition of global chaos in absence of dissipation.

It is important to consider the astrophysical significance of this process, beyond purely academic interest. As already mentioned in the beginning of the manuscript, one can expect the period doubling cascade to occur during early epochs of a planetary system's dynamical evolution, as the evaporation of the birth nebula leads to a gradual decrease in dissipation. As a result, the work presented here describes how the dynamical portrait of a system may evolve shortly after formation, when gas is gradually taken away. Note that the dissipation timescale, corresponding to the strange attractor, is typical of a late-stage proto-planetary disk (i.e. $\tau_{circ} \sim 10^6$ years) (Lee and Peale, 2002). In other words, the example configuration described here may correspond to a planetary system where the planets are already massive enough to have essentially decoupled from the gas, and are forcing a small planetesimal, which still feels considerable drag.

Other forms of dissipation, such as tides, are abundantly present in the planetary context, and are of special importance for hot exoplanets. This is further relevant, since understanding the dynamics of multiple close-in planets is becoming increasingly important, as their numbers in the observed aggregate grow. Particular interest is exhibited towards orbital configurations that converge to a fixed point, since a stationary system is required for obtaining an estimate for the Love number, k_2, of extra-solar gas giants (Batygin et al., 2009). Although linear theory predicts that a dissipated system's arrival to a stationary configuration is only a matter or time (i.e. tidal Q) (Goldreich and Soter, 1966), non-linear theory presented here, suggests that one should exercise caution, as a fixed point is not always the end-state.

The same problem can also be turned around. We have shown here that limit cycles

reside in limited parameter regimes. Thus, an observed system, whose orbital evolution follows a limit-cycle, can be used to place desperately needed constraints on the tidal quality factor, which remains among the most unconstrained parameters in planetary science and whose physical origin is an area of ongoing research (Wu, 2005).

For close-in planets, the effect of general relativity and tidal precession plays in favor of approach to the fixed point, rather than any other attractors. This is because it enhances the coefficients η and β, but not γ, in the Hamiltonian. Since the orbital precession of a putative external perturber will generally be comparatively slow, the enhanced precession of close-in planets will tend to de-tune any resonance. As the relative amplitude of the external perturbations is diminished, the dynamics approaches the $e_p = 0$ phase-space portrait seen in Figure 2A. Naturally, this will also lead to stabilization of the orbits. It is noteworthy that such an effect is also at play in the solar system, as additional precession from GR places the free precession of Mercury further from the ν_5 secular resonance, diminishing its chances of ejection (Batygin and Laughlin, 2008; Laskar and Gastineau, 2009).

Finally, it is worthwhile to consider the limitations of the presented model. Indeed, we have approached the problem by utilizing a classical perturbation theory, where only a few relevant terms are retained in the disturbing function. As already mentioned above, this approach was necessary for the exploration of parameter space, as the efficiency offered by conventional direct integration is not sufficient. Although the approach we take here breaks down at high eccentricities, in the solar system, it has been successful in capturing the important physical processes that govern chaotic motion (Sidlichovsky, 1990). The recent work of Lithwick and Wu (2010) has further confirmed this to be true in the case

of Mercury's orbit. Thus, we expect that inclusion of the full disturbing function will only modify our findings on a quantitative level. However, future numerical confirmation and re-eavluation of the work done here will surely be a fruitful venture, especially if performed in the context of a particular observed system.

9.7 Appendix: Coefficients of the Hamiltonian

In this work, we choose to write the coefficients, such that they appear in the equations of motion without pre-factors. The notation used here is identical to that of Murray and Dermott (1999), i.e. b denotes a Laplace coefficient, α is the semi-major axis ratio, and $\mathcal{D} \equiv \partial/\partial\alpha$. Throughout the paper, we account for the induced precession that arises from the $m = 15m_\oplus$, $\alpha_p = 1/2$ perturber as well as Jupiter, with $\alpha_J = 0.178$, but not Saturn. Evaluation of the formulae below shows that Saturn's effect is negligible. The eccentricity forcing, taken into account is solely due to the $m = 15m_\oplus$ perturber. The resulting formulae read:

$$\eta = \frac{n}{4}\left(\frac{m_p}{M_\star}\frac{a}{a_p}\left(2\alpha_p\mathcal{D} + \alpha_p^2\mathcal{D}^2\right)b_{\frac{1}{2}}^{(0)}(\alpha_p) + \frac{m_J}{M_\star}\frac{a}{a_J}\left(2\alpha_J\mathcal{D} + \alpha_J^2\mathcal{D}^2\right)b_{\frac{1}{2}}^{(0)}(\alpha_J)\right) \tag{9.8}$$

$$\beta = \frac{n}{32}\left(\frac{m_p}{M_\star}\frac{a}{a_p}\left(4\alpha_p^3\mathcal{D}^3 + \alpha_p^4\mathcal{D}^4\right)b_{\frac{1}{2}}^{(0)}(\alpha_p) + \frac{m_J}{M_\star}\frac{a}{a_J}\left(4\alpha_J^3\mathcal{D}^3 + \alpha_J^4\mathcal{D}^4\right)b_{\frac{1}{2}}^{(0)}(\alpha_J)\right) \tag{9.9}$$

$$\gamma = \frac{n}{4}\frac{m_p}{M_\star}\frac{a}{a_p}\left(2 - 2\alpha_p\mathcal{D} - \alpha^2\mathcal{D}^2\right)b_{\frac{1}{2}}^{(1)}(\alpha_p) \tag{9.10}$$

The first and the second terms in η and β arise from the $m = 15m_\oplus$ perturber and Jupiter respectively. All terms in γ correspond to the $m = 15m_\oplus$ perturber.

Acknowledgments We thank K. Tsiganis for numerous useful discussions and Oded Aharonson for carefully reviewing the manuscript. Additionally, we thank the anonymous referees for useful suggestions.

Bibliography

Albers, D. J., Sprott, J. C. 2006. Routes to chaos in high-dimensional dynamical systems: A qualitative numerical study. Physica D Nonlinear Phenomena 223, 194-207.

Laskar, J. 1996. Large Scale Chaos and Marginal Stability in the Solar System. Celestial Mechanics and Dynamical Astronomy 64, 115-162.

Laskar, J. 1994. Large-scale chaos in the solar system. Astronomy and Astrophysics 287, L9-L12.

Laskar, J. 1989. A numerical experiment on the chaotic behaviour of the solar system. Nature 338, 237.

Laskar, J., Gastineau, M. 2009. Existence of collisional trajectories of Mercury, Mars and Venus with the Earth. Nature 459, 817-819.

Lovis, C., and 13 colleagues 2011. The HARPS search for southern extra-solar planets. XXVIII. Up to seven planets orbiting HD 10180: probing the architecture of low-mass planetary systems. Astronomy and Astrophysics 528, A112.

Batygin, K., Laughlin, G. 2011. Resolving the sin(I) Degeneracy in Low-mass Multi-planet Systems. The Astrophysical Journal 730, 95.

Batygin, K., Bodenheimer, P., Laughlin, G. 2009. Determination of the Interior Structure of Transiting Planets in Multiple-Planet Systems. The Astrophysical Journal 704, L49-L53.

Batygin, K., Laughlin, G., Meschiari, S., Rivera, E., Vogt, S., Butler, P. 2009. A Quasi-stationary Solution to Gliese 436b's Eccentricity. The Astrophysical Journal 699, 23-30.

Batygin, K., Laughlin, G. 2008. On the Dynamical Stability of the Solar System. The Astrophysical Journal 683, 1207-1216.

Beauge, C., Ferraz-Mello, S. 1993. Resonance trapping in the primordial solar nebula - The case of a Stokes drag dissipation. Icarus 103, 301-318.

Gonczi, R., Froeschle, C., Froeschle, C. 1982. Poynting-Robertson drag and orbital resonance. Icarus 51, 633-654.

Henrard, J., Lemaitre, A. 1983. A second fundamental model for resonance. Celestial Mechanics 30, 197-218.

Henrard, J., Morbidelli, A. 1993. Slow crossing of a stochastic layer. Physica D Nonlinear Phenomena 68, 187-200.

Henrard, J., Henrard, M. 1991. Slow crossing of a stochastic layer. Physica D Nonlinear Phenomena 54, 135-146.

Bruhwiler, D. L., Cary, J. R. 1989. Diffusion of particles in a slowly modulated wave. Physica D Nonlinear Phenomena 40, 265-282.

Murray, N., Holman, M. 1999. The Origin of Chaos in the Outer Solar System. Science 283, 1877.

Ott, E. 1993. Chaos in dynamical systems. Cambridge: Cambridge University Press, —c1993 .

Goldreich, P., Soter, S. 1966. Q in the Solar System. Icarus 5, 375-389.

Mardling, R. A. 2007. Long-term tidal evolution of short-period planets with companions. Monthly Notices of the Royal Astronomical Society 382, 1768-1790.

Morbidelli, A. 2002. Modern celestial mechanics : aspects of solar system dynamics. Modern celestial mechanics : aspects of solar system dynamics, by Alessandro Morbidelli. London: Taylor Francis, 2002, ISBN 0415279399 .

Lithwick, Y., Wu, Y. 2010. Theory of Secular Chaos and Mercury's Orbit. ArXiv e-prints arXiv:1012.3706.

Sidlichovsky, M. 1990. The existence of a chaotic region due to the overlap of secular resonances nu5 and nu6. Celestial Mechanics and Dynamical Astronomy 49, 177-196.

Chirikov, B. V. 1979. A universal instability of many-dimensional oscillator systems. Physics Reports 52, 263-379.

Murray, C. D., Dermott, S. F. 1999. Solar system dynamics. Solar system dynamics by Murray, C. D., 1999 .

Wisdom, J. 1980. The resonance overlap criterion and the onset of stochastic behavior in the restricted three-body problem. The Astronomical Journal 85, 1122-1133.

Wu, Y., Goldreich, P. 2002. Tidal Evolution of the Planetary System around HD 83443. The Astrophysical Journal 564, 1024-1027.

Lee, M. H., Peale, S. J. 2002. Dynamics and Origin of the 2:1 Orbital Resonances of the GJ 876 Planets. The Astrophysical Journal 567, 596-609.

Wu, Y. 2005. Origin of Tidal Dissipation in Jupiter. II. The Value of Q. The Astrophysical Journal 635, 688-710.

Benettin, G., Galgani, L., Strelcyn, J.-M. 1976. Kolmogorov entropy and numerical experiments. Physical Review A 14, 2338-2345.

Chapter 10

Formation and Evolution of Planetary Systems in Presence of Highly Inclined Stellar Perturbers

Originally published as:
Batygin, K., Morbidelli, A., & Tsiganis, K. 2011, Astronomy & Astrophysics, 533, A7

10.1 Abstract

The presence of highly eccentric extrasolar planets in binary stellar systems suggests that the Kozai effect has played an important role in shaping their dynamical architectures. However, the *formation* of planets in inclined binary systems poses a considerable theoretical challenge, as orbital excitation due to the Kozai resonance implies destructive, high-velocity collisions among planetesimals. To resolve the apparent difficulties posed by Kozai resonance, we seek to identify the primary physical processes responsible for inhibiting the action of Kozai cycles in protoplanetary disks. Subsequently, we seek to understand how newly-formed planetary systems transition to their observed, Kozai-dominated dynamical states. The main focus of this study is on understanding the important mechanisms at play. Thus, we rely primarily on analytical perturbation theory in our calculations.

Where the analytical approach fails to suffice, we perform numerical N-body experiments. We find that theoretical difficulties in planet formation arising from the presence of a distant ($\tilde{a} \sim 1000$AU) companion star, posed by the Kozai effect and other secular perturbations, can be overcome by a proper account of gravitational interactions within the protoplanetary disk. In particular, fast apsidal recession induced by disk self-gravity tends to erase the Kozai effect, and ensure that the disk's unwarped, rigid structure is maintained. Subsequently, once a planetary system has formed, the Kozai effect can continue to be wiped out as a result of apsidal precession, arising from planet-planet interactions. However, if such a system undergoes a dynamical instability, its architecture may change in such a way that the Kozai effect becomes operative. The results presented here suggest that planetary formation in highly inclined binary systems is not stalled by perturbations, arising from the stellar companion. Consequently, planet formation in binary stars is probably no different from that around single stars on a qualitative level. Furthermore, it is likely that systems where the Kozai effect operates, underwent a transient phase of dynamical instability in the past.

10.2 Introduction

Among the most unexpected discoveries brought forth by a continually growing collection of extra-solar planets has been the realization that giant planets can have near-parabolic orbits. Since the seminal discovery of 16Cygni B (Cochran et al., 1997), followed by HD80606 (Naef et al., 2001), much effort has been dedicated to understanding the dynamical origin and evolution of systems with highly eccentric planets. In particular, it has

been understood that in presence of a companion star on an inclined orbital plane, the most likely pathway to production of such extreme planet eccentricities is via Kozai resonance (Eggleton and Kiseleva-Eggleton, 2001).

The Kozai resonance was first discovered in the context of orbital dynamics of highly-inclined asteroids forced by Jupiter, and has been subsequently recognized as an important process in sculpting the asteroid belt (Kozai, 1962) as well as being the primary mechanism by which long-period comets become Sun-grazing (Bailey et al., 1992; Thomas and Morbidelli, 1996). Physically, the Kozai resonance corresponds to extensive excursions in eccentricity and inclination of a test particle forced by a massive perturber, subject to conservation of the third Delaunay momentum $H = \sqrt{1-e^2}\cos(i)$ (where e is the eccentricity and i is the inclination), and libration of its argument of perihelion ω around $\pm 90^\circ$. A necessary criterion for the resonance is a sufficiently large inclination ($i > \arccos\sqrt{3/5}$) relative to the massive perturber's orbital plane, during the part of the cycle where the test-particle's orbit is circular.

By direct analogy with the Sun-Jupiter-asteroid picture, the Kozai resonance can give rise to variation in orbital eccentricity and inclination of an extra-solar planet, whose orbit, at the time of formation, is inclined with respect to a stellar companion of the planet's host star (Wu and Murray, 2003). In the systems mentioned above (16Cygni B, HD80606) the stellar companions' (e.g. 16Cygni A, HD80607) proper motion has been verified to be consistent with a binary solution. Other examples of planets in binary stellar systems are now plentiful (e.g. γ Cephei (Hatzes et al., 2003), HD 196885 (Correia et al., 2008), etc) with binary separation spanning a wide range ($\tilde{a} \sim 10 - 1000$ AU). However, all planets

whose eccentricities are expected to have been excited by the Kozai resonance with the companion star are in wide binaries.

If a Kozai cycle is characterized by a sufficiently small perihelion distance, the eccentricity of the planet may subsequently decay tidally, yielding a pathway to production of hot Jupiters, whose orbital angular momentum vector is mis-aligned with respect to the stellar rotation axis (Fabrycky and Tremaine, 2007). The presence of such objects has been confirmed via observations of the Rossiter-McLaughlin effect (McLaughlin, 1924), leading to a notion that Kozai cycles with tidal friction are responsible for generating at least some misaligned systems (Winn et al., 2010; Morton and Johnson, 2011).

Kozai cycles may have also played an important role in systems where a stellar companion is not currently observed. Indeed, one can envision an evolutionary history where the binary companion gets stripped away as the birth cluster disperses. In fact, such a scenario may be rather likely, as the majority of stars are born in binary systems (Duquennoy and Mayor, 1991). In this case, a Kozai cycle can be suddenly interrupted, causing the planet's eccentricity to become "frozen-in."

In face of the observationally suggested importance of Kozai cycles during early epochs of planetary systems' dynamical evolution, the *formation* of planets in presence of a massive, inclined perturber poses a significant theoretical challenge (Larwood et al., 1996; Marzari et al., 2009; Thébault et al., 2010). After all, in the context of the restricted problem (where only the stars are treated as massive perturbers), one would expect the protoplanetary disk to undergo significant excursions in eccentricity and inclination due to the Kozai resonance, with different temporal phases at different radial distances, resulting

in an incoherent structure. Such a disk would be characterized by high-velocity impacts among newly-formed planetesimals, strongly inhibiting formation of more massive objects (planetary embryos) (Lissauer, 1993).

Damping of eccentricities due to gas-drag has been considered as an orbital stabilization process. However, excitation of mutual inclination among neighboring annuli renders this mechanism ineffective (Marzari et al., 2009). Ultimately, in the context of a restricted model, one is forced to resort to competing time-scales for formation of planetesimals and dynamical excitation by the companion star. Such an analysis suggests that although possible, planetary formation in binary systems is an unlikely event.

Here, we show that the theoretical difficulties in planet formation arising from the companion star, posed by the Kozai effect and other secular perturbations, can be resolved by a proper account of the self-gravity of the proto-planetary disk (i.e. planetesimals embedded in a gaseous disk). During the preparation of this manuscript, a paper was published (Fragner et al., 2011) addressing the role of the gravity of a gas disk on the relative motion of embedded planetesimals, with hydrodynamical simulations. The work of Fragner et al. (2011) neglects the gravitational effects of the gas-disk onto itself, and therefore considers a case where pressure and viscosity keep the disk more coherent against external perturbations than would be possible with self-gravity alone (Fragner and Nelson, 2010). This parameter regime is characteristic of systems where external stresses are strong enough to partially overcome the role of self-gravity, but not that of the internal forces of the fluid. Examples of such systems include binary stars with moderate separations (60 AU in the simulations of Fragner et al. (2011)).

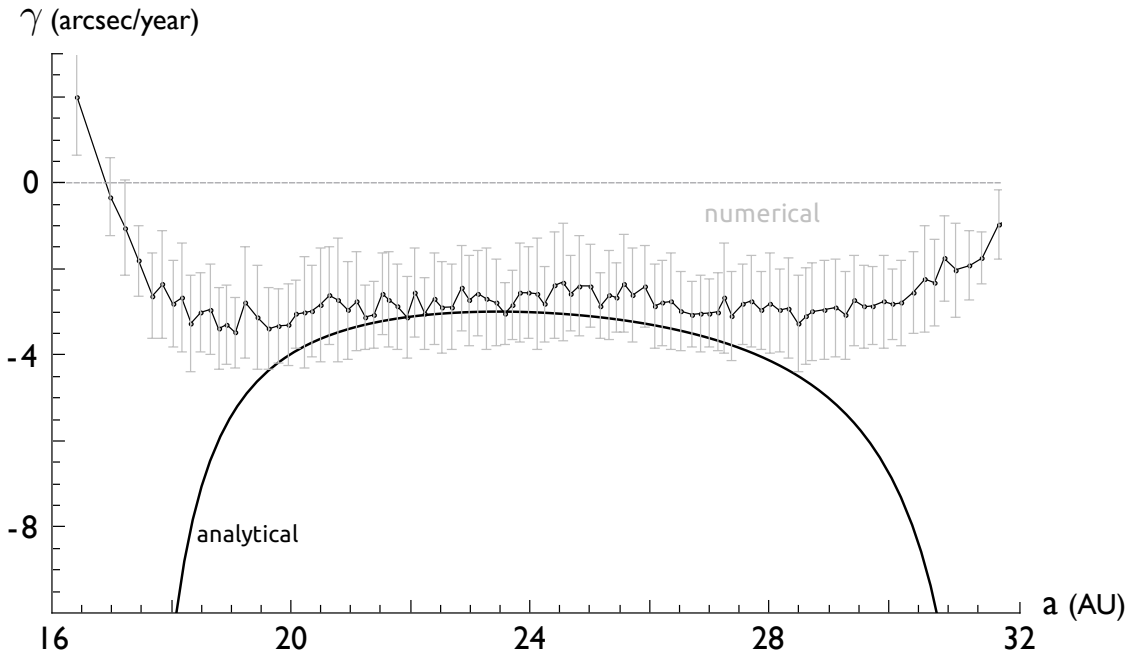


Figure 10.1: An example of apsidal precession,γ, in a self-gravitating disk. Here the disk is assumed to contain $50M_{\oplus}$ between 16 and 32 AU, characteristic of a typical post-formation debris disk in the Nice model of solar system formation (Tsiganis et al., 2005). The solid curve shows the precession rate predicted by eqs. (1)-(3), as a function of semi major axis. The dots and error bars show the results of a numerical calculation, integrating 3,000 equal-mass particles with a softening parameter of $\epsilon \approx 0.005$ AU to smooth the effects of their mutual close encounters. The disk was binned into 100 annuli in a and the mean frequency of the longitude of pericenter $\dot{\varpi}$ was measured from the time-series of ϖ of the particles in each bin (dots) as well as its variance (error bars). Note that the precession frequency of a self-gravitating disk is negative.

In systems of this sort, the dynamics of the planetesimals tends to be somewhat different from those of the gas. Consequently, gas-drag induces size-sorted orbital evolutions, ultimately leading to high-velocity impacts among planetesimals of different sizes. This again, limits the prospects for accretion. Conversely, in the present paper we consider binary separation on the order of $\tilde{a} \sim 1000$ AU, consistent with the cases of 16Cyg B and HD HD80606. This allows us to show, with a simple analytic approach, that gravity is a sufficient mechanism to maintain orbital coherence and planetary growth, without any need to account for other forces acting inside the disk. In fact, we show that one of the primary effects of self-gravity is to induce a fast, rigid recession in the longitudes of perihelion and ascending node of the disk. This allows for planetary formation to take place, as if the secular perturbations arising from the stellar companion were not present. It is noteworthy that such a process is in play, for instance, in the Uranian satellite system, where the Kozai effect arising from the Sun is wiped out owing to secular interactions among the satellites and the precession arising from Uranus' oblateness (Morbidelli, 2002).

Furthermore, we show that even after the formation process is complete, and the disk has evaporated, the Kozai effect may continue to be wiped out by the orbital precession, arising from planet-planet interactions. This is again in line with the example of the outer solar system, where interactions among the giant planets erase a Kozai-like excitation due to the galactic tide (Fabrycky and Tremaine, 2007). However, if such a planetary system undergoes a dynamical instability, which leads to a considerable change in system architecture, it may evolve to a state where the Kozai resonance is no-longer inhibited.

The purpose of this work is to identify the important physical processes at play, rather

than to perform precise numerical simulations. Consequently, we take a primarily analytical approach in addressing the problem. The plan of this paper is as follows. In section 2, we compute the precession rate, arising from the self-gravity of the disk and show that it is copiously sufficient to impede the Kozai effect. In section 3, we show that under secular perturbations from the companion star, the reference plane of the disk precesses rigidly, implying an un-warped structure. In section 4, we show how an initially stable two-planet system enters the Kozai resonance after a transient instability causes one of the planets to be ejected from the system. We summarize and discuss our results in section 5.

10.3 Kozai Resonance in a Self-Gravitating Disk

We begin by considering the secular dynamics of planetesimals in an isolated, flat nearly-circular disk of total mass $M_{\rm disk}$ around a Sun-like ($M_\star = 1 M_\odot$) star. Due to a nearly-null angular momentum deficit, secular interactions within the disk will not excite the eccentricities and inclinations significantly. Rather, as already mentioned above, the primary effect of disk self-gravity is to induce a fast, retrograde apsidal precession.

Our calculation of the induced precession follows the formalism of Binney and Tremaine (1987), originally developed in the context of galactic dynamics. We work in terms of a polar coordinate system, where the radial coordinate is logarithmic ($\rho = \ln r$) and ϕ denotes the polar angle. The reduced potential due to a disk surface density σ reads:

$$\Phi = -\frac{\mathcal{G}}{\sqrt{2}} \int_{-\infty}^{\infty} \int_{0}^{2\pi} \frac{\mathbf{e}^{\rho/2} \sigma}{\sqrt{\cosh(\rho - \rho') - \cos(\phi - \phi')}} d\phi' d\rho' \ , \tag{10.1}$$

where $\mathcal{G}$ is the gravitational constant. We assume $\sigma \propto r^{-1}$. Consequently, axial symmetry is implicit, and the potential is only a function of ρ. The characteristic frequencies of a planetesimal in the disk are the mean motion, n, and the radial frequency, κ:

$$\begin{aligned} n &= \frac{1}{a}\left(\frac{\partial \Phi}{\partial r}\right) \\ \kappa &= \frac{3}{a}\left(\frac{\partial \Phi}{\partial r}\right) + \left(\frac{\partial^2 \Phi}{\partial r^2}\right) \end{aligned} \tag{10.2}$$

where a is semi-major axis. The apsidal precession that results from self-gravity, γ, can then be written as the difference between the mean motion and radial frequencies

$$\gamma \equiv n - \kappa. \tag{10.3}$$

In practice, the calculation of γ is performed by breaking up the disk into cells and computing the derivatives discretely. Following Levison and Morbidelli (2007), we split the disk into 1000 logarithmic radial annuli, and take the angular cell width to be $\Delta\phi = 0.5°$. We assume the disk edges to be $a_{in} = 0.5$AU and $a_{out} = 50$AU, although the results are not particularly sensitive to these choices. The resulting precession in the disk is roughly uniform in a, except for the edges, where this linear theory breaks down. Numerical experiments of debris disks, where self-gravity is taken into account directly, however, show that the precession rate at the edges is also roughly uniform and quantitatively close to that elsewhere in the disk (see Figure 1). In other words, the disk's apsidal precession is approximately rigid. Consequently, for the purposes of this work, we take the precession rate evaluated at $a = 10$AU, to be the characteristic γ for the entire disk.

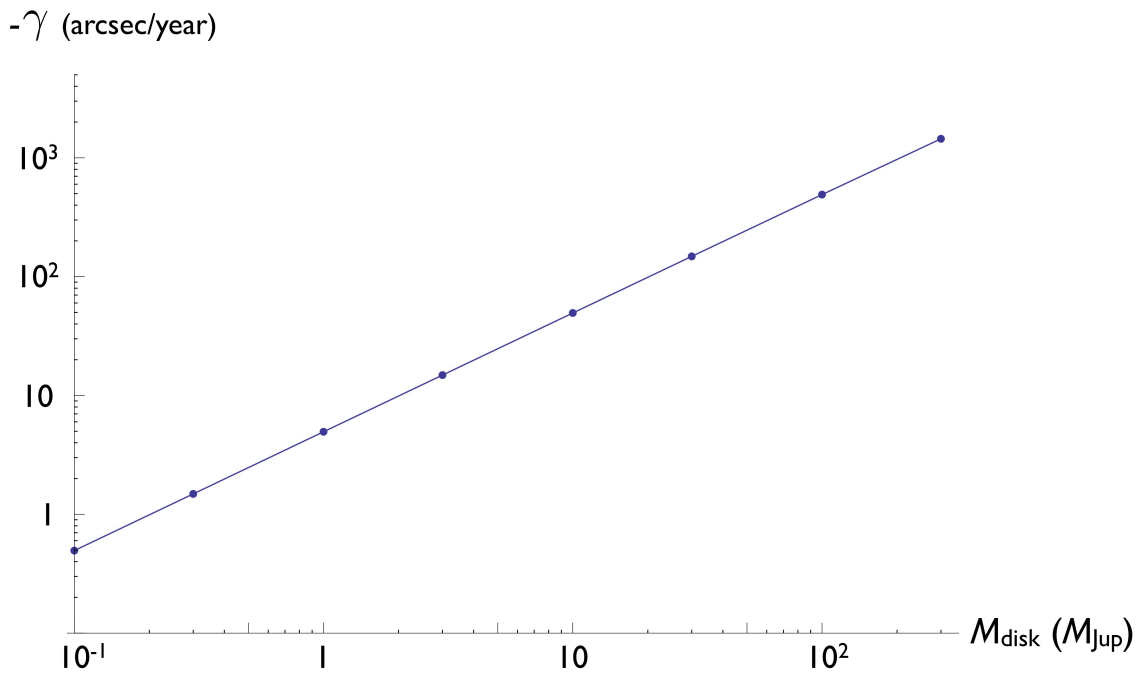


Figure 10.2: Apsidal recession of a self-gravitating disk. The recession rate, γ is plotted as a function of disk mass. Blue points are the model results. The points are well fit by a linear functional relationship $\gamma = -2.4 \times 10^{-5}(M_{\rm disk}/M_{\rm Jup})$ rad/year.

Generally, typical protoplanetary disks contain $M_{\rm disk} \sim 10 - 100 M_{\rm Jup}$ at the time of formation, in gas and planetesimals. We have calculated the characteristic precession rate for a planetesimal embedded in such a disk, for total disk mass range, spanning roughly two orders of magnitude, between $M_{\rm disk} = 0.1 M_{\rm Jup}$ and $M_{\rm disk} = 300 M_{\rm Jup}$. Figure 2 shows the relationship between γ and $M_{\rm disk}$. Note that unlike typical planetary systems, where secular interactions among planets give rise to positive apsidal precession, γ of a self-gravitating disk is negative. Quantitatively, for the assumed disk geometry, the precession rate is well fit by the functional relationship $\gamma = -2.4 \times 10^{-5}(M_{\rm disk}/M_{\rm Jup})$ rad/year. Having computed the characteristic precession rate, we can now write down the orbit-averaged Hamiltonian of a planetesimal in the disk.

We work in terms of canonically conjugated action-angle Delaunay variables

$$\begin{aligned} G &= \sqrt{a(1-e^2)}\,, \qquad g = \omega \\ H &= \sqrt{a(1-e^2)}\cos i\,, \quad h = \Omega \end{aligned}$$

where the inclination i is measured relative to an arbitrary reference plane and Ω is the longitude of the node of the disk relative to such a plane. In the analysis that follows, we shall take the binary star's orbital plane to be the reference plane. The Hamiltonian is simply

$$\mathcal{K}^{SG} = \gamma G\,. \tag{10.4}$$

$\mathcal{K}^{SG}$ describes an eccentric precessing orbit on a fixed orbital plane (i.e. the plane of the disk).

Let us now incorporate the perturbations from the stellar companion into the Hamiltonian. Due to a considerable orbital separation between the protoplanetary disk and the perturber, the interactions between the two will be secular in nature. For our purposes here, we take the perturber to lie on a circular, inclined orbit. A circular perturber implies that, to leading order, potential eccentricity excitations in the disk would arise exclusively from the Kozai resonance. The free orbital precession induced by the companion can be approximated as (Murray and Dermott, 1999)

$$\dot{g}_{\rm free} \simeq \frac{3n}{2}\frac{\tilde{m}}{M}\left(\frac{a}{\tilde{a}}\right)^3 \tag{10.5}$$

where $\tilde{m}$ and $\tilde{a}$ are the perturber's mass and semi-major axis and M is the mass of the

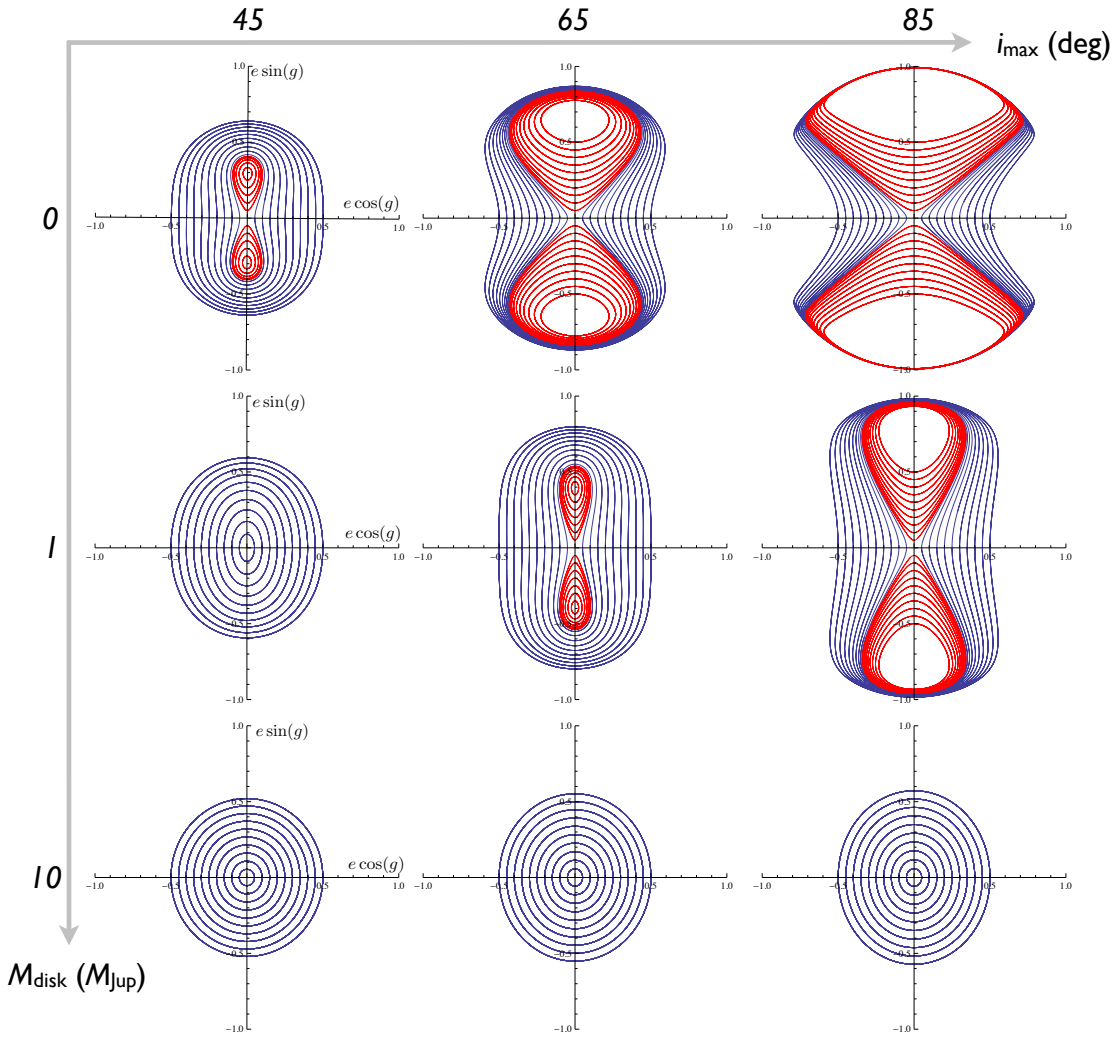


Figure 10.3: Dynamical phase-space portraits for a planetesimal in protoplanetary disks of various masses, perturbed by a stellar companion at various inclinations showing Kozai resonance. The eccentricity vector is plotted in cartesian coordinates on each panel ($x = e\cos g, y = e\sin g$). Regions of libration of argument of perihelion are shown as red curves, while blue curves depict circulation. The top panels represent a mass-less disk, middle panels correspond to a $M_{\rm disk} = 1 M_{\rm Jup}$ disk and the bottom panels show a $M_{\rm disk} = 10 M_{\rm Jup}$ disk. Note that the Kozai resonance disappears as the disk mass is increased.

central star. In the following, we assume that

$$\dot{g}_{\rm free} \ll -|\gamma| \ , \tag{10.6}$$

and neglect $\dot{g}_{\rm free}$ altogether. This condition is satisfied for well-separated binary systems ($a/\tilde{a} \ll 1$). Note that the lack of orbital eccentricity of the perturber is a mathematical convenience that makes the calculation more straight forward, without modifying our conclusions on Kozai dynamics qualitatively[1].

We take the stellar companion's orbital semi-major axis to be $a = 1000$AU and take the inclination as well as disk mass to be variable parameters. This choice is motivated by the estimate of the orbital separation between HD80606 and HD80607 (Eggenberger et al., 2003). It is noteworthy, that the particular choice of a does not have significant consequences on the dynamics of the disk, beyond setting the time-scale on which the Kozai effect operates, provided that it is large enough that condition (10.6) is satisfied. In this section, we shall assume that the nodal reference plane of the disk precesses rigidly, and the disk remains un-warped. In other words, we assume that no mutual inclination is excited between neighboring disk annuli. This feature is implicitly essential to our argument, and we will justify this assumption quantitatively in the next section.

In accord with the reasoning outlined above, we solely retain the Kozai term in the disturbing potential of the stellar companion. Consequently, the planetesimal's Hamiltonian

[1] An eccentric star would induce some forced eccentricity in the disk. However, because of the fast rigid apsidal precession of the disk, the forced eccentricity will be small and the disk will maintain coherence.

now reads (Kinoshita and Nakai, 1999)

$$\begin{aligned}\mathcal{K}^{SGK} &= \gamma G + \frac{a\tilde{m}\tilde{n}^2}{(M+\tilde{m})}\left\{\frac{15(a-G^2)(G^2-H^2)\cos(2g)}{G^2}\right. \\ &- \left.\frac{(5a-3G^2)(G^2-3H^2)}{G^2}\right\} \end{aligned} \tag{10.7}$$

where $\tilde{n}$ is the stellar perturber's mean motion. Notice that in the Hamiltonian above, γ could be factorized, so that the magnitude of the perturbation would be proportional to $\tilde{n}/\gamma$. Since γ is a linear function of the disk mass, this illustrates that it is equivalent to have a closer binary companion (larger $\tilde{n}$) to a proportionally less massive disk.

With this simplified dynamical model in place, we can now explore the effect of self-gravity on the orbital excitation of the planetesimals in the disk due to the Kozai resonance. We study three choices of perturber inclination: $i = 45°$, $i = 65°$ and $i = 85°$. To obtain a dynamical portrait of the system, we proceed as follows. Because the variable h does not appear in (10.7), the action H is a constant of motion. On each $H = \text{constant}$ surface, the Hamiltonian, $\mathcal{K}^{SGK}$, describes a one-degree of freedom system, in the variables G, g. Simultaneously, because the Hamiltonian is also a constant of motion, the dynamics is described by the level curves of the Hamiltonian. For simplicity, we show the dynamics in cartesian coordinates ($x = e\cos g, y = e\sin g$) in Figure 10.3, where e is computed from the definition of G, for the assumed value of a (here $a = 10$AU). Given that H is constant on each panel, a given eccentricity also yields the inclination. The panels are identified by the value of the inclination i_{max} that corresponds to $e = 0$ for the given value of H (i.e. the inclination of the star relative to the initial, circular disk). Similarly, the maximal value of e on each panel corresponds to $i = 0$.

It is useful to begin with a discussion of a mass-less disk as this configuration is often assumed in formation studies. The corresponding plots are shown as the top panel of Figure 3. In this case, there is no added precession ($\gamma = 0$) so the Kozai resonance is present for all considered choices of inclination. The phase-space portraits show that any orbit which starts out at low eccentricity (near the origin) will follow a trajectory which will eventually lead to a highly eccentric orbit, regardless of initial phase. In particular, for $i_{\rm max} = 45^\circ$, a particle which starts out on a circular orbit will attain $e_{max} \simeq 0.4$. For $i_{\rm max} = 65^\circ$, $e_{max} \simeq 0.85$ and for $i_{\rm max} = 85^\circ$, $e_{max} \simeq 1$. The resulting high-velocity collisions render formation of planetary embryos ineffective. Consequently, one should not expect planets to form under the mass-less disk approximation.

Let us now consider a $M_{\rm disk} = 1 M_{\rm Jup}$ disk. The phase-space portraits of this system are shown as the middle panels of Figure 3. Although the quoted value corresponds to a very low-mass disk, the situation is considerably different from the mass-less case. For $i_{\rm max} = 45^\circ$, the Kozai resonance is no longer effective[2]. Thus, an initially nearly-circular orbit will retain its near-zero eccentricity, allowing for planetary formation to take place. The Kozai resonance still operates in the $i_{\rm max} = 65^\circ$ and $i_{\rm max} = 85^\circ$ cases, but the maximum eccentricities are now lower ($e_{max} \simeq 0.55$ and $e_{max} \simeq 0.95$ respectively) compared to the mass-less disk scenario.

Finally, the bottom panels of Figure 2 show the phase-space portraits of a $M_{\rm disk} = 10 M_{\rm Jup}$ disk. Here, the Kozai resonance is completely wiped out, for all values of $i_{\rm max}$.

[2]Interestingly, the disappearance of the Kozai separatrix is not exactly symmetric with respect to the sign of γ. If γ is negative, it immediately acts to erase the Kozai effect. However, a small positive γ (i.e. $\gamma = 10^{-5}$ for $i_{max} = 45^\circ$) can act to enhance to Kozai effect. The effect however rapidly turns over for faster positive precession (i.e. $\gamma > 10^{-4}$).

Particularly, circular orbits remain circular (the center of each panel is a stable equilibrium point). Strictly speaking, these calculations describe the dynamics of planetesimals embedded in the disk. However, if planetesimals remain circular, the gaseous component of the disk must do so as well because it feels the same gravitational potential. On the other hand, if the Kozai resonance forces the planetesimals to acquire a considerable eccentricity during their evolution (the cases with low disk mass in Figure 10.3 or, equivalently, cases with a close stellar companion) the gas-disk may remain more circular than the planetesimals, thanks to its additional dissipative forces. This is the situation illustrated in Fragner et al. (2011), where a differential evolution of planetesimals and gas, leads to size-dependent gas-drag forces.

In conclusion, recalling that $10M_{\rm Jup}$ is a lower-bound for the mass of a typical protoplanetary disk, this analysis suggests that planetary formation can take place in well separated binary systems like 16Cyg and HD80606-7 as if secular perturbations, arising from the companion star were not present.

10.4 Rigid Precession of a Self-Gravitating Disk

In the previous section, we showed that a self-gravitating disk is not succeptible to excitation by the Kozai resonance. However, in order for our argument to be complete, it remains to be shown that the assumptions of rigid precession of the disk's nodal reference plane, as well as the lack of the excitation of mutual inclination within the disk, hold true. To justify our assumptions, it is sufficient to consider a nearly circular self-gravitating disk and show that it is characterized by rigid nodal precession, since we have already shown that a flat

disk will remain circular under external perturbations.

Intuitively, one can expect a rigidly precessing flat disk, from adiabatic invariance. Consider an isolated self-gravitating disk where mutual inclinations (inclinations with respect to the instantaneous mid-plane), $\hat{i}$, are initially small (i.e. $\sin\hat{i} \sim \hat{i} \ll 1$). Forced by self-gravity, the mutual inclinations within the disk will be modulated on a characteristic (secular) timescale related to the precession of the longitude of the node, $\hat{\Omega}$, relative to the disk mid-plane. One can define the action,

$$J = \oint \hat{i} d\hat{\Omega} = \text{const.} \tag{10.8}$$

which represents the phase-space area bounded by a secular cycle (Neishtadt, 1984). If the disk is subjected to an external perturbation (such as the torquing from a stellar companion), whose characteristic timescale is much longer than the secular timescale on which self-gravity modulates the mutual inclinations (the so-called adiabatic condition), J will remain a conserved quantity (Henrard and Morbidelli, 1993). This implies that the mutual inclinations within the disk will remain small. This is true for each annulus of the disk in which the adiabatic condition is fulfilled. Consequently, the disk will remain unwarped and the disk mid-plane will precess rigidly at constant inclination relative to the binary star's orbital plane.

The same idea can be illustrated more quantitatively in the context of classical Laplace-Lagrange secular theory. One possible way to view the inclination dynamics of disk is by modeling the disk as a series of massive self-gravitating rings, adjacent to one-another. Note that the same technique cannot be directly applied to eccentricities, since it would

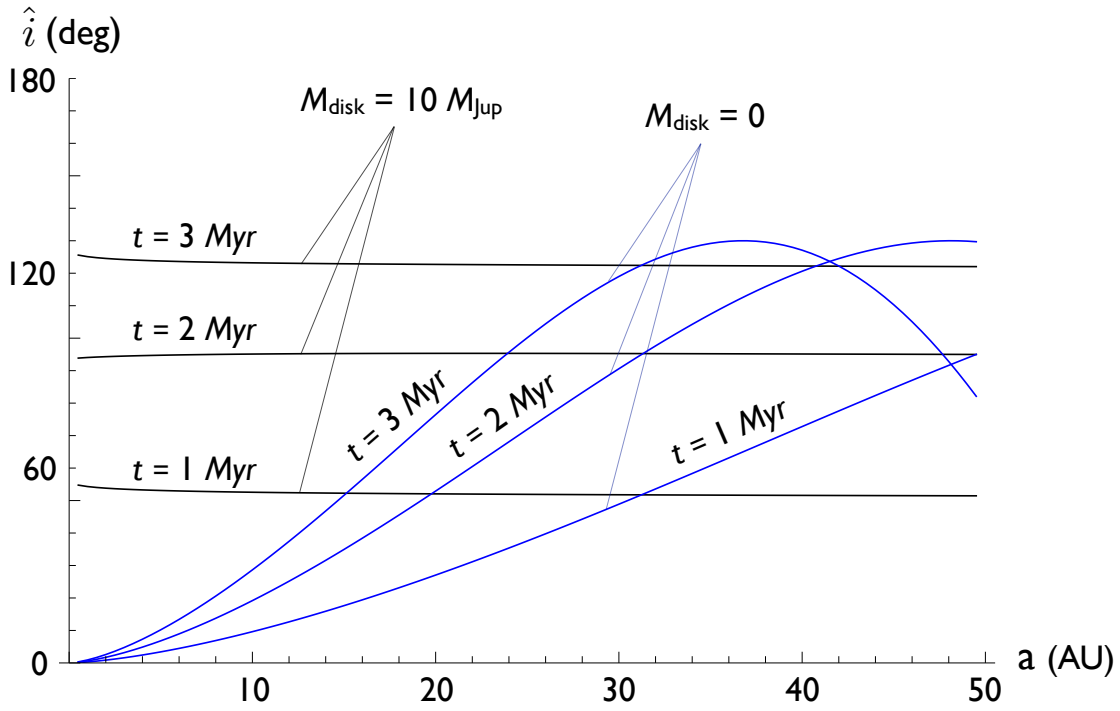


Figure 10.4: Inclination structure of a mass-less (blue) and a self-gravitating $M_{\rm disk} = 10 M_{\rm Jup}$ disks. Here the inclination is measured relative to the original plane of the disk. The inclination is shown as a function of semi major axis a at $t = 1, 3,$ and 5 Myr. Note that the mass-less disk is considerably warped due to the perturbations from the companion star, while the self-gravitating disk maintains a uniform inclination. in this case, the growth of inclination with time is due to the rigid precession of the disk relative to the binary star plane. The inclination returns back to zero after a precession period.

predict a positive apsidal precession, whereas in reality the apsidal precession would be negative, as shown in the previous section. This is because, as soon as the eccentricity is non zero, a ring would start to intersect the adjacent rings, violating the assumption on which the Laplace-Lagrange secular theory is based.

The scaled Hamiltonian of a given annulus j, where exclusively secular terms up to second order in inclination have been retained, reads (Murray and Dermott, 1999)

$$\mathcal{K}_j^{LL} = \frac{1}{2} B_{jj} i_j^{'2} + \sum_{j=1, j\neq k}^{N} B_{jk} i_j^{'} i_k^{'} \cos(\Omega_j^{'} - \Omega_k^{'}) \tag{10.9}$$

where the primed quantities are expressed with respect to an fixed inertial plane (here the initial plane of the disk). In this approach, the disk is broken up into $N-1$ annuli whereas the N^{th} index corresponds to the stellar companion. The coefficients B_{jj} and B_{jk} take the form

$$\begin{aligned} B_{jj} &= -\frac{n_j}{4}\sum_{k=1,k\neq j}^{N}\frac{m_k}{M_\star+m_j}\alpha_{jk}\bar{\alpha}_{jk}b^{(1)}_{3/2}(\alpha_{jk}) \\ B_{jk} &= \frac{n_j}{4}\frac{m_k}{M_\star+m_j}\alpha_{jk}\bar{\alpha}_{jk}b^{(1)}_{3/2}(\alpha_{jk}) \end{aligned} \tag{10.10}$$

where m_j denotes the mass of a given annulus j if $j<N$, $m_N=\tilde{m}$, $\alpha=a_j/a_k$, $\bar{\alpha}=\alpha$ if perturbation is external and $\bar{\alpha}=1$ otherwise, while $b^{(1)}_{3/2}$ is the Laplace coefficient of the first kind. Similarly to equation (6), the diagonal terms in the $\mathbf{B}$ matrix correspond to the free nodal precession rates. To crudely account for the large mutual inclination of the stellar companion, we reduced its mass by a factor of $\sin(i)$ because in the context of second-order theory, it is appropriate to only consider the *projection* of its mass onto the disk's reference plane. Rewriting the above Hamiltonian in terms of cartesian coordinates ($q=i^{'}\cos\Omega^{'}, p=i^{'}\sin\Omega^{'}$), the first-order perturbation equations ($\dot{p}=\partial\mathcal{K}/\partial q, \dot{q}=-\partial\mathcal{K}/\partial p$) yield an eigen-system that can be solved analytically (see Ch. 7 of Murray and Dermott (1999)). In our calculations, we choose $N=101$ and again consider a $\sigma\propto r^{-1}$ surface density across the disk.

We take the orbital properties of the stellar companion to be the same as those discussed in the previous section and take the initial mutual inclination between the disk and the stellar companion to be $i=65^{\circ}$. Indeed, the evaluation of the solution for various disk masses shows that the disk precesses rigidly, if the disk mass exceeds $M_{\rm disk}\gtrsim 1M_{\rm Jup}$.

This threshold is in rough quantitative agreement with the numerical models of Fragner et al. (2011). Figure 4 shows the evolution of the inclination as a function of semi-major axis, of a massless (blue) disk as well as a self-gravitating (black) $M_{\rm Jup} = 10M_{\rm Jup}$ disk at various epochs. The reference plane for the measure of the inclination is the initial plane of the disk. We see that for the mass-less disk the inclination varies considerably with semi-major axis, which means that the disk is significantly warped, as one would expect in the context of a standard restricted 3-body problem. However, the inclination of a self-gravitating disk is nearly constant in semi-major axis, depicting an unwarped, rigid structure. Note, however, that the inclination changes with time. this is because the disk is precessing with a constant inclination relative to the plane of the binary star, so that the disk's inclination relative to its initial plane has to change periodically, over a precession period, from $i_{\rm min} = 0^\circ$ to a maximum of $i_{\rm max} = 130^\circ$ and back.

In conclusion, the assumption of untwisted structure, that we employed in the previous section when calculating the excitation by the Kozai resonance, is valid for massive disks perturbed by distant stellar companions, such as the ones that we consider in this paper. This conclusion does not apply only to the planetesimal disk, but also to the gas disk, for the same reasons mentioned at the end of the previous section.

10.5 Production of Highly Eccentric Planets

In the two preceeding sections, we have shown that in a binary stellar system, a self-gravitating disk avoids dynamical excitation, arising from the stellar companion, even if inclined. As a result, one can expect that formation of planetary systems is generally not

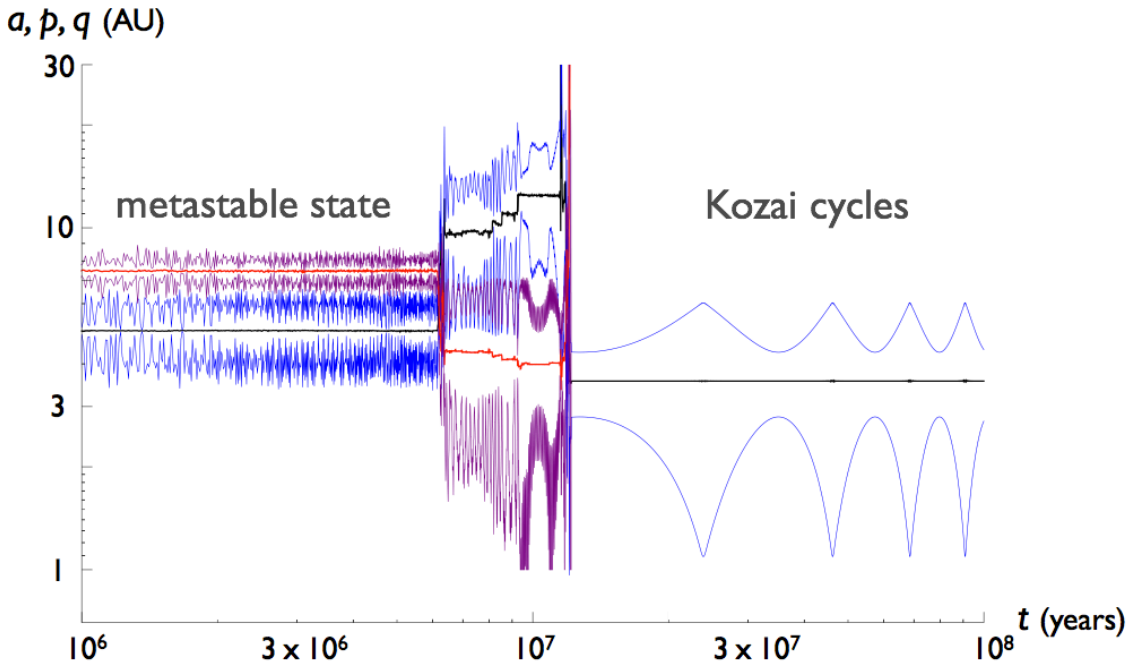


Figure 10.5: Orbital evolution of a two-planet system and its transition into the Kozai resonance via an instability. The figure shows the semi-major axes, as well as perihelion and aphelion distances as functions of time. The planets initially start out in a metastable configuration which is protected from Kozai resonance by apsidal precession, arising from planet-planet interactions. Following ~ 12Myr of dynamical evolution, the planets suffer a dynamical instability, during which the initially outer planet is ejected. Consequently, the remaining planet enters the Kozai resonance.

inhibited. Furthermore, even after the disappearance of the gas, we expect that the Kozai effect can continue to be wiped out as a result of apsidal precession, induced by planet-planet interactions. As already discussed in the introduction, this is the case for the planets of the solar system with respect to the galactic tide, or the satellites of Uranus relative to solar perturbations. Consequently, the final issue we need of address is how planets, such as HD80606b and 16Cygni Bb, do eventually end up undergoing Kozai cycles.

The evolutionary path that a planetary system can take between the birth nebula stage and the Kozai stage is necessarily non-unique. One obvious possibility is that only a single large planet forms in the disk and as the gas evaporates, self-gravity of the disk becomes insufficient to wipe out the Kozai resonance. Such a scenario, although possible, is probably far from being universal, since the observed multiplicity in planetary systems (Mugrauer et al., 2010), as well as theoretical considerations (Armitage, 2010), suggest that protoplanetary disks rarely produce only a single body.

However, an alternative picture can be envisioned: a multiple system forms and after the dispersion of the disk, still protects itself from the Kozai cycles, exerted by the companion star, through self-induced apsidal precession. Then, following a dormant period ,a dynamical instability occurs, removing all planets except one, and therefore the remaining object starts to experience the Kozai resonance. Such a scenario would be considerably more likely, since dynamical instabilities are probably common among newly-formed planetary systems (Ford and Rasio, 2008; Raymond et al., 2009). In particular, over the last decade or so, it has been realized that a transient dynamical instability has played an important role in shaping the architecture of the solar system (Thommes et al., 1999; Tsiganis et al.,

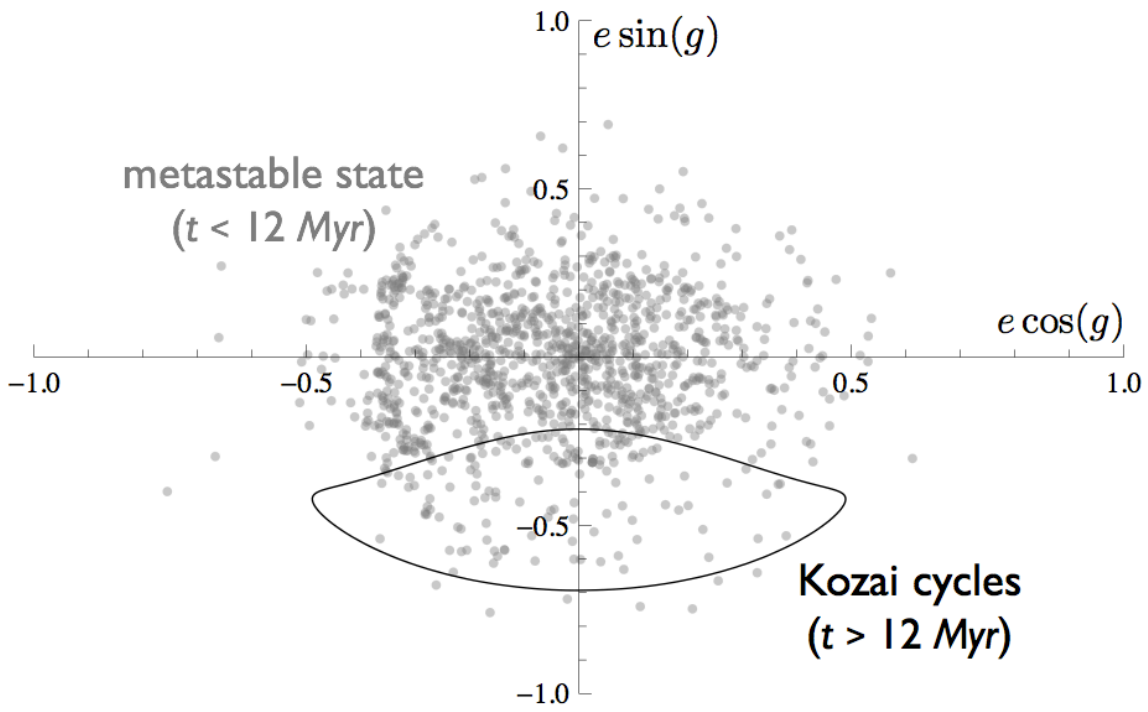


Figure 10.6: Phase-space plot of the inner planet, corresponding to the orbital evolution, shown in Figure 4. Prior to the instability ($t < 12$Myr), the motion of the planet (shown as gray points) is non-resonant. However, after a the outer planet gets ejected, the remaining planet enters the Kozai resonance (shown as a black line).

2005; Morbidelli et al., 2007). Moreover, planet-planet scattering has been suggested to be an important process in explaining the eccentricity distribution of extra-solar planets (Jurić and Tremaine, 2008) as well as the misalignment of planetary orbits with stellar rotation axes (Morton and Johnson, 2011).

The usefulness of analytical methods is limited when it comes to the particular study of dynamical instabilities, so one must resort to numerical methods. Below, we demonstrate a numerical proof-of-concept of the scenario outlined above.

The system we considered was a pair of giant planets, both with $m_p = 1M_J$ around a sun-like ($M_\star = 1M_\odot$) star, perturbed by a $\tilde{m} = 2M_\odot$ companion. The planets were initialized on near circular orbits ($e_1 = e_2 = 0.01$) with $a_1 = 5$AU and $a_2 = 7.5$AU in the same

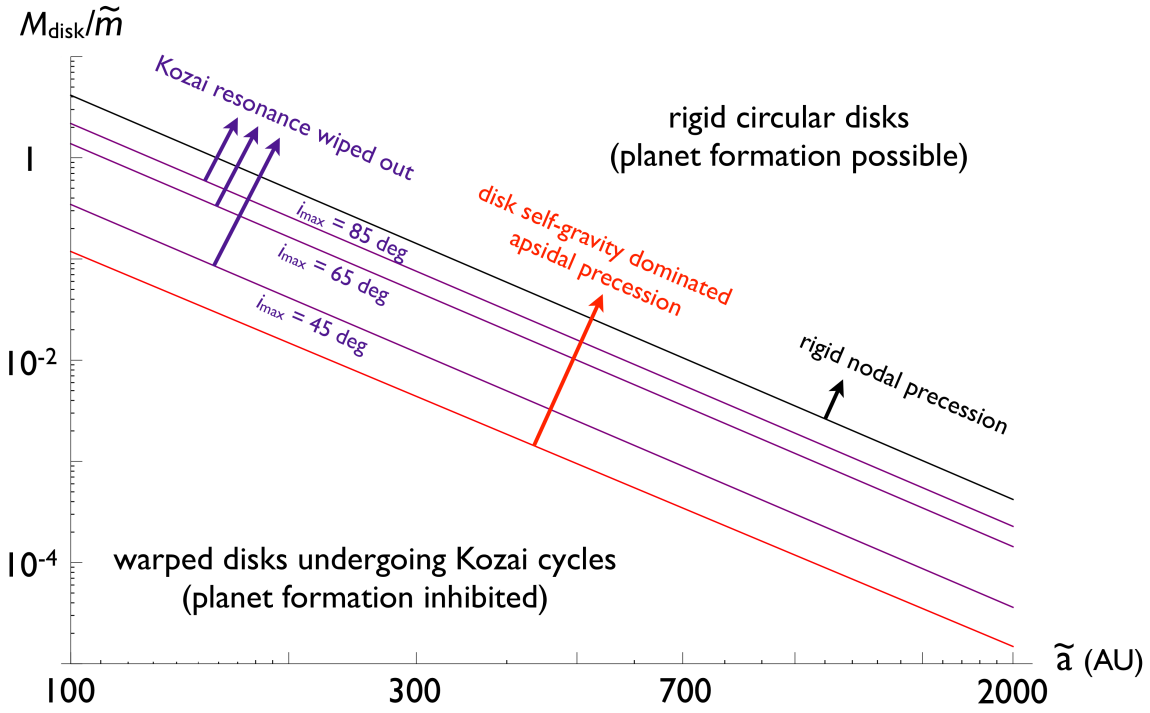


Figure 10.7: Domain of applicability of the arguments presented in this paper. The red curve shows the dividing line between disk-dominated and stellar companion-dominated apsidal precession (as in section 2). The three purple curves illustrate the disappearance of the Kozai separatrix, for various choices of maximal inclination (as in section 3). The black curve delineates the boundary between rigid precession of the disk's mid-plane and a warped structure (as in section 4). Successful formation of planets can take place in well-separated binary systems where disk self-gravity dominates over perturbations from the stellar companion.

plane. The stellar companion was taken to be on a circular $\tilde{a} = 1000$AU orbit, inclined by $i = 80^\circ$ with respect to the orbital plane of the planets. We performed the simulation using a modified version of SyMBA (Duncan et al., 1998) in which a companion star is set on a distant, fixed circular orbit. The timestep was chosen to be 0.2y and throughout the integration, the fractional energy error remained below $\Delta E/E \leqslant 10^{-5}$. The system was evolved over 10^8 years.

The orbital evolution of the system is shown in Figure 5. As can be seen, the system appears stable for the first ~ 6 Myr, with no sign of Kozai oscillations. However, at ~ 6

Myr, the system becomes unstable because the planets‘ orbital proximity (initial orbital separation is only ~ 7 Hill radii) prevents them from remaining stable on long timescales (Chambers et al., 1996). The eccentricities of the planetary orbits grow and eventually, the planets begin to experience close encounters with each other. At ~ 12 Myr, one of the two planets is ejected onto a hyperbolic orbit. The remaining planet, now alone and with no apsidal precession, gets captured into the Kozai resonance, and starts undergoing large, coupled oscillations in eccentricity and inclination. It is noteworthy that had the remaining planet ended up on an orbit with smaller semi-major axis, general relativistic precession could have wiped out the Kozai effect (Wu and Murray, 2003; Fabrycky and Tremaine, 2007). Additionally, although in our setup, the stellar companion was initialized with a high inclination, this is not a necessary condition for the scenario, since the scattering event can generate mutual inclination. The transition from non-resonant motion to that characterized by Kozai cycles is depicted in Figure 6, where orbital parameters prior to the second planet's ejection are shown as gray dots and the resonant motion is shown as a black curve. Note the similarity of resonant motion computed numerically, to that computed analytically, shown in Figure 3.

10.6 Discussion

In this paper, we have addressed the issue of how planetesimals could preserve relative velocities that are slow enough to allow planet accretion to take place, in binary stellar systems. Particularly, we focused on highly inclined systems where Kozai resonance with the perturbing stellar companion have been thought to disrupt the protoplanetary disk and

inhibit planet formation (Marzari et al., 2009). Here, we have shown, from analytical considerations, that fast apsidal precession, which results from the disk self-gravity, wipes out the Kozai resonance and ensures rigid precession of the disk's nodal reference plane.

It is useful to consider the domain of applicability of the criteria discussed here. Namely, the trade-off between stellar binary separation and the perturbing companion's mass should be quantified. The region of parameter space (binary separation $\tilde{a}$ vs disk mass to perturber mass ratio) where self-gravity suppresses secular excitation from the binary companion is delineated in Figure 7. The red curve shows the dividing line between disk-dominated and stellar companion-dominated apsidal precession (as in section 2). The three purple curves illustrate the disappearance of the Kozai separatrix, for various choices of maximal inclination (as in section 3). The black curve delineates the boundary between rigid precession of the disk's mid-plane and a warped structure (as in section 4).

As can be deduced from Figure 7, for distant stellar companions ($\tilde{a} \sim 1000$ AU), the required total disk mass is of order $M_{\rm disk} \sim$ 1-10M_J (depending on the perturber's mass), considerably less than or comparable to, the total mass of the minimum mass solar nebula. This implies that generally, protoplanetary disks in binary stars can maintain roughly circular, unwarped and untwisted structures. Consequently, we can conclude that planetary formation in wide binary systems is qualitatively no different from planetary formation around single stars.

After the formation of planets is complete and the gaseous nebula has dissipated, the Kozai effect can continue to be inhibited as a result of orbital precession induced by planet-planet interactions. However, as the numerical experiment presented here suggests, if a

planetary system experiences a transient dynamical instability that leaves the planets on sufficiently well-separated orbits, the planets can start undergoing Kozai cycles. An evolutionary sequence of this kind can explain the existence of orbital architectures characterized by highly eccentric planets, such as those of HD 80606 and 16 Cygni B (Eggleton and Kiseleva-Eggleton, 2001; Wu and Murray, 2003).

The work presented here resolves, at least in part, a pressing dynamical issue of planetary formation in highly inclined binary systems. As an avenue for further studies, the analytical results presented here should be explored numerically in grater detail. Particularly, hydrodynamic simulations, such as those presented by Fragner et al. (2011) can be used to quantitatively map out the parameter space that allows for planetary systems to form successfully.

The study presented here has further consequences beyond an explanation of planet formation in wide binary systems. Particularly, the model of instability-driven evolution of newly-formed systems into the Kozai resonance has substantial implications for orbital misalignment with the parent star's rotation axis. In fact, Kozai cycles with tidal friction produce a particular distribution of orbit-spin axis angles (Fabrycky and Tremaine, 2007). This distribution differs significantly from that produced by the planet-planet scattering scenario (Nagasawa et al., 2008). This distinction has been used to statistically infer the dominant process by which misaligned hot Jupiters form (Morton and Johnson, 2011). However, the model presented here suggests that the two distributions should be intimately related, as planet-planet scattering provides the initial condition from which Kozai cycles originate. Consequently, a quantitative re-examination of the orbit-spin axis misalignment

angle distribution, formed by Kozai cycles with tidal friction that originate from a scattered orbital architecture, and subsequent comparison of the results with observations of the Rossiter-McLaughlin effect will likely yield new insights into dynamical evolution histories of misaligned hot Jupiters.

Acknowledgments We thank the referee, Y. Wu for useful suggestions.

Bibliography

Armitage, P. J. 2010. Astrophysics of Planet Formation. Astrophysics of Planet Formation, by Philip J. Armitage, pp. 294. ISBN 978-0-521-88745-8 (hardback). Cambridge, UK: Cambridge University Press, 2010. .

Bailey, M. E., Chambers, J. E., Hahn, G. 1992. Origin of sungrazers - A frequent cometary end-state. Astronomy and Astrophysics 257, 315-322.

Binney, J., Tremaine, S. 1987. Galactic dynamics. Princeton, NJ, Princeton University Press, 1987, 747 p. .

Chambers, J. E., Wetherill, G. W., Boss, A. P. 1996. The Stability of Multi-Planet Systems. Icarus 119, 261-268.

Correia, A. C. M., and 11 colleagues 2008. The ELODIE survey for northern extra-solar planets. IV. HD 196885, a close binary star with a 3.7-year planet. Astronomy and Astrophysics 479, 271-275.

Cochran, W. D., Hatzes, A. P., Butler, R. P., Marcy, G. W. 1997. The Discovery of a Planetary Companion to 16 Cygni B. The Astrophysical Journal 483, 457.

Duquennoy, A., Mayor, M. 1991. Multiplicity among solar-type stars in the solar neigh-

bourhood. II - Distribution of the orbital elements in an unbiased sample. Astronomy and Astrophysics 248, 485-524.

Duncan, M. J., Levison, H. F., Lee, M. H. 1998. A Multiple Time Step Symplectic Algorithm for Integrating Close Encounters. The Astronomical Journal 116, 2067-2077.

Eggenberger, A., Udry, S., Mayor, M. 2003. Planets in Binaries. Scientific Frontiers in Research on Extrasolar Planets 294, 43-46.

Eggleton, P. P., Kiseleva-Eggleton, L. 2001. Orbital Evolution in Binary and Triple Stars, with an Application to SS Lacertae. The Astrophysical Journal 562, 1012-1030.

Fabrycky, D., Tremaine, S. 2007. Shrinking Binary and Planetary Orbits by Kozai Cycles with Tidal Friction. The Astrophysical Journal 669, 1298-1315.

Ford, E. B., Rasio, F. A. 2008. Origins of Eccentric Extrasolar Planets: Testing the Planet-Planet Scattering Model. The Astrophysical Journal 686, 621-636.

Fragner, M. M., Nelson, R. P. 2010. Evolution of warped and twisted accretion discs in close binary systems. Astronomy and Astrophysics 511, A77.

Fragner, M. M., Nelson, R. P., Kley, W. 2011. On the dynamics and collisional growth of planetesimals in misaligned binary systems. Astronomy and Astrophysics 528, A40.

Hatzes, A. P., Cochran, W. D., Endl, M., McArthur, B., Paulson, D. B., Walker, G. A. H., Campbell, B., Yang, S. 2003. A Planetary Companion to γ Cephei A. The Astrophysical Journal 599, 1383-1394.

Henrard, J., Morbidelli, A. 1993. Slow crossing of a stochastic layer. Physica D Nonlinear Phenomena 68, 187-200.

Kozai, Y. 1962. Secular perturbations of asteroids with high inclination and eccentricity. The Astronomical Journal 67, 591.

Jurić, M., Tremaine, S. 2008. Dynamical Origin of Extrasolar Planet Eccentricity Distribution. The Astrophysical Journal 686, 603-620.

Kinoshita, H., Nakai, H. 1999. Analytical Solution of the Kozai Resonance and its Application. Celestial Mechanics and Dynamical Astronomy 75, 125-147.

Larwood, J. D., Nelson, R. P., Papaloizou, J. C. B., Terquem, C. 1996. The tidally induced warping, precession and truncation of accretion discs in binary systems: three-dimensional simulations. Monthly Notices of the Royal Astronomical Society 282, 597-613.

Levison, H. F., Morbidelli, A. 2007. Models of the collisional damping scenario for ice-giant planets and Kuiper belt formation. Icarus 189, 196-212.

Lissauer, J. J. 1993. Planet formation. Annual Review of Astronomy and Astrophysics 31, 129-174.

Marzari, F., Thébault, P., Scholl, H. 2009. Planet formation in highly inclined binaries. Astronomy and Astrophysics 507, 505-511.

McLaughlin, D. B. 1924. Some results of a spectrographic study of the Algol system.. The Astrophysical Journal 60, 22-31.

Morbidelli, A. 2002. Modern celestial mechanics : aspects of solar system dynamics. Modern celestial mechanics : aspects of solar system dynamics, by Alessandro Morbidelli. London: Taylor & Francis, 2002, ISBN 0415279399

Morbidelli, A., Tsiganis, K., Crida, A., Levison, H. F., Gomes, R. 2007. Dynamics of the

Giant Planets of the Solar System in the Gaseous Protoplanetary Disk and Their Relationship to the Current Orbital Architecture. The Astronomical Journal 134, 1790-1798.

Morton, T. D., Johnson, J. A. 2011. Discerning Exoplanet Migration Models Using Spin-Orbit Measurements. The Astrophysical Journal 729, 138.

Mugrauer, M., Neuhäuser, R., Ginski, C., Eisenbeiss, T. 2010. Multiplicity study of exoplanet host stars. Highlights of Astronomy 15, 694-694.

Murray, C. D., Dermott, S. F. 1999. Solar system dynamics. Solar system dynamics by Murray, C. D., 1999 .

Nagasawa, M., Ida, S., Bessho, T. 2008. Formation of Hot Planets by a Combination of Planet Scattering, Tidal Circularization, and the Kozai Mechanism. The Astrophysical Journal 678, 498-508.

Naef, D., and 11 colleagues 2001. HD 80606 b, a planet on an extremely elongated orbit. Astronomy and Astrophysics 375, L27-L30.

Neishtadt, A. 1984. The separation of motions in systems with rapidly rotating phase. Journal of Applied Mathematics and Mechanics 48, 133-139.

Raymond, S. N., Armitage, P. J., Gorelick, N. 2009. Planet-Planet Scattering in Planetesimal Disks. The Astrophysical Journal 699, L88-L92.

Thébault, P., Marzari, F., Augereau, J.-C. 2010. Debris discs in binaries: a numerical study. Astronomy and Astrophysics 524, A13.

Thomas, F., Morbidelli, A. 1996. The Kozai Resonance in the Outer Solar System and

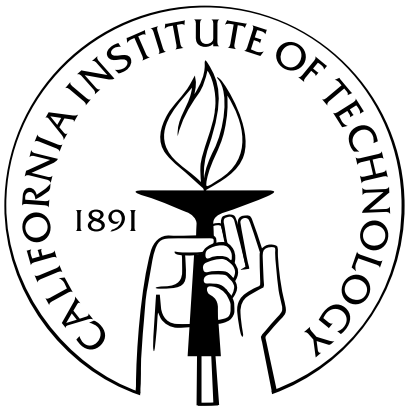
the Dynamics of Long-Period Comets. Celestial Mechanics and Dynamical Astronomy 64, 209-229.

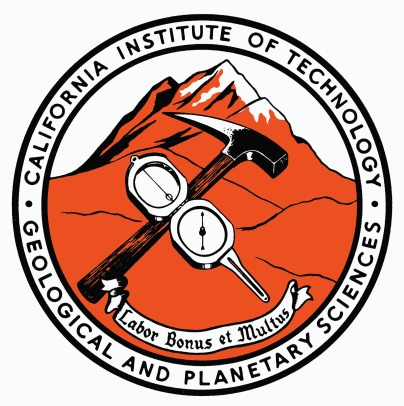
Thommes, E. W., Duncan, M. J., Levison, H. F. 1999. The formation of Uranus and Neptune in the Jupiter-Saturn region of the Solar System. Nature 402, 635-638.

Tsiganis, K., Gomes, R., Morbidelli, A., Levison, H. F. 2005. Origin of the orbital architecture of the giant planets of the Solar System. Nature 435, 459-461.

Winn, J. N., Fabrycky, D., Albrecht, S., Johnson, J. A. 2010. Hot Stars with Hot Jupiters Have High Obliquities. The Astrophysical Journal 718, L145-L149.

Wu, Y., Murray, N. 2003. Planet Migration and Binary Companions: The Case of HD 80606b. The Astrophysical Journal 589, 605-614.


CALIFORNIA INSTITUTE OF TECHNOLOGY
1891


CALIFORNIA INSTITUTE OF TECHNOLOGY
GEOLOGICAL AND PLANETARY SCIENCES
Labor Bonus et Multus