Capture of planets into mean-motion resonances and the origins of extrasolar orbital architectures

Konstantin Batygin

Division of Geological and Planetary Sciences, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA

Accepted 2015 May 7. Received 2015 May 6; in original form 2015 April 3

ABSTRACT
The early stages of dynamical evolution of planetary systems are often shaped by dissipative processes that drive orbital migration. In multi-planet systems, convergent amassing of orbits inevitably leads to encounters with rational period ratios, which may result in establishment of mean-motion resonances. The success or failure of resonant capture yields exceedingly different subsequent evolutions, and thus plays a central role in determining the ensuing orbital architecture of planetary systems. In this work, we employ an integrable Hamiltonian formalism for first order planetary resonances that allows both secondary bodies to have finite masses and eccentricities, and construct a comprehensive theory for resonant capture. Particularly, we derive conditions under which orbital evolution lies within the adiabatic regime, and provide a generalized criterion for guaranteed resonant locking as well as a procedure for calculating capture probabilities when capture is not certain. Subsequently, we utilize the developed analytical model to examine the evolution of Jupiter and Saturn within the protosolar nebula, and investigate the origins of the dominantly non-resonant orbital distribution of sub-Jovian extrasolar planets. Our calculations show that the commonly observed extrasolar orbital structure can be understood if planet pairs encounter mean-motion commensurabilities on slightly eccentric ($e \sim 0.02$) orbits. Accordingly, we speculate that resonant capture among low-mass planets is typically rendered unsuccessful due to subtle axial asymmetries inherent to the global structure of protoplanetary discs.

Key words: methods: analytical – celestial mechanics – planets and satellites: dynamical evolution and stability.

1 INTRODUCTION
The orbital architectures of the Solar system’s planetary and satellite populations, as well as the currently known aggregate of extrasolar planets, exhibit numerous peculiar features. In principle, each specific attribute of the galactic planetary census entails delicate constraints that inform the dominant mechanisms responsible for its inception and subsequent evolution. Therefore, theoretical interpretation of the observed state of planetary systems ultimately holds the key to understanding their origins.

Among the most striking characteristics inherent to the orbital distribution of planets and satellites is the modest predisposition for orbital resonances, or loosely speaking, pairs of orbits with nearly rational periods. In the Solar system, the preference for commensurability among satellite pairs (over randomly distributed orbits) is well-recognized, and was first pointed out more than half a century ago by Roy & Ovenden (1954) (see also Goldreich 1965; Dermott 1973).

In the extrasolar realm, the radial-velocity sub-sample points to a mild overabundance of resonant giant planets, typically at orbital radii in excess of $\sim 1$ au (Wright et al. 2011; Winn & Fabrycky 2014). Meanwhile, the period-ratio distribution of an extensive collection of sub-Jovian planets discovered by Kepler and other surveys exhibits notable enhancements at values slightly outside of exact 2:1 and 3:2 commensurabilities (Fabrycky et al. 2014). Instances of multi-resonant chains, akin to the Galilean satellites, have also been found.1

Collectively, these observations imply that planet formation environments can be congenial towards emergence of resonant systems. However, the overwhelming majority of observed commensurabilities are unlikely to be truly primordial in nature. Instead, resonances generally result from dissipative convergent evolution of pairs of...

1 Examples of such systems include GJ 876, Kepler-79 and Kepler-223.
orbits that follows the initial phase of conglomeration (Goldreich 1965; Allan 1969, 1970; Sinclair 1970, 1972).

There exists a multitude of physical mechanisms that may cause orbits to slowly approach one-another. Specifically, migration facilitated by tidal dissipation (Goldreich & Soter 1966; Peale 1976, 1986; Yoder & Peale 1981), interactions with a gaseous disc (Goldreich & Tremaine 1980; Crida, Sándor & Kley 2008; Kley & Nelson 2012), and scattering of debris within a planetesimal swarm (Fernandez & Ip 1984; Malhotra 1993, 1995; Kirsh et al. 2009) are the most frequently quoted transport processes relevant to planets and satellites.

While convergent migration is required for resonances to congregate, this process alone does not guarantee successful resonant locking. In fact, the outcome of an encounter with a mean-motion commensurability depends on the intrinsic parameters of the system, such as the planet–star and planet–planet mass ratios, the convergence rate, as well as the degree of pre-encounter orbital excitation. However, in spite of the uncertainty in the prospect of resonant capture itself, the post-encounter evolution of the system may depend critically on its result.

As an example of resonances’ decisive nature, consider the fate of a pair of giant planets, at an epoch when the gaseous protoplanetary disc has not yet dispersed. Although individually the planetary orbits would decay towards the central star along with the accretionary flow of the gas, successful capture of such objects into mean-motion resonance can reverse the overall inward drift and allow the planets to remain at large orbital radii2 (Masset & Snellgrove 2001; D’Angelo & Marzari 2012). As another example, assembly of sub-Jovian planets into resonances facilitates stabilization by the phase-protection mechanism (Greenberg 1977) and allows chains of planets to coherently migrate inwards without encountering each other3 (Terquem & Papaloizou 2007; Cresswell & Nelson 2008; see also Lee & Peale 2002). On the contrary, failure of objects to capture into resonance leads to an impulsive change in the system’s orbital properties (Tittmore & Wisdom 1988, 1989, 1990) and may trigger large-scale dynamical instabilities (Tsiganis et al. 2005; Morbidelli et al. 2007; Batygin & Brown 2010). Thus, quantification of the capture probability of planets into mean-motion resonance is central to understanding the early evolution of planetary systems.

Over the last half century, numerous efforts have been made to develop a complete understanding of resonant capture. Following the initial studies of Goldreich (1965), Goldreich & Soter (1966), Allan (1969, 1970), Sinclair (1972), Greenberg (1973), the first analytical formulation of the capture probability within the framework of the circular restricted three-body problem (i.e. where a massive planet is assumed to orbit the central star on a circular trajectory and perturb a massless test-particle) was proposed4 by Yoder (1973, 1979) and independently by Neishtadt (1975). Fully employing the use of adiabatic invariants, the formalism for analytic determination of capture probabilities was generalized by Henrard (1982) (see also Henrard & Lemaître 1983; Lemaître 1984) and subsequently simplified by Borderies & Goldreich (1984) (see also Peale 1986; Malhotra 1988). Additional studies aimed at extending resonant capture theory to the non-adiabatic regime have since also been undertaken by Friedland (2001), Quillen (2006), Ketchum, Adams & Bloch (2011) and Ogihara & Kobayashi (2013).

While the paradigm of the circular restricted three-body problem lends itself easily to theoretical analysis, it is not directly applicable to the ultimately relevant issue of resonant capture of massive planets with non-circular orbits. Accordingly, to attack this more complicated problem, numerous authors have resorted to the use of numerical experiments. To this end, simulations that mimic the dissipative effects with fictitious forces (Lee & Peale 2002; Hahn & Malhotra 2005; Terquem & Papaloizou 2007; Reim & Papaloizou 2010) as well as self-consistent hydrodynamical (Kley et al. 2005; Papaloizou & Szuszkiewicz 2005; Crisida et al. 2008; Cresswell & Nelson 2008), self-gravitational (Moore, Quillen & Edgar 2008; Moore & Quillen 2011) and tidal calculations (Mardling 2008; Lithwick & Wu 2012; Batygin & Morbidelli 2013a) have been performed.

Although the published simulations represent important advancements in the understanding of planet formation and evolution, such calculations continue to leave analytical development to be desired for two reasons. First, numerical experiments are typically tailored towards particular systems with specific choices of physical parameters. Without the knowledge of underlying theoretical relationships, it becomes difficult to translate the obtained results to another case without additional modelling. Secondly, despite substantial advances in computational technologies, high-resolution numerical experiments remain much too computationally expensive to practically serve as a replacement for theory.

Consequently, in order to explain why planetary systems possess the orbital architecture that they have, it is necessary to generalize the existing resonance capture theory to the physical domain of the unrestricted elliptic three-body problem. This is the primary purpose of this work. Ultimately, with a more complete model in hand, we shall delineate parameter regimes within which resonant capture is expected to be successful, and aim to understand why resonant orbits are neither ubiquitous nor absent within the galactic planetary census.

The paper is structured as follows. In Section 2, we generalize the existing first-order resonance capture theory to allow for both planets to have masses as well as finite eccentricities. Additionally, we obtain a criterion which dictates whether or not the dissipative evolution lies in the adiabatic regime. In Section 3, we examine some immediate consequences of the model. Particularly, we construct maps of capture probabilities relevant to adiabatic encounters and derive specific adiabatic thresholds for disc-driven migration, planetesimal-driven migration, and tidal evolution. Subsequently, we analytically consider the orbital evolution of Jupiter and Saturn while submerged within the protosolar nebula, and derive critical eccentricities above which sub-Jovian extrasolar planets fail to capture into mean-motion resonances. We summarize and discuss our results in Section 4.

2 ANALYTICAL THEORY

The calculation we aim to perform is perturbative in nature. Accordingly, we begin with a basic description of the dynamics. It is standard practice in celestial mechanics to treat planet–planet interactions as perturbations to Keplerian motion. As such, planet–planet potential is typically expanded as a Fourier series in the orbital angles and a power series in the planetary eccentricities and inclinations (Laskar & Robutel 1995; Ellis & Murray 2000; Laskar & Boué 2010).
In the vicinity of a given resonance, it is sensible to average over all short-periodic terms (i.e. those that vary on an orbital time-scale) in the expansion, and retain only slowly varying (compared to the orbital period) harmonics, which include those associated with the resonant interaction (Murray & Dermott 1999). This procedure yields a simplified Hamiltonian that approximates the full Hamiltonian near a commensurability.

2.1 Model Hamiltonian

In its most rudimentary form (i.e. to first order in planetary masses and eccentricities), the planar \( k = 1 \) resonant Hamiltonian reads:

\[
H = -\frac{GM_1 m_1}{2a_1 - \frac{GM_2 m_2}{2a_2} - \frac{Gm_1 m_2}{a_2}} \times \left[ f_{1s}^{(1)} e_1 \cos(\kappa_2 - (k - 1)\lambda_1 - \sigma_1) \right.
+ \left. f_{2s}^{(2)} e_2 \cos(\kappa_2 - (k - 1)\lambda_1 - \sigma_2) \right].
\]

In the above expression, \( k \) is an integer greater than unity, \( G \) is the gravitational constant, \( M_s \) is the mass of the central star, \( m \) is the planetary mass, \( e \) is the semi-major axis, \( \lambda \) is the mean longitude, \( \sigma \) is the longitude of perihelion, and \( f_{ns}, s \) are functions of order unity that (weakly) depend on the semi-major axis, \( H \), the planetary mass, and \( \lambda \) (see Callegari & Yokoyama 2007 for explicit expressions). The subscripts 1 and 2 denote the inner and outer planets, respectively.

Upon inspection it is immediately clear that the Hamiltonian (1) contains two resonant arguments that appear in two separate harmonics. Although at first glance this Hamiltonian appears to comprise a typical two degree of freedom system, it was shown to be integrable by Sessin (1981) and Sessin & Ferraz-Mello (1984). The published calculations of Sessin & Ferraz-Mello (1984) are rather involved, as they were performed within the framework of the Hori (1966) perturbation method. However, the demonstration of integrability was greatly simplified by Wisdom (1986) and Henrard et al. (1986), who showed that instead of constructing formal perturbation series, it is possible to consolidate the two harmonics into a single term with the aid of a reducing canonical transformation\(^5\) that corresponds to a rotation in phase space (see also Henrard & Lemaître 2005).

By employing the reducing transformation, we shall cast the Hamiltonian into a form that is synonymous to a first-order Andoyer Hamiltonian, also known as the second fundamental model for resonance (Henrard & Lemaître 1983; Morbidelli 2002). In this work, we will not spell out the derivation in excruciating detail, as it is available elsewhere (Sessin & Ferraz-Mello 1984; Henrard et al. 1986; Wisdom 1986; Ferraz-Mello 2007; Batygin & Morbidelli 2013b; Deck, Payne & Holman 2013; Delisle, Laskar & Correia 2014). Instead, we shall restrict ourselves to sketching out the important steps. In both notation and substance, we shall closely follow the derivation outlined by Batygin & Morbidelli (2013b).

As a first step, we define the canonical variables

\[
\kappa = \Lambda_1 + \frac{k - 1}{k} \Lambda_2, \quad \kappa = \lambda_1
\]

\[
\Theta = \Lambda_2/k, \quad \theta = k\lambda_2 - (k - 1)\lambda_1
\]

\[
x_i = e_i \sqrt{\Lambda_i} \cos(-\sigma_i), \quad y_i = e_i \sqrt{\Lambda_i} \sin(-\sigma_i)
\]

where \( \Lambda = m_1 \sqrt{GM_1 a_1} \) and the square brackets \( [\cdot] \) denote an evaluation at nominal resonant semi-major axis (i.e. \( a_1 = (k/(k - 1))^3/2[a_1] \)). Upon expanding the Keplerian part of the Hamiltonian to second order in \( \Lambda \), (which corresponds to a perturbation that is first order in \( e \)), neglecting semi-major axis variation in the disturbing part of the Hamiltonian (since it is already first order in \( e \)), and dropping constant terms, the expression (1) takes the form:

\[
H = 3[h_1](k - 1)\kappa \Theta - \frac{3}{2} \left( [h_1](k - 1)^2 + [h_2]k^2 \right) \Omega^2
- (\alpha x_1 + \beta x_2) \cos(\Theta) + (\alpha y_1 + \beta y_2) \sin(\Theta).
\]

Here, \( [h] = [n]/[\Lambda] = 1/\sqrt{m_1 a_1^3} \) is the inverse moment of inertia of a circular orbit, and

\[
\alpha = \frac{G^2 M_1 m_1^3}{[\Lambda_1^2]} f_{1s}^{(1)} \sqrt{[\Lambda_1]}, \quad \beta = \frac{G^2 M_1 m_1^3}{[\Lambda_2^2]} f_{2s}^{(2)} \sqrt{[\Lambda_2]}
\]

are constants that encapsulate the strengths of the individual harmonics present in the Hamiltonian.

Hamiltonian (3) is independent of the angle \( \kappa \), meaning that the action \( \kappa \) is a constant of motion. However, the number of degrees of freedom is still too great for integrability. Defining a canonical rotation (Wisdom 1986; Henrard et al. 1986; Ferraz-Mello 2007)

\[
u_1 = \frac{\alpha x_1 + \beta x_2}{\sqrt{\alpha^2 + \beta^2}}, \quad v_1 = \frac{\alpha y_1 + \beta y_2}{\sqrt{\alpha^2 + \beta^2}},
\]

\[
u_2 = \frac{\beta x_1 - \alpha x_2}{\sqrt{\alpha^2 + \beta^2}}, \quad v_2 = \frac{\beta y_1 - \alpha y_2}{\sqrt{\alpha^2 + \beta^2}}
\]

and the associated action-angle coordinates

\[
u_i = \sqrt{2\Phi_i} \cos(\phi_i), \quad v_i = \sqrt{2\Phi_i} \sin(\phi_i),
\]

the Hamiltonian is reduced to a form that contains only a single harmonic:

\[
H = 3[h_1](k - 1)\kappa \Theta - \frac{3}{2} \left( [h_1](k - 1)^2 + [h_2]k^2 \right) \Theta^2
- \sqrt{\alpha^2 + \beta^2} \sqrt{2\Phi_1} \cos(\phi_1 + \Theta).
\]

This reduction identifies a second integral of motion, that is \( \Phi_2 \). Employing a contact transformation, we define a final set of action-angle variables

\[
\Psi = \Phi_1, \quad \psi = \phi_1 + \theta
\]

\[
\Omega = \Theta - \Psi, \quad \sigma = \theta
\]

which renders \( \mathcal{H} \) independent of \( \sigma \). Accordingly, identifying \( \Omega \) as the final conserved quantity and dropping constant terms, the Hamiltonian is cast into an integrable form:

\[
\mathcal{H} = \frac{3}{2} \left( [h_1](k - 1)^2 + [h_2]k^2 \right) \Psi^2
- \sqrt{\alpha^2 + \beta^2} \sqrt{2\Psi} \cos(\psi).
\]

The constants that precede each of the terms in the Hamiltonian are not truly independent. Thus, as a concluding step in the preparation of the Hamiltonian, we shall scale the action and time, in order to introduce a single resonance proximity parameter, \( \delta \). To maintain symplecticity, we scale all of the actions present in the

---

\(^5\) Evidently, a transformation of this sort was first proposed by Poincaré (1899).
Hamiltonian, as well as the Hamiltonian itself by the same constant factor, $\eta$:

$$\tilde{\mathcal{H}} = \mathcal{H}/\eta \quad \tilde{\Psi} = \Psi/\eta \quad \tilde{K} = K/\eta \quad \tilde{\Omega} = \Omega/\eta. \quad (10)$$

We choose the expression for $\eta$ such that the coefficient upfront the term that is quadratic in action is half as big as the one that precedes the harmonic:

$$\eta = \left(\frac{(a^2 + b^2)}{9[\eta_1((k-1)^2 + [\eta_1]k^2)^2]}\right)^{1/3} = m_1 m_2 \sqrt{GM_1 a_1} \times \left[\left(\frac{(f_{\text{rx}})^2 m_1 + (f_{\text{rx}})^2 (k/(k-1))^2 \mu_2}{9 M_2^2 (k^2 m_1 + (k-1)^2 k^2/\mu_2)^2}\right)^{1/3}\right]. \quad (11)$$

Subsequently, we change the unit of time to $3n([\eta_1((k-1)^2 + [\eta_1]k^2)/2$, and divide $\tilde{\mathcal{H}}$ by the same factor. Writing the proximity parameter as

$$\delta = -([\eta_1]k^2(3 + 2\tilde{\Omega}) + [\eta_1](k-1)) \times (3k - 2\tilde{K} + 2(k-1)\tilde{\Omega} - 3)) \times ([\eta_1](k-1)^2 + [\eta_1]k^2)^{-1}, \quad (12)$$

the Hamiltonian takes on the familiar form:

$$\tilde{\mathcal{H}} = 3(\delta + 1)\tilde{\Psi} - \tilde{\Psi}^2 - 2\sqrt{2\tilde{\Psi}} \cos(\tilde{\psi}). \quad (13)$$

With the exception of the signs,7 this is the Hamiltonian considered by Yoder (1973, 1979), Neishtadt (1975), Henrard (1982) and Borderies & Goldreich (1984). Note that this expression for the Hamiltonian is written in terms of dimensionless variables. Thus, the quantity $\eta$ (which has units of angular momentum) encapsulates all of the information regarding how the dynamics scales with mass ratios and physical sizes of the orbits.

### 2.2 Equilibria of the Hamiltonian

Some insight into the properties of the Hamiltonian can be obtained by considering its equilibria. Defining the Cartesian coordinates

$$p = \sqrt{2\tilde{\Psi}} \cos(\tilde{\psi}) \quad q = \sqrt{2\tilde{\Psi}} \sin(\tilde{\psi}), \quad (14)$$

the Hamiltonian can be written as:

$$\tilde{\mathcal{H}} = 3(\delta + 1) \left(\frac{p^2 + q^2}{2}\right) - \left(\frac{p^2 + q^2}{2}\right)^2 - 2p. \quad (15)$$

It can be understood from equation (15) that the equilibrium points must reside on the $p$-axis, since $\tilde{\psi} = 0, \pi$ for $\partial \tilde{\mathcal{H}}/\partial \tilde{\psi} = 0$. Thus, setting $q = 0$, the equilibrium equation reads:

$$3(\delta + 1)p - p^3 - 2 = 0. \quad (16)$$

The cubic equation (16) admits three roots, one of which is always real and negative, while the other two are real (and positive) only above a critical bifurcation value of $\delta$, which can be computed by setting the two positive roots equal to each other. For Hamiltonian (13), the bifurcation occurs at $\delta_C = 0$. Strictly speaking, $\delta \geq \delta_C$ defines a condition for the existence of resonances, since a homoclinic curve (i.e. separatrix), necessary for identification of resonant trajectories, only exists for values of $\delta$ greater than or equal to $\delta_C$ (see e.g. Delisle et al. 2012). To this end, it is worth noting that although the functional form for $\delta$ is complicated, it monotonically grows as orbits approach each other convergently. Thus, an increase in the value of $\delta$ unilaterally signals evolution into the resonance.

The three roots of equation (16) are shown as functions of the resonance proximity parameter. The elliptic equilibrium point that is always real and negative is shown in blue. The elliptic and hyperbolic fixed points that only exist above the bifurcation value of $\delta \geq \delta_C$ are shown in black and red, respectively. Extrinsically driven convergent evolution of the orbits corresponds to a gradual increase in $\delta$.

![Figure 1.](http://mnras.oxfordjournals.org) Equilibria of Hamiltonian (13). The three roots of equation (16) are shown as functions of the resonance proximity parameter. The elliptic equilibrium point that is always real and negative is shown in blue. The elliptic and hyperbolic fixed points that only exist above the bifurcation value of $\delta \geq \delta_C$ are shown in black and red, respectively. Extrinsically driven convergent evolution of the orbits corresponds to a gradual increase in $\delta$.

The three roots of equation (16) can be classified as follows. The negative root (shown with a blue line) always corresponds to a stable (elliptical) equilibrium point and, for $\delta \geq \delta_C$, is enveloped by resonant trajectories. The positive root, whose value is closer to zero (shown with a black line), also corresponds to a stable fixed point but lies at the centre of the inner circulation region of the phase-space portrait. The larger positive root (shown with a red line) corresponds to an unstable (hyperbolic) equilibrium. For reference, phase-space portraits of the Hamiltonian for $\delta = -1, \delta = \delta_C = 0$ and $\delta = 1$ are depicted in Fig. 2, where the fixed points are colour-coded in the same way as in Fig. 1.

Note that the unstable equilibrium defines a point where the inner and outer branches of the separatrix join. For this reason, its value is particularly important for the evaluation of capture probabilities. Written explicitly, the corresponding expression for the unstable fixed point reads:

$$p_{\text{eq}}^* = \left(\sqrt[3]{3\delta + \delta - i\sqrt{3}} \left(\sqrt{-\delta(\delta + 3) + 3} + 1\right)\right)^{2/3} \times \left(\sqrt{-\delta(\delta + 3) + 3} + 1\right)^{1/3} + i\sqrt{3} + 1 \times \left(2\sqrt[3]{-\delta(\delta + 3) + 3} + 1\right)^{-1}, \quad (17)$$

where $i = \sqrt{-1}$. Recall that this expression is purely real only for $\delta \geq 0$.

### 2.3 Conditions for guaranteed capture

As already stated above, within the context of Hamiltonian (13), the slow convergence of two initially non-resonant orbits towards

---

6 This is a slightly different scaling from that employed by Batygin & Morbidelli (2013b).

7 Signs only determine if the resonant equilibrium point resides at $\tilde{\psi} = 0$ or $\tilde{\psi} = \pi$, and are thus unimportant to the discussion at hand.
a mean-motion commensurability can be envisioned as a gradual increase in $\delta$ from values below $\delta_C$ to values above $\delta_C$ (Peale 1986). Provided that dissipative processes responsible for secular changes in planetary architecture act on much longer time-scales than the dynamical time-scales associated with resonant motion, we can define an adiabatic invariant (Henrard 1982; Neishtadt 1984)

$$\mathcal{J} = \oint \Psi d\psi = \oint p \, dq,$$

which is conserved as long as the trajectory does not encounter a homoclinic curve (Lichtenberg & Lieberman 1983). Physically, $\mathcal{J}$ corresponds to the area occupied by the orbit in phase space.

For a given value of $\delta$, the conservation of $\mathcal{J}$ dictates the energy level on which the trajectory resides. This is demonstrated in Fig. 2 where a set of orbits characterized by the same $\mathcal{J}$ are shown on phase-space portraits with different values of $\delta$. The condition for guaranteed capture into resonance stems directly from the fact that the adiabatic invariant is preserved as long as the orbit does not encounter a separatrix.

Of the three thickened circulating trajectories shown in panel A of Fig. 2, consider the one with the smallest radius (i.e. the orbit whose engulfed area is shaded blue and labelled $\mathcal{J}$). This orbit is bound to become a resonant trajectory (specifically the one corresponding to $\mathcal{J}$ in panels B and C) because the phase-space area it occupies is smaller than the area occupied by the separatrix at $\delta = \delta_C$ (we shall call this area $\mathcal{J}_C$), and the total phase-space area occupied by the resonance is minimized when the separatrix first appears (Henrard 1982; Morbidelli 2002). Furthermore, because the area occupied by resonant trajectories only grows with increasing $\delta$, the libration amplitude of the trajectory in question shall only decrease with time (Yoder 1979; Peale 1986). For reference, $\mathcal{J}_C$ is shown in green shade in Fig. 2(B).

Let us calculate $\mathcal{J}_C$. Setting $\delta = \delta_C$, from equation (17) we obtain $p^*_{eq} = 1$. Accordingly, the value of the Hamiltonian that corresponds to the separatrix, $\mathcal{H}^*_{eq}$, is given by

$$\mathcal{H}^*_{eq} = 3(\delta_C + 1)\left(\frac{p^*_{eq}}{2}\right)^2 - \left(\frac{p^*_{eq}}{4}\right)^4 - 2p^*_{eq} = -\frac{3}{4}. \quad (19)$$

With a value of $\mathcal{H}^*_{eq}$ in hand, the relationship between $\tilde{\Psi}$ and $\psi$ along the separatrix at $\delta = \delta_C$ can be obtained from equation (13).

Integration along the homoclinic curve shows that the phase-space area occupied by the separatrix is given by

$$\mathcal{J}_C = 6 \left[ \arcsin \left( \frac{1}{(p^*_{eq})^{3/2}} \right) + \frac{\pi}{2} \right] = 6\pi. \quad (20)$$

The same procedure yields the area occupied by any trajectory (identified by the value of $\mathcal{H}$) for any value of $\delta$. The evaluation of $\mathcal{J}$ is particularly trivial in the limit where $\delta \ll \delta_C$. In this case, the phase-space portrait is composed of concentric circles, centred on the stable fixed point near the origin. Therefore, provided some value of $\Psi = \frac{p^2}{2}$, we have

$$\mathcal{J} = \pi \cdot p^2 = 2\pi \cdot \Psi. \quad (21)$$

Naturally, this limit is quite relevant for convergently migrating planets or satellites, as it represents the dynamical state of the system far away from mean-motion commensurability.

The above discussion indicates that the evaluation of $\Psi$ far away from resonance is sufficient to gauge whether or not guaranteed resonant capture will take place. The calculation of $\Psi$ itself requires the specification of planetary eccentricities, longitudes of perihelia

---

*Alternatively, an equivalent relationship between $p$ and $q$ can be obtained from equation (15).*

---

**Figure 2.** Evolution of the phase-space portrait of the Hamiltonian during convergent orbital migration. Panels A, B and C depict level curves of equation (13) for $\delta = -1$, 0, and 1, respectively. The equilibrium points are shown with small dots and are colour-coded in the same way as in Fig. 1. While libration of the critical angle is possible in panel A, formal resonant trajectories only exist in panels B and C, where the separatrix is shown as a thick red curve. Across the panels, three values of the phase-space area (adiabatic invariant) are accentuated: sub-critical (labelled $\mathcal{J}$ and shown as a shaded blue region), critical (labelled $\mathcal{J}_C$ and emphasized with a green shade in panel B), and super-critical (corresponding to the area occupied by the outer branch of the separatrix on panel C). Trajectories encircling the critical and super-critical areas are shown on panel A as thickened red lines. A circulating orbit engulfing a super-critical area is similarly shown in panel B. While smooth adiabatic entry into resonance is possible for initially circulating trajectories occupying a phase-space area equal to, or smaller than $\mathcal{J}_C$, those occupying a larger area must cross the separatrix to be captured, as shown on panel C. On panel B, the minimal and maximal excursions of the action along the separatrix (which correspond to the resonance width at the onset of resonance) are labelled $p^*_{eq}$ and $p_*$, respectively.
as well as planet–star mass ratios. Provided the appropriate quantities, we can write the expression for \( \Psi \) in terms of orbital elements and thus formulate a practically useful criterion for guaranteed adiabatic capture within the framework of the unrestricted three-body problem. Specifically, we have

\[
\left[ \frac{3 \xi M_\odot}{2 (k - 1)^2 m_2 + k^2 m_1 \xi} \right]^{2/3} \left( \frac{f_{\text{res}}^{(1)}}{m_2 + \xi} \right)^2 e_1^2 + \left( \frac{f_{\text{res}}^{(2)}}{m_2 + \xi} \right)^2 e_1^2 \leq 6,
\]

which is more readily interpreted, is conserved on the secular dynamics of the planets outside of the resonant domain. Moreover, neither \( \Psi \) nor \( \Phi_2 \) (identifying equation 5) is individually conserved in the secular domain, and large-scale variations of the two quantities are indeed possible in principle (see Batygin & Morbidelli 2013b). On the other hand, the angular momentum deficit

\[
A = \eta \Psi + \Phi_2 = m_1 \sqrt{G M_\odot [a]} \frac{e_1^2}{2} + m_2 \sqrt{G M_\odot [a]} \frac{e_2^2}{2},
\]

which is more readily interpreted, is conserved on the secular domain.

Because \( \Phi_2 \geq 0 \), we have: \( \Psi \leq A/\eta \). Taking advantage of this, we can formulate an excessively strong, but very simple version of the criterion for guaranteed capture:

\[
\frac{A}{\eta} \leq 3.
\]  

It is important to note that the maximal extent of orbital excitation that allows for guaranteed capture is a function of the secondary to primary mass ratios. In particular, equations (22) and (24) imply that the maximum allowed eccentricities decrease with decreasing mass ratio, a relationship that may be anticipated intuitively.

### 2.3.1 Circular restricted problem as a special case

Although expression (22) is derived here for the first time, a related criterion for guaranteed adiabatic capture is well known within the framework of the circular restricted three-body problem (see Murray & Dermott 1999). Provided that equation (22) is a more general criterion, it should reduce to the restricted limit as one of the planetary masses is taken to vanish, while the orbit of the other planet is assumed to approach a circle.

Considering interior resonances first, we set \( m_1 = 0 \) and \( e_2 = 0 \) in equation (22) to obtain:

\[
e_1 \leq \sqrt{6} \left( \frac{m_2}{3M_\odot} \frac{f_{\text{res}}^{(1)}}{(k - 1)^{3/2}} \right)^{1/3}.
\]  

Similarly, for the case of exterior resonances we set \( m_2 = 0 \) and \( e_1 = 0 \), which yields

\[
e_2 \leq \sqrt{6} \left( \frac{m_1}{3M_\odot} \frac{f_{\text{res}}^{(2)}}{k^{3/2}} \right)^{1/3}.
\]  

Both of these expressions are in direct agreement with those derived by assuming a circular restricted formalism from the beginning (Peale 1986; Malhotra 1988).

It is noteworthy that within the framework of the restricted problem, the aforementioned capture criteria for interior and exterior resonances are obtained from different Hamiltonians. Specifically, the analysis typically begins with a variant of equation (1) where only a single resonant harmonic is identified. Subsequently, the corresponding expressions are individually cast into the form of Hamiltonian (13) and a calculation of capture probabilities is carried through. In this work, we have utilized the reducing transformation (5) to derive the previously known criteria as limiting cases of a more comprehensive formalism. Our analysis thus generalizes previous results.

### 2.4 Probabilistic capture

If the criterion (22) is not satisfied, resonant capture is not certain. This is because at sufficiently high values of the action \( \Psi \), the phase-space area occupied by the orbit exceeds the critical area occupied by the separatrix at the inception of the resonance, i.e. \( J > J_C \) at \( \delta = \delta_c \). Consequently, unlike the ‘smooth’ transition to resonance discussed above, in this case the trajectory must cross the separatrix in order to reach the resonant domain. However, passage across a critical curve is a fundamentally probabilistic process, and may advect the trajectory into the inner circulation region of the phase-space portrait.

To determine the probabilistic outcome of an encounter of a \( J > J_C \) trajectory with a homoclinic curve, one must first relate the non-resonant initial conditions to the value of the proximity parameter at which the transition will take place. Utilizing the conservation of the adiabatic invariant (equation 21), we can connect the value of the action \( \Psi_0 \) far away from resonance (i.e. at \( \delta = \delta_c \), where the trajectory in phase space can be approximated by a circle) to the value of the proximity parameter at which the encounter will take place \( \delta_{\text{enc}} \). This can be done by matching the phase-space area occupied by the trajectory to that engulfed by the separatrix at \( \delta_{\text{enc}} \). The appropriate expression reads (Henrard 1993):

\[
\Psi_0 = \frac{1}{2} \left[ \frac{3 \xi M_\odot}{2 (k - 1)^2 m_2 + k^2 m_1 \xi} \right]^{2/3} \left( \frac{f_{\text{res}}^{(1)}}{m_2 + \xi} \right)^2 e_1^2 \times \left( \frac{f_{\text{res}}^{(2)}}{m_2 + \xi} \right)^2 e_2^2 + 2 f_{\text{res}}^{(1)} f_{\text{res}}^{(2)} e_1 e_2 \cos(\Delta \varpi) \right]
\]

\[
= \frac{3}{\pi} \left( \delta_{\text{enc}} + 1 \right) \left( \frac{1}{\left( p_{\text{eq}}^* \right)^{3/2}} \right) \right] \]  

\[
+ \sqrt{\left( p_{\text{eq}}^* \right)^3 - 1} \right].
\]  

Recalling that \( p_{\text{eq}}^* \) is a single-valued function of \( \delta \) (see equation 17), the above expression unequivocally links the starting condition of the system to the phase-space portrait at which the encounter takes place. Fig. 3(A) shows \( J/2\pi = \Psi_0 \) as a function of \( \delta_{\text{enc}} \), obtained from equation (27).

With the above relationship at hand, we can evaluate the probability of resonant capture for arbitrary initial conditions. Generally, adiabatic capture probability, \( P \), is given by the ratio of the rates of changes of the phase-space areas occupied by the resonant domain of the phase-space portrait to that occupied by the outer branch of the separatrix (see Fig. 2 C). Following Borderies &
Goldreich (1984), the explicit expression\(^9\) for \(\mathcal{P}\) that corresponds to the Hamiltonian (13) reads:

\[
\mathcal{P} = 2 \left[ 1 + \frac{\pi}{2 \arcsin \left( \frac{\rho_{\text{eq}}}{\sqrt{3}} \right)^{3/2}} \right]^{-1}. \tag{28}
\]

Fig. 3(B) depicts the resonant capture probability as a function of \(\tilde{\Psi}_0\). It should be noted that in the case of probabilistic capture, the transition through the separatrix necessarily breaks the conservation of \(J\). This means that the resonant encounter will lead to an impulsive change in orbital parameters, irrespective of whether capture is successful. To this end, note that passage through the resonance (see Fig. 2 C) diminishes the overall phase-space area, implying a less dynamically excited post-encounter state.

\subsection{2.5 The adiabatic threshold}

The entirety of the above discussion assumes that the dissipative evolution of the proximity parameter is so slow that at any point, a ‘frozen’ (i.e. constant \(\delta\)) representation of the dynamics provides a good approximation to the real evolution (Wisdom 1985; Henrard & Caranicolas 1990). This approximation holds in a regime where the resonant libration frequency greatly exceeds the extrinsic resonance crossing rate. Within the framework of the adopted model, both quantities can be analytically derived, thereby providing an explicit criterion for adiabatic dynamics.

To calculate the resonant libration frequency, we expand the Hamiltonian (13) to second order in \(\tilde{\Psi}\) and \(\tilde{\psi}\) around the equilibrium point \((\tilde{\Psi}, \pi)\) that corresponds to the definite negative root of equation (16). Dropping constant terms, we have

\[
\tilde{\mathcal{H}} = -\left[ 1 + \frac{1}{2 \sqrt{2}[\tilde{\Psi}]^{3/2}} \right] (\tilde{\Psi} - [\tilde{\Psi}])^2 - \sqrt{2[\tilde{\Psi}]}(\tilde{\psi} - \pi)^2
+ \left( 3 + 3\delta - 2[\tilde{\Psi}] + \sqrt{\frac{2}{[\tilde{\Psi}]}} \right) (\tilde{\Psi} - [\tilde{\Psi}]). \tag{29}
\]

The fixed-point condition dictates that in the above expression, the term linear in \(\tilde{\Psi}\) must vanish. This yields a relationship between the proximity parameter and the equilibrium action:

\[
\delta = \frac{2}{3} [\tilde{\Psi}] - \frac{1}{3} \sqrt{\frac{2}{[\tilde{\Psi}]}} - 1. \tag{30}
\]

Introducing new canonical variables (e.g. Batygin & Morbidelli 2013b):

\[
\Psi = (\tilde{\Psi} - [\tilde{\Psi}]) \left( \frac{\sqrt{2} + 4[\Psi]^{1/2}}{4 \sqrt{2}[\Psi]^{1/2}} \right)^{1/4}, \\
\tilde{\psi} = (\psi - \pi) \left( \frac{\sqrt{2} + 4[\Psi]^{1/2}}{4 \sqrt{2}[\Psi]^{1/2}} \right)^{-1/4}, \tag{31}
\]

we obtain a Hamiltonian that is synonymous to that of a simple harmonic oscillator:

\[
\tilde{\mathcal{H}} = -\frac{\omega}{2} (\Psi^2 + \tilde{\psi}^2), \tag{32}
\]

where

\[
\omega = \sqrt{\frac{2 + 4[\Psi]^{1/2}}{[\Psi]}}. \tag{33}
\]

is identified as the resonant libration frequency (Lichtenberg & Lieberman 1983).

Utilizing equation (30), \(\omega\) can be expressed as a function of the proximity parameter, \(\delta\), and this relationship is shown in Fig. 4. It is trivial to show that the value of \(\omega\) is minimized at \(\delta = -1\) (corresponding to \([\tilde{\Psi}] = 2^{-1/3}\)) and quantitatively evaluates to \(\omega_{\text{min}} = 2^{2/3}\sqrt{3}\). For larger values of \(\delta\), \(\omega\) slowly increases. For example, at the onset of resonance (i.e. at \(\delta = 0\) when the separatrix first appears), \(\omega = 3\). Recalling the scalings introduced in equations (10) and (11), we adopt the value of the libration frequency at \(\delta_C\) to formulate a practically useful criterion for adiabatic evolution.\(^{10}\)

\(^{10}\)Non-adiabatic evolution will cause the trajectory to jump to the inner circulation region without being trapped into resonance. So, it is sensible to consider the libration frequency at a value of \(\delta\) at which the inner circulation region first appears, i.e. \(\delta_C\). At this value of \(\delta\), \(\omega\) exceeds its minimum value by \(~10\) per cent.
of the crossing points is the bifurcated equilibrium, from equations (17) and (19), we have
\[
\Delta \Psi = \left( \frac{p_s}{2} - \frac{p_s}{2} \right) \eta = 4 \eta,
\]
where \(p_s = -3\) signifies the negative intersection point of the separatrices.

From conservation of the integrals \(K\) and \(\Omega\) (see equations 2 and 8), it follows that the maximal resonant variation of the first Poincaré momenta is
\[
\Delta \Lambda_1 = -(k - 1) \Delta \Psi \eta
\]
\[
\Delta \Lambda_2 = k \Delta \Psi \eta.
\]

The quantity of interest for our calculation is the ratio of mean motions:
\[
\chi = \frac{n_2}{n_1} \simeq \frac{[n_1]^2}{[n_1]} \left( 1 + 3 \frac{(\Lambda_1 - [\Lambda_1])}{[\Lambda_1]} - 3 \frac{(\Lambda_2 - [\Lambda_2])}{[\Lambda_2]} \right).
\]

Accordingly, the above expressions define a resonance width in \(\chi\):
\[
\Delta\chi_{\text{res}} = \frac{[n_1]^2}{[n_1]} \left( 1 - 3(k-1)\Delta \Psi \eta \frac{\Delta \Psi \eta}{[\Lambda_1]} - 3 k \frac{\Delta \Psi \eta}{[\Lambda_1]} \right)
\]
\[
= -12 \eta \frac{k([\Lambda_1] + [\Lambda_2]) - [\Lambda_1]}{[\Lambda_1] [\Lambda_2]}
\]

Let us now turn to dissipative evolution. A conventional way to parametrize extrinsically facilitated drifts in semi-major axes is to adopt the following form (Lee & Peale 2002):
\[
\frac{da_1}{dt} = \frac{a_1}{\tau_1} \quad \frac{da_2}{dt} = \frac{a_2}{\tau_2}.
\]

From these expressions, it follows that the associated rate of change in the ratio of mean motions is
\[
\frac{d\chi}{dt} = -\frac{3}{2} \frac{(\tau_1 - \tau_2)}{\tau_1 \tau_2} \frac{n_2}{n_1}.
\]

By replacing the derivative on the LHS by a fraction of finite differences, and equating the change in \(\chi\) to the resonance width, we obtain a corresponding segment of time required for dissipative forces to carry the orbits across the resonance (Friedland 2001; Goldreich & Schlichting 2014):
\[
\Delta t \approx 8 \eta \frac{[\Lambda_1] [\Lambda_2]}{\tau_1 \tau_2} \frac{k([\Lambda_1] + [\Lambda_2]) - [\Lambda_1]}{\tau_1 [\Lambda_1]}
\]

The condition for adiabatic evolution is thus defined: the characteristic time-scale of orbital convergence, \(\Delta t\), must exceed the libration period \(\tau_{\text{lib}}\). In terms of physical parameters, the adiabatic criterion reads:
\[
\frac{\tau_{\text{lib}}}{\Delta t} = \frac{(k-1)^{4/9} \pi (\tau_1 - \tau_2)}{2 (3)^{2/3} [n_1]^2} \frac{\tau_1 \tau_2}{M_* \left( \frac{(k-1)/k)^{1/3} \left( f_{11}^{(2)2} \right)^2 m_1 + \left( f_{11}^{(2)} \right)^2 m_2 \right)^{2/3}} \left( \frac{(k-1)/k)^{1/3} \left( f_{11}^{(2)2} \right)^2 m_1 + \left( f_{11}^{(2)} \right)^2 m_2 \right)^{2/3}
\]

Similarly to the preceding discussion of the expression for \(\tau_{\text{lib}}\), equation (43) is approximately independent of how mass is
distributed between the secondary bodies. Instead, the criterion depends sensitively on the ratio of the total planetary mass to the mass of the central object. Accordingly, in the same spirit as equation (35), we can obtain a simplified expression for the adiabatic criterion:

$$\frac{\tau_{\text{lib}}}{\Delta t} \approx \frac{\pi}{2|n_1|} \frac{(\tau_1 - \tau_2)}{\tau_1 \tau_2} \left( \frac{M_s}{m_1 + m_2} \right)^{4/3} \times \left[ \frac{(k - 1)^{-2/9}}{(\sqrt{3} f_{\text{rot}})^{1/3}} \right] \lesssim 1. \tag{44}$$

We note, however, that dependence on the planet–planet mass ratio can still enter equations (43) and (44) through particular expressions for $\tau_1$ and $\tau_2$.

The analytical expression for the adiabatic threshold derived above exhibits the characteristic $-4/3$ power-law dependence on the mass ratio. Unsurprisingly, the same power law holds for the restricted problem (Friedland 2001; Quillen 2006; Goldreich & Schlichting 2014), and has been observed to arise in numerical $N$-body simulations (Ketchum et al. 2011; Ogihara & Kobayashi 2013). Fig. 6(A) shows the dependence of the adiabatic criterion on the distribution of masses, while the dependence on the cumulative secondary to primary mass ratio is shown in Fig. 6(B). Also over-plotted on panel B are numerical estimates of the adiabatic threshold for two equal-mass bodies corresponding to the $2:1$ resonance obtained by Ogihara & Kobayashi (2013). Clearly, the agreement between theory and simulation is more than satisfactory.

3 RESULTS

With the crucial features of the analytical theory defined, we are now in a position to examine the generic consequences of the model. Additionally, we shall aim to understand how observations of exoplanetary systems and simulations of the early dynamical evolution of the Solar system fit into the framework of the developed formalism. We shall begin by outlining some generalities.

3.1 General results

3.1.1 Capture probability maps

The formulae presented in the previous section allow for the evaluation of $P$ at a computationally negligible cost. Taking advantage of this, we are provided with an opportunity to rapidly delineate the probability of capture in a given section of parameter space, a task that would be impossible to accomplish with brute-force numerical simulations. To this end, the only quantity we need to specify is $\Psi$ far away from resonance. However, $\Psi$ itself represents a rather complicated combination of orbital and physical parameters, and involves the specification of planet–star and planet–planet mass ratios, orbital eccentricities, as well as the difference in the longitudes of perihelia.

Although the specification of mass ratios and eccentricities is unavoidable, one would ideally like to be agnostic with respect to specifying the difference in longitudes of perihelia of the orbits, $\Delta \sigma$. Unfortunately, the outcome of resonant encounters is not insensitive to $\Delta \sigma$. As an example, consider a pair of planets with star–planet mass ratios equal to that of Jupiter and Saturn in the vicinity of a $2:1$ commensurability (this example will also be relevant to the discussion presented below). Utilizing equation (22), we can delineate curves that correspond to guaranteed capture on a $e_1$ versus $e_2$ diagram, as shown in Fig. 7. Clearly, parameter space associated with certain capture is minimized for $\Delta \sigma = \pi$ and maximized for $\Delta \sigma = 0$.

Thus, if the knowledge of $\Delta \sigma$ is readily accessible, one may proceed with the estimation of capture probabilities in the conventional fashion. If not however, it is sensible to assume that this difference of angles is uniformly distributed between 0 and $2\pi$, and correspondingly average over $\Delta \sigma$, leading to the following definition of the apsidally averaged capture probability:

$$\langle P \rangle = \frac{1}{2\pi} \int_0^{2\pi} P \, d\Delta \sigma. \tag{45}$$

With this definition at hand, the number of input parameters is reduced to four (specifically, they are the total secondary mass $m_1 + m_2$, the secondary mass ratio $m_1/m_2$, and the two orbital
K. Batygin

Figure 6. The adiabatic threshold as a function of plane–planet and planet–star mass ratios. As in Fig. 5, panel A depicts the dependence of the adiabatic criterion on the distribution of mass among the planets, while panel B shows its \(-4/3\) power-law dependence on the ratio of the total planet mass to the central mass. In addition to the theoretical relationships, numerically computed adiabatic threshold data corresponding to equal-mass bodies in a 2:1 resonance (Ogihara & Kobayashi 2013) are shown with blue points. Curves corresponding to the 2:1, 3:2 and 4:3 resonances are shown as red, blue, and green lines, respectively.

Figure 7. Conditions for guaranteed adiabatic capture of Jupiter and Saturn into the 2:1 mean-motion resonance. Each of the plotted curves corresponds to a distinct assumed difference in the apsidal lines, \(\Delta \varpi\). The \(e_1 - e_2\) parameter space associated with certain capture is clearly minimized at \(\Delta \varpi = \pi\) and maximized at \(\Delta \varpi = 0\). We have calculated the averaged probability across a range of parameters for the 2:1, 3:2 and the 4:3 mean-motion resonances. In particular, we have mapped out \(\langle P \rangle\) on a chain of \(e_1\) versus \(e_2\) diagrams, corresponding to total secondary masses of \((m_1 + m_2)/M_\star = 10^{-3}, 10^{-4}, 10^{-5}\) and secondary mass ratios of \(m_1/m_2 = 1/9, 1, 9\). The results are shown in Figs 8–10.

Upon examination of Figs 8–10, a few features immediately stand out. First, the overall capture probability decreases with \(k\), meaning that capture into the 2:1 mean-motion resonance is the most generous with respect to the degree of pre-encounter orbital excitation. A second readily apparent quality is the weak dependence of \(\langle P \rangle\) on the distribution of masses among the two bodies. That is, the difference between the values of eccentricities that characterize guaranteed capture is smaller than a factor of 2, depending on whether the majority of the mass resides in the inner or the outer planet. Moreover, the dependence on \(m_1/m_2\) further subsides with increasing \(k\). Thirdly, it is clear that the probability of capture most strongly depends on the total planetary mass of the system.

The construction of capture probability maps, in principle, removes the necessity for computationally expensive numerical experiments, provided that the assumptions inherent to the theory are satisfied for the problem at hand. The results also highlight an important attribute of resonant dynamics: while substantial orbital excitation (\(e \sim 0.1\)) does not inhibit capture for giant planets, rather minute (\(e \sim 0.02\)) eccentricities may be sufficient to obstruct the onset of resonant evolution of Earth-like terrestrial planets. We shall return to this point again below.

3.1.2 Critical rates of convergent evolution

An implicit assumption inherent to the calculation of capture probability maps depicted in Figs 8–10 is that orbital convergence is much slower than angular momentum and energy exchange via resonant interaction. As shown in the previous section, it is possible to write down an analytic criterion for adiabatic evolution. Accordingly, we shall now apply this criterion to the most common mechanisms responsible for orbital migration (namely interactions of planets with gaseous discs, scattering of planetesimals, and tidal evolution), and derive a series of constraints on physical parameters inherent to each physical setting.
Figure 8. A series of maps of the apsidally averaged capture probability corresponding to the 2:1 mean-motion resonance, for a variety of planet–planet and planet–star mass ratios. The colour scale marks the averaged capture probability, with orange and blue corresponding to high and low values, respectively. Additionally, on each plot a $\langle P \rangle = 100$ per cent, curve below which capture is certain, is shown with a thick red line, while a series of contours corresponding to $\langle P \rangle = 75$ per cent, 50 per cent, 35 per cent, 25 per cent, 15 per cent, 10 per cent are shown as gold curves. Note that the range of the plots increases with total secondary to primary mass. Specifically, $e_{\text{max}} = 0.08, 0.2, 0.35$ for $(m_1 + m_2)/M_\star = 10^{-5}, 10^{-4}, 10^{-3}$, respectively.

(i) Disc-Driven Migration
Gravitational interactions between a planet and a gaseous disc are well-known to yield a time-irreversible exchange of angular momentum (Goldreich & Tremaine 1980). At the detailed level, the rate and direction of migration are determined by the physical properties of the disc (e.g. density and entropy profiles) as well as the planetary mass (Paardekooper et al. 2010; Bitsch & Kley 2011; Paardekooper, Baruteau & Kley 2011). Depending on the latter, disc-driven orbital migration typically falls into one of two12 distinct regimes (Ward 1997; Papaloizou & Larwood 2000). For planets that are not sufficiently massive to clear out their own orbital neighbourhoods and thereby carve out gaps in their gaseous discs, migration proceeds in the so-called ‘type-I’, or linear regime. We consider this mode of orbital transport first.

A rudimentary way to define a planetary mass below which type-I migration applies is to consider the viscous gap-opening criterion:

$$\frac{m}{M_\star} \lesssim \sqrt{\frac{27\pi}{8}} \left(\frac{h}{a}\right)^5 \bar{\alpha},$$

where $h$ is the scale-height of the disc, and $\bar{\alpha}$ is the Shakura–Sunayev viscosity parameter (Armitage 2010). The quoted expression is

12 For the purposes of this work, we shall neglect the so-called ‘type-III’ mode of orbital transport.
Figure 9. Same as Fig. 8, but for the 3:2 mean-motion resonance.

derived by equating resonant and viscous torques in the planetary neighbourhood. A more sophisticated treatment of the gap-opening process can be found in Crida, Morbidelli & Masset (2006).

Under the simplifying assumptions of a power-law surface density profile and an isothermal equation of state, the direction of type-I migration is strictly inwards, and occurs on a characteristic time-scale (Tanaka, Takeuchi & Ward 2002):

$$\tau_I = -\frac{1}{s} \frac{M_2^2}{m \Sigma a^2} \left(\frac{h}{a}\right)^2 \frac{1}{n},$$

(47)

where $s$ is a dimensionless constant that depends on the structure of the disc.

Motivated by observations of Andrews et al. (2013), here we shall assume that the surface density profile takes the form of a classical Mestel disc, i.e. $\Sigma = \Sigma_0 a_0/a$, where $\Sigma_0$ is the surface density at some reference semi-major axis $a_0$ (Mestel 1963). With this choice for $\Sigma$, we have $s = 3.8$ (Tanaka et al. 2002). Additionally, for simplicity we shall assume that the disc aspect ratio, $h/a$, is constant throughout the region of interest.

From the form of equation (47), it is immediately clear that if two planets are undergoing type-I orbital decay, a necessary requirement for convergent migration is $m_2 > \zeta m_1$. Assuming that this criterion is satisfied, we evaluate the decay time-scales (47) at nominal resonance, such that $a_1 = ((k-1)/k)^{2/3} = \zeta^2 a_2$, and employ equation (44) to obtain an expression for the adiabatic threshold:

$$\frac{M_{\text{disc}}}{M_*} \lesssim \frac{4}{s} \left(\frac{h}{a}\right)^2 \left(\frac{k-1}{k}\right)^{11/9} \times \frac{3^{2/3} (m_1 + m_2)^{4/3}}{((k-1)^2 k)^{1/3} m_1 - (k-1)m_2} M_1^{1/3},$$

(48)
where $M_{\text{disc}} = 2\pi \Sigma_0 a_0 a_2$ is the disc mass contained interior to the outer planetary orbit. This relationship is depicted for a particular mass ratio and disc aspect ratio in Fig. 11.

It is worth noting that while the resonant dynamics themselves tend to only depend on the cumulative planetary mass, the above expression is indeed strongly dependent on the mass ratio. The additional dependence on the disc surface density profile and separation from the central star yields a diverse range of physical parameters which may fall either into the strongly adiabatic or strongly non-adiabatic regimes (see e.g. Papaloizou & Szuszkiewicz 2005; Cresswell & Nelson 2008; Ketchum et al. 2011).

In this definition, it is implicitly assumed that $a_2$ greatly exceeds the truncation radius of the disc.

Having considered convergent migration of sub-Jovian planets, let us now examine the case where the planetary mass is sufficient to gravitationally clear out a substantial gap around its co-orbital neighbourhood. In this situation, the planet is shepherded towards the central part of the gap where all torques instantaneously cancel. Maintaining this configuration, the planet drifts inwards along with the accretionary flow of the gas (Lin & Papaloizou 1986; Crida & Morbidelli 2007; Morbidelli & Crida 2007). Therefore, the orbital evolution rate in this so-called ‘type-II’ regime is largely controlled by the global angular momentum transport within the disc, and

13 In this definition, it is implicitly assumed that $a_2$ greatly exceeds the truncation radius of the disc.

14 Strictly speaking, this is only true if the planet mass is not large enough to disrupt the process of viscous accretion. In other words, the dominant portion of the angular momentum budget within the planet’s neighbourhood must reside in the disc material. Quantitatively, this criterion is satisfied.
Two requirements bracket the product of decay proceeds at a slower rate than type-II migration of the inner body. Consequently, the outer body must not be so small that its type-I orbital migration will never experience convergent resonant encounters. Instead, convergent evolution is only possible if a less massive object (that migrates in the type-I regime) on an external orbit catches up to a more massive, slowly migrating planet. For this scenario to hold true, the outer planet ($m_2$) must fail the gap-opening criterion (46).

For resonant capture to be successful within the framework of this setup, the system must satisfy two additional criteria. Specifically, the outer body must not be so small that its type-I orbital decay proceeds at a slower rate than type-II migration of the inner body. Simultaneously, the outer body must not be massive enough for its migration rate to overwhelm the adiabatic limit (43). These two requirements bracket the product of $m_2$ and $M_{\text{disc}}$ from below and above:

$$\frac{3\pi}{s} \frac{k}{k - 1} \left(\frac{h}{a}\right)^4 \tilde{\alpha} \lesssim \left(\frac{m_2}{M_*}\right) \left(\frac{M_{\text{disc}}}{M_*}\right).$$

$$\lesssim 4 \left(\frac{27(k - 1)}{s}\right)^{2/3} \left(\frac{f_{\text{esc}}}{M_*}\right)^{4/3} \left(\frac{h}{a}\right)^2.$$

A graphic representation of this criterion for a Jupiter-mass inner planet is shown in Fig. 12.

In obtaining equation (50), we have made a series of approximations. In particular, we used a simplified form of the adiabatic limit as before. Subsequently, we assumed that the majority of the planetary mass is contained within the inner body. Finally, for the sake of a less cumbersome expression, we assumed that convergent evolution is dominated by type-I migration of the outer body in the second inequality. These simplifications generally hold as long as the bracketed quantity is not too close to either limit.

$$\left|m\right| \lesssim 4\pi \Sigma a^2,$$ where $\Sigma$ is the gas surface density immediately outside the gap (Baruteau et al. 2014).

Figure 11. The adiabatic threshold for two objects undergoing type-I migration in an isothermal disc with a Mestel-like surface density profile. While this particular figure shows the adiabatic limit for a secondary mass ratio of $m_1/m_2 = 2/3$ ($\xi = 3/5$), it should be kept in mind that unlike the expression for $\tau_{\text{lib}}$, equation (48) does depend strongly on this quantity.

Here, $\nu$ is the kinematic viscosity of the gas (Armitage 2010).

Equation (49) implies that within the context of the envisioned (simplified) picture, two planets that are simultaneously undergoing type-II migration will never experience convergent resonant encounters. Instead, convergent evolution is only possible if a less massive object (that migrates in the type-I regime) on an external orbit catches up to a more massive, slowly migrating planet. For this scenario to hold true, the outer planet ($m_2$) must fail the gap-opening criterion (46).

For resonant capture to be successful within the framework of this setup, the system must satisfy two additional criteria. Specifically, the outer body must not be so small that its type-I orbital decay proceeds at a slower rate than type-II migration of the inner body. Simultaneously, the outer body must not be massive enough for its migration rate to overwhelm the adiabatic limit (43). These two requirements bracket the product of $m_2$ and $M_{\text{disc}}$ from below and above:

$$\frac{3\pi}{s} \frac{k}{k - 1} \left(\frac{h}{a}\right)^4 \tilde{\alpha} \lesssim \left(\frac{m_2}{M_*}\right) \left(\frac{M_{\text{disc}}}{M_*}\right).$$

$$\lesssim 4 \left(\frac{27(k - 1)}{s}\right)^{2/3} \left(\frac{f_{\text{esc}}}{M_*}\right)^{4/3} \left(\frac{h}{a}\right)^2.$$

A graphic representation of this criterion for a Jupiter-mass inner planet is shown in Fig. 12.

In obtaining equation (50), we have made a series of approximations. In particular, we used a simplified form of the adiabatic limit as before. Subsequently, we assumed that the majority of the planetary mass is contained within the inner body. Finally, for the sake of a less cumbersome expression, we assumed that convergent evolution is dominated by type-I migration of the outer body in the second inequality. These simplifications generally hold as long as the bracketed quantity is not too close to either limit.

$$\left|m\right| \lesssim 4\pi \Sigma a^2,$$ where $\Sigma$ is the gas surface density immediately outside the gap (Baruteau et al. 2014).

Figure 12. The adiabatic threshold for two objects, where the inner planet has the planet–star mass ratio identical to Jupiter and migrates in the type-II regime, while the outer planet of mass $m_2$ undergoes type-I migration in an isothermal Mestel-like disc. The red, blue, and green curves labelled by resonant period ratios correspond to the adiabatic criteria and are given by the RHS of equation (50). The intersection between the cyan line (given by the LHS of equation 50) and the grey curve of slope 2 denotes the minimum mass, below which type-II migration proceeds at a faster rate than type-I migration, leading to divergent evolution. The assumed disc parameters are quoted in the figure and roughly mimic those of the minimum mass solar nebula. Note that for the utilized fiducial parameter choices, Saturn fails the adiabatic criterion for the 2:1 resonance with Jupiter, but satisfies the threshold that corresponds to the 3:2 resonance. However, capture into the 2:1 resonance is possible, given a diminished surface density profile.

For typical disc parameters (e.g. $h/a \sim 0.05; \tilde{\alpha} \sim 0.001$), the extrema in equation (50) differ by a little more than an order of magnitude. However, it should be noted that without invoking unreasonable quantities, it is possible to make the range of planetary masses for which adiabatic capture may take place, exceptionally small.

(ii) Planetesimal-Driven Migration

In the post-nebular stage of planetary system evolution, torques derived from interactions with the gas are non-existent. However, large-scale migration can still occur as a consequence of asymmetric scattering of planetesimals by a more massive object (Fernandez & Ip 1984; Malhotra 1995). Within the context of the early dynamical evolution of the outer Solar system, this process may have played a crucial role, as scattering-facilitated divergent evolution of Jupiter and Saturn has been invoked as a means of igniting a transient dynamical instability (Tsiganis et al. 2005; Batygin & Brown 2010).

A prominent setting where planetesimal-driven migration can lead to substantial orbital changes is a gas-depleted planetesimal disc with embedded planetary embryos. In light of this, it is worthwhile to examine if resonant locking among such embryos can occur. A fiducial time-scale on which planetesimal-driven migration operates is (Ida et al. 2000; Kirsh et al. 2009)

$$|\tau_p| = \frac{M_*}{2 \Sigma a^2 \bar{n}}.$$
where \( \Sigma \) now refers to planetesimal surface density, for which we have once again assumed a Mestel-like radial profile. A detailed study of planetesimal-driven migration was recently carried out by Minton & Levison (2014), who identified five criteria that must be satisfied for this mode of orbital transport to be self-sustained. Here, we shall not dwell extensively on these requirements, but simply assume that they are satisfied for at least one of the secondary objects of interest.

The direction of planetesimal-driven migration can be inward or outward. To this end, Kirsh et al. (2009) find a preference for orbital decay over growth; however Minton & Levison (2014) find no statistically significant predisposition for either direction. From equation (51), it is clear that the inner orbit changes on a shorter time-scale. Therefore, if both orbits decay at rates given by equation (51), the evolution is bound to be divergent. Reversing this argument, one may envisage that convergent evolution is possible if both orbits grow. However, in this case the inner planet would be migrating in the wake of the outer planet, through a pre-excited disc. Without invoking very rapid dissipation that would dynamically cool the disc, such a scenario would likely violate the disc eccentricity criterion (see Minton & Levison 2014), halting the inner orbit’s migration. Consequently, it would appear that given nominal parameters, convergent evolution is possible either if the inner orbit evolves outwards while the outer evolves inwards, or if one of the orbits remains stationary while being approached by the other.

In principle, each of these situations will be characterized by its own, subtly different adiabatic threshold. However, in interest of succinctness we shall only consider the case where both orbits actively approach one another, noting that the corresponding expressions for the case where one of the planets remains stationary are quantitatively similar, within a factor of \( \sim 2 \). Combining expressions (51) and (43), we have

\[
\frac{M_{\text{disc}}}{M_*} \lesssim \frac{2^{2/3}}{\pi} \frac{k}{k - 1 + (k(k - 1)^{2})^{1/3}} \times \left( \frac{f^1_0(m_1 + m_2)}{M_*} \right)^{4/3}. \tag{52}
\]

It is interesting to note that while expression (52) provides an upper bound on the disc mass, the mass-ratio criterion for operation of planetesimal-driven migration requires the mass of the migrating body to not exceed its encounter mass (equivalently, isolation mass; Lissauer 1987) by more than a factor of \( \sim 3 \) (Kirsh et al. 2009), thereby providing a lower bound. Given that the multiplicative factor of \((m_1 + m_2)/M_*\) is of the order of unity for all resonances of interest and that the stopping mass is approximately \((m/M_\text{stop}) \approx 12(M_{\text{disc}}/M_*)^{3/2}\), we can formulate the following rough criterion for scattering-facilitated adiabatic resonant encounters:

\[
\frac{M_{\text{disc}}}{M_*} \lesssim \left( \frac{m_1 + m_2}{M_*} \right)^{4/3} \lesssim 27 \left( \frac{M_{\text{disc}}}{M_*} \right)^2. \tag{53}
\]

Despite its crudeness, the above expression is informative, and is shown in Fig. 13. In particular, it dictates that in order to simultaneously satisfy the adiabatic criterion and the mass-ratio criterion, the planetesimal disc mass must exceed \(M_{\text{disc}}/M_* \gtrsim 0.04\). Such a mass ratio is characteristic of gaseous protoplanetary discs, and thus exceeds the mass ratio inherent to the solid component of the disc by approximately two orders of magnitude. Therefore, equation (53) implies that without appealing to special configurations (such as the primordial multi- resonant state of the outer Solar system), planetary migration facilitated by scattering of planetesimal cannot lead to convergent adiabatic resonant encounters.

![Figure 13](http://mnras.oxfordjournals.org/Downloaded from). The adiabatic threshold for planetesimal-driven migration. As in previous figures, the adiabatic criteria are shown with red, blue, and green lines. However, these curves lie well below the mass-ratio criterion (shown with a cyan line) which must be exceeded for migration to be self-sustaining. This means that planetary embryos that migrate through long-lived planetesimal discs generally do not suffer adiabatic resonant encounters.

\[Q \lesssim \frac{1}{k^2} \left( \frac{a_*}{R_p} \right)^5 \frac{3^{1/3} \pi}{2} \frac{M_p}{f^{(1)}(m_1 + m_2)^{4/3}} \times \frac{(k^{1/3} m_1 - (k - 1)^{1/3} m_2)}{(k - 1)^{4/3} M_p}. \tag{55}
\]

\[16 \text{ See, however, Peale & Lee (2002) and Canup & Ward (2002) for an alternative, circumplanetary disc-driven view on the assembly of the Laplace resonance among the Galilean satellites.}

\[M_{\text{disc}}/M_* \lesssim \left( \frac{m_1 + m_2}{M_*} \right)^{4/3} \lesssim 27 \left( \frac{M_{\text{disc}}}{M_*} \right)^2. \tag{53}
\]

\[\text{Figure 13. The adiabatic threshold for planetesimal-driven migration. As in previous figures, the adiabatic criteria are shown with red, blue, and green lines. However, these curves lie well below the mass-ratio criterion (shown with a cyan line) which must be exceeded for migration to be self-sustaining. This means that planetary embryos that migrate through long-lived planetesimal discs generally do not suffer adiabatic resonant encounters.}

\[\text{planetary migration facilitated by scattering of planetesimal cannot lead to convergent adiabatic resonant encounters.}

\[\text{(iii) Tidal Evolution}

A final mode of convergent migration that we shall consider here is that facilitated by tidal dissipation. This mechanism is particularly relevant to planetary satellites, as orbital changes induced by tides raised on the host planet are considered to be the dominant driving mechanism\(^{16}\) responsible for resonant pairs of satellites in the Solar system (Goldreich 1965; Peale 1976, 1986; Henrard 1983). Correspondingly, we shall frame the following discussion in the satellite–planet context, keeping in mind that similar arguments can apply to close-in planets and their host stars (see e.g. Adams & Bloch 2015).

\[\text{Following the works of Allan (1969) and Dermott, Malhotra & Murray (1988), we shall assume that changes in the semi-major axis induced by satellite tides can be neglected in favour of their planetary counterparts. In this case, the characteristic migration time-scale for an orbit residing beyond the synchronous radius is given by (Murray & Dermott 1999)}

\[\tau_1 = \frac{1}{k^3} \frac{Q M_p}{k^2 R_p} \left( \frac{a}{R_p} \right)^5 \left( \frac{1}{n} \right), \tag{54}
\]

where \(Q \) is the specific dissipation function, \(k_2\) is the Love number, and \(R_p\) is the physical radius of the planet.

\[\text{Contrary to the case of type-I migration discussed above, the expression (54) implies that the outer satellite mass cannot exceed the inner satellite mass by more than } m_2 \lesssim m_1 (k/k - 1)^{1/2} \text{ for evolution to remain convergent. Substitution of the tidal time-scale into equation (44) yields the following expression for the adiabatic threshold:}

\[\frac{Q}{k^2} \left( \frac{a_*}{R_p} \right)^5 \lesssim \left( \frac{3}{2} \right)^{3/2} \frac{M_p}{f^{(1)}(m_1 + m_2)^{4/3}} \times \frac{k^{1/3} m_1 - (k - 1)^{1/3} m_2}{(k - 1)^{4/3} M_p}. \tag{55}\]
Figure 14. The adiabatic threshold corresponding to tidal evolution of an equal-mass pair of satellites. For illustrative purposes, in this figure we considered the inner object’s orbit to approximately coincide with the Roche radius, yielding a maximal migration rate. Even in this extreme case, a reasonable dissipation efficiency will yield adiabatic evolution.

To evaluate the physical meaning of equation (55), it is instructive to consider a limiting parameter regime that maximizes the migration rate. Accordingly, let us envision that the total secondary mass is concentrated entirely in the inner satellite which orbits immediately outside the Roche limit [i.e. \(a_2 \approx (5/2) R_p\)], and the planet is a homogenous fluid body with a Love number of \(k_2 \approx 3/2\). Setting the specific dissipation function to a minimal\(^{17}\) value of \(Q \sim 10\), we can obtain a critical mass ratio \((m_1/M_p)_{\text{crit}}\) below which the adiabatic limit is broken. Upon evaluation, we obtain \((m_1/M_p)_{\text{crit}} \sim 10^{-7}, 10^{-5}\), and \(10^{-3}\) for the 2:1, 3:2, and 4:3 resonances, respectively.

While these values already lie well below the typical range of natural\(^{19}\) satellite–planet mass ratios, it is important to keep in mind that we have invoked unreasonably rapid tidal evolution (as an example, for Jupiter, \(k_2/Q \sim 10^{-5}\) whereas we have used \(k_2/Q \sim 3/20\); Goldreich & Soter 1966; Lainey et al. 2009) in order to derive them. In other words, the obtained quantities are gross overestimates of the practically relevant critical mass ratios. Recall also that the rate of tidal evolution diminishes with \((m/M)\) meaning that the critical mass ratio may also lie in a regime where \(\tau_e\) greatly exceeds the age of the system, rendering this quantity meaningless. Therefore, for all practical purposes it is likely safe to assume that tidally facilitated resonant encounters will almost always lie in the adiabatic regime. For completeness, Fig. 14 shows the adiabatic threshold for an equal-mass pair of satellites.

The calculations presented above yield a series of practically useful criteria that inform whether or not convergent migration within a given physical setting adheres to the adiabatic limit. In interpreting this discussion however, it is important to keep in mind that for the sake of definitiveness, we have limited our calculations to simplified formulae that describe nominal dissipative evolution rates of planets. In reality, the detailed physics of disc–planet interactions, planetesimal scattering and tidal evolution can be rather complex (see Efroimsky & Williams 2009; Kley & Nelson 2012; Baruteau et al. 2014; Minton & Levison 2014 for in-depth discussions). Indeed, substantial deviations from the quoted prescriptions can occur within a more comprehensive treatment. As a result, the derived criteria should be viewed as approximate, rather than conclusive.

### 3.2 Specific applications

#### 3.2.1 Resonant capture of Jupiter and Saturn in the protosolar nebula

As already mentioned above, type-II migration generally causes planets to spiral in towards their host stars. In light of this, the following question naturally arises: why are Jupiter and Saturn where they are today? As an answer to this question, it was shown by Masset & Snellgrove (2001) that the simultaneous dynamics of Jupiter and Saturn, when submerged in the protosolar nebula, alter the isolated migration picture qualitatively. In particular, simulations show that Jupiter and Saturn exhibit convergent orbital evolution, and eventually get captured in resonance. With a resonant lock established, the planets carve out a mutual gap, which (owing to a particular planet–planet mass ratio) alters the torque balance of the system and facilitates joint outward migration of the resonant pair (Morbidelli & Crida 2007).

Although resonant reversal of orbital decay was initially proposed to account for the retention of Jupiter and Saturn at large orbital radii, such a sequence of events smoothly connects the nebular stage of Solar system evolution to the compact initial conditions of the Nice model (Morbidelli et al. 2007; Batygin & Brown 2010; Nesvorný & Morbidelli 2012), and provides natural explanations for the comparatively small mass of Mars (Walsh et al. 2011), the composition of the Asteroid belt (O’Brien et al. 2014), and the non-existence of otherwise common close-in Super-Earths inside of Mercury’s orbit (Batygin & Laughlin 2015). Moreover, it has been proposed that a similar mechanism is ubiquitously responsible for halting large-scale migration of giant exoplanets that reside at wide orbital separations (Morbidelli 2013).

Following the pioneering study of Masset & Snellgrove (2001), early dynamical evolution of Jupiter and Saturn has been studied by Morbidelli & Crida (2007), Pierens & Nelson (2008), D’Angelo & Marzari (2012), Pierens et al. (2014) using hydrodynamical simulations. In agreement with the initial experiments of Masset & Snellgrove (2001), most studies find that independent of initial conditions, the common outcome is passage through the 2:1 resonance and capture into the 3:2 resonance. The prevalent speculation (and indeed this turns out to be correct) within the aforementioned works is that failure of the 2:1 resonance to capture Jupiter and Saturn arises from non-adiabatic evolution. Given the well-defined nature of the problem, it is instructive to examine how the results of numerical simulations fit into the framework of the developed analytic theory.

As a first step, let us imagine that the adiabatic criterion is satisfied. In this regime, we can use equation (22) to inform the conditions under which resonant capture is not guaranteed. Specific curves corresponding to the 2:1 Jupiter–Saturn resonance are shown in Fig. 7. Evidently, even in the conservative case, pre-encounter eccentricities in excess of \(e \gtrsim 0.1\) are required to reduce the adiabatic capture probability below \(\sim 75\) per cent. Conventional theory (e.g. Goldreich & Tremaine 1980; Ward 1997) suggests that dissipative evolution of planets within axisymmetric discs tends to damp

\(^{17}\) The smallest possible value that \(Q\) can take on is unity. Physically, this would correspond to complete dissipation of all energy stored within a single tidal cycle. However, for the weak-friction theory [within the context of which equation (54) is obtained] to apply, \(Q\) must greatly exceed unity (Hut 1981).

\(^{18}\) Note that the assumptions of low eccentricity and inclination inherent to the formulation of the Hamiltonian (1) render our treatment largely inapplicable to irregular satellites of the Solar system.
orbital eccentricities and inclinations, meaning that capture into the 2:1 resonance is highly likely if the adiabatic limit is satisfied. Recall also that the 2:1 resonance allows for guaranteed capture at higher eccentricity than the 3:2 resonance, meaning that if the system fails to capture into the 2:1 resonance due to being too dynamically pre-excited, it will also likely fail to capture into the 3:2 resonance.

Let us now evaluate the critical rates of convergent evolution for Jupiter and Saturn in the protosolar nebula, as dictated by equation (50). Adopting the typically quoted aspect ratio of h/a = 0.05, a viscosity parameter of a = 10^{-3}, a surface density of Σ_0 = 2000 g cm^{-2} at a_0 = 1 au, and an encounter semi-major axis of a_{2} ≃ 2.5 au (the envisioned ‘tacking’ radius), the minimum mass of the outer planet above which the orbits will approach each other is m_{2} ≳ 3M_{⊕}. This is in excellent agreement with the simulation results of Pierens & Nelson (2008) who find the minimum mass to be m_{2} ≳ 3.5M_{⊕}.

Using the same expression we find the threshold mass, above which the adiabatic criterion is violated for the 2:1 resonance, is m_{2} ≳ 60M_{⊕}, or about 2/3 of Saturn’s mass. For the 3:2 resonance, the critical value increases to m_{2} ≳ 140M_{⊕}, which comfortably exceeds Saturn’s mass. Again, these estimates are in perfect alignment with the numerical experiments of Pierens & Nelson (2008) who observe capture into the 2:1 resonance for m_{3} = 30–40M_{⊕} and capture into the 3:2 resonance for m_{2} = 80–100M_{⊕}. Thus, our analytical theory robustly conforms to the results of numerical experiments and analytically demonstrates that the recurrent capture of Saturn into a 3:2 rather than a 2:1 mean-motion resonance is indeed an outcome of the violation of the adiabatic threshold for the 2:1, but not the 3:2 resonance.

Equation (50) further implies that capture into the 2:1 resonance is indeed possible, provided a reduced disc surface density. As an example, in addition to results corresponding to fiducial parameters quoted above, Fig. 12 also shows a supplementary (dashed) curve that represents a disc with Σ_0 = 900 g cm^{-2}. The fact that capture is permitted in this regime is fully consistent with the simulation results of Pierens et al. (2014), who numerically obtain capture into the 2:1 resonance for sub-nominal surface densities.

Naturally, the analytic model developed herein does not take into account the full richness of the possibilities that can occur in real protoplanetary discs. For example, simulations show that after Saturn passes through the 2:1 resonance, it directly invades the disc gap opened by Jupiter. As a result, the diminished local surface density of the gas further reduces its migration rate, allowing for an even more robust satisfaction of the adiabatic criterion (Morbidelli & Crida 2007). A similar reduction in Saturn’s migration rate can in principle occur if Saturn successfully opens its own gap, such that its migration rate is still faster than the strict type-II rate given by equation (49), but only slightly so. As is made clear by equation (46), this can be achieved by reducing the aspect ratio or the viscosity of the disc.

### 3.2.2 Paucity of resonances among close-in sub-Jovian planets

Having examined the well-studied problem of nebular evolution of Jupiter and Saturn, let us now consider a somewhat less transparent phenomenon. A pivotal realization that has emerged as a consequence of the recent identification of thousands of extrasolar planets by the Kepler transit survey is that the dominant mode of planet formation within the Galaxy is one that produces planets substantially smaller than Jupiter and Saturn with orbital periods below ~100 d (Youdin 2011; Howard et al. 2012). These close-in sub-Jovian planets are estimated to orbit approximately half of all Sun-like stars, and often comprise tightly packed multi-planet systems (Mayor et al. 2011; Batalha et al. 2013; Lissauer et al. 2014). Accordingly, understanding the architectural origins of such systems holds the key to identifying the dominant processes at play during the epoch of planet formation.

The orbital distribution of close-in sub-Jovian planets is for the most part devoid of orbital resonances. While there exist substantial enhancements in the periods ratios immediately outside of the 2:1 and 3:2 commensurabilities (and such systems can indeed be attributed to resonant evolution in presence of dissipative forces; Delisle et al. 2012; Lithwick & Wu 2012; Batygin & Morbidelli 2013a), these planet-pairs comprise a minority within the observational sample (Fabrycky et al. 2014). This fact is puzzling when viewed within the context of the theoretical expectation that type-I migration should typically generate crowded resonant chains in protoplanetary discs (Cresswell & Nelson 2008).

One way to prevent resonant capture is to invoke a sufficiently massive disc such that the adiabatic criterion inherent to type-I migration (equation 48) is not satisfied (Quillen 2006; Mustill & Wyatt 2011). Although this will indeed provide an effective avenue towards halting the production of resonant chains of planets, this process alone is unlikely to explain the observed exoplanet architectures in a satisfactory manner for two reasons. First, the disc mass itself decreases in time, generally on a time-scale that is much longer than typical type-I migration time-scales (Haisch, Lada & Lada 2001; Herczeg & Hillenbrand 2008). This means that even if a given disc starts out in a state where resonant capture is hampered by overly rapid migration, such a disc will eventually evolve to a state where the resonant condition (48) is satisfied. Secondly, it has been shown that even in massive discs, type-I migration is not ubiquitously fast, as suggested by conventional linear calculations (Paardekooper et al. 2010, 2011; Bitsch & Kley 2011; Bitsch, Boley & Kley 2013). The main deviation from linear theory arises from the fact that entropy gradients in radiative discs alter the disc–planet interaction in such a way as to create zones where type-I migration greatly slows down, reverses, or ceases altogether, allowing for convergent congregation of low-mass planets (McNeil, Duncan & Levison 2005; Papaloizou & Szuszkiewicz 2005; Matsumoto, Nagasawa & Ida 2012).

An alternative means of generating a largely non-resonant orbital distribution is to invoke processes that tend to destabilize mean-motion commensurabilities. To this end, Adams, Laughlin & Bloch (2008), followed by Ketchum et al. (2011) and Rein & Papaloizou (2009), proposed that stochastic perturbations due to turbulent forcing within protoplanetary discs may inhibit long-lived resonances (see also Laughlin, Steinacker & Adams 2004, for a related discussion). Another possibility, recently put forward by Goldreich & Schlichting (2014), is that the combined effects of semi-major axis and eccentricity damping may render resonant configurations unstable on long time-scales. While both of these proposals are in principle reasonable, they ultimately rely on processes that are poorly constrained. Specifically, these mechanisms inherently depend on quantities such as the duty-cycle of turbulent eddies and the relative damping rates of orbital eccentricities and semi-major axes, both of which remain scantily understood and constitute active fields of research (Bai & Stone 2013; Bitsch et al. 2015). As a result, in the spirit of Occam’s razor, it is tempting to inquire if a simpler mechanism that can preclude resonant capture exists.

---

19 See Goldreich & Sari (2003) and Tsang (2014) for an alternative view.
Keeping in mind the arguments presented above, let us only consider the adiabatic regime. As we already saw in the beginning of this section, capture is guaranteed only below some critical extent of pre-encounter orbital excitation, and capture probability decreases rapidly above this value (Fig. 3 B). How eccentric must have close-in planet pair been to preferentially skip over resonances without locking? To obtain a rough answer to this question, consider the following calculation.

Suppose that by default, planets in protoplanetary discs migrate not on exactly circular orbits but with some small eccentricities, \( \bar{e} \). Further, let us further speculate that \( \bar{e} \) will be similar for two planets that are about to encounter a mean-motion resonance, as it is set by external factors rather than direct planet–planet interactions. Provided that planetary and stellar masses are known, the critical value of pre-encounter eccentricity, \( \bar{e}_{\text{crit}} \), can be computed by means of equation (22), setting \( \Delta \sigma = \pi \). We have performed this calculation for all confirmed planet-pairs detected by the Kepler transit survey.

Fig. 15 shows the distributions of \( \bar{e}_{\text{crit}} \) corresponding to the 2:1, 3:2 and 4:3 resonances for Kepler planets. Clearly, the pre-encounter orbital eccentricities required to render capture uncertain are ubiquitously low: \( \bar{e}_{\text{crit}} \simeq 0.01–0.03 \). More importantly however, our simple estimate for \( \bar{e}_{\text{crit}} \) is astonishingly close to the observed rms eccentricities of the Kepler sample. Specifically, from their analysis of transit timing variations, Wu & Lithwick (2013) and Hadden & Lithwick (2014) find that three quarters of the considered sub-sample of planets conforms to an eccentricity distribution with an rms free eccentricity of \( e \sim 0.02 \), while the remaining quarter is characterized by substantially higher values. Although caution must be exercised in interpreting the present eccentricities of sub-Jovian planets as having been inherited from their natal discs, the conspicuous similarity of the theoretically required values to the corresponding observations is suggestive, and opens a previously unexplored avenue for explaining the origins of the orbital distribution.

4 DISCUSSION

Both theoretical and observational lines of inquiry suggest that mean-motion resonances play a central role in the formation and long-term evolution of planetary systems (see e.g. Morbidelli et al. 2007; Rivera et al. 2010; Deck et al. 2012; Goldreich & Schlichting 2014 and the reference therein). While this fact has been well recognized as a consequence of countless numerical experiments (Quillen 2006; Ketchum et al. 2011; Ogihara & Kobayashi 2013), until now a comprehensive qualitative understanding of the conditions for capture of planets into resonances had remained elusive. In this work, we have taken steps towards formulating a model for capture into resonance in a generic, analytical way. Our development provides an underlying framework within which the outcomes of numerical experiments can be interpreted. Moreover, we provide a series of simple criteria which can be used to inform the outcome of dynamical simulations.

As a practical recipe for theoretical analysis of resonant encounters, we propose the following order of calculation.

(i) Does the convergent evolution rate of the system satisfy the adiabatic condition? For the specific cases of orbital migration driven by interactions with a gaseous disc, scattering of planetesimals, or tidal evolution, expressions (48) or (50), (53), and (55) may be used, respectively. Alternatively, the generalized adiabatic criterion (43), or its simplified form (44) can be employed. If the criterion is violated, capture will not take place.

(ii) If the resonant encounter lies in the adiabatic regime, is capture certain? This can be assessed using expression (22). If knowledge of the differences in the apsidal lines, \( \Delta \sigma \), is available, the guaranteed capture equation can be used directly. If \( \Delta \sigma \) is not known, it is sensible to assume \( \Delta \sigma = \pi \), as this yields the most conservative estimate.

(iii) If adiabatic capture is not guaranteed, what is the capture probability? If \( \Delta \sigma \) is known, the initial condition far away from resonance can be related to the location of the unstable fixed point of the resonance at separatrix-crossing using expression (27). The corresponding probability can then be calculated using equation (28). If \( \Delta \sigma \) is unspecified, the averaged capture probability given by equation (45) can be calculated.

In this work, we have utilized the developed formalism to consider two specific applications. First, we analysed the numerically well modelled evolution of Jupiter and Saturn’s orbits in the primordial solar nebula (Pierens et al. 2014 and the references therein). Within the context of this problem, our theoretical arguments confirmed the previously insinuated notion (e.g. Morbidelli & Crida 2007) that the tendency of Jupiter and Saturn to lock into the 3:2 rather than 2:1 resonance is a consequence of Saturn’s rapid migration and the associated violation of the adiabatic criterion. We subsequently considered the origins of the dominantly non-resonant orbital distribution of close-in sub-Jovian multi-planet systems.
discovered by Kepler (Batalha et al. 2013; Fabrycky et al. 2014). Specifically, we showed that if low-mass planets generally reside on slightly eccentric orbits (i.e. $e \geq 0.02$) when submerged in their natal discs, the chances of resonant capture are greatly diminished. Although the theoretically derived threshold eccentricities (below which resonant capture is hindered) are almost identical to the present observationally inferred values within the Kepler sample (Wu & Lithwick 2013), we have not specified the physical origin of these subtle deviations from circular orbits. In principle, there exists a large number of dynamical mechanisms that may lead to such excitation. Among them are orbital excitations by turbulent forcing (Nelson & Papaloizou 2004; Adams & Bloch 2009) and interactions with distant massive planets which may themselves experience violent dynamical instabilities within protoplanetary nebulae (Lega, Morbidelli & Nesvorný 2013).

An arguably more intriguing idea is that steady-state non-zero eccentricities may stem directly from the interactions of planets with their natal disc. A body of recent literature has shown that contrary to the conventional theoretical simplification of perfectly circular protoplanetary nebulae, real discs exhibit substantial deviations away from axial symmetry and are indeed believed to be globally lopsided (Brown et al. 2009; Tang et al. 2012; Casassus et al. 2013; Fukagawa et al. 2013; Isella et al. 2013; van der Marel et al. 2013; Bruderer et al. 2014; Pérez et al. 2014). Although observations taken in the dust continuum wavelength band severely over-represent the corresponding non-axisymmetric over-densities of gas (which holds the vast majority of the disc mass), the gas disc eccentricities required to explain the ALMA observations (Mittal & Chiang 2015) may very well be sufficient to also perturb low-mass planets on to slightly non-circular orbits. While in this work, we have taken initial steps towards exploring this idea, undoubtedly much additional effort is required to quantitatively assess the viability of this hypothesis.

Our analytical study of resonant capture complements a number of recent developments that employ the same integrable formalism for resonant dynamics. In a closely related study (Batygin & Morbidelli 2013b), we showed how resonant dynamics can be represented in an intuitive geometric way, and demonstrated that divergent resonant encounters (where capture necessarily fails) leave the system in a persistent apsidally anti-aligned state. Using the same resonant encounters (where capture necessarily fails) leave the system in a persistent apsidally anti-aligned state. Using the same multiplicity (Duchêne & Kraus 2013).

(i) To obtain a better relationship between the specified orbital conditions far away from resonance and those at the resonant encounter, it may be fruitful to characterize the pre-encounter secular evolution. This may better inform the true value of the action $\mathcal{J}$ at the time of the appearance of the separatrix.

(ii) A more complete version of the model may incorporate the effects of external eccentricity damping or excitation (e.g. Lee & Peale 2002; Goldreich & Schlichting 2014). Direct modulation of the eccentricities will lead to corresponding changes in the value of the action and may cause $\mathcal{J}$ to shift from a value smaller than $\mathcal{J}_c$ to a value that exceeds it, or vice versa. It is possible that the inclusion of this effect will not affect the formulation of capture probabilities because one can envision defining time-dependent canonical coordinates, in which $\mathcal{J}$ is conserved by construction and the evolution rate of $δ$ is correspondingly accelerated or diminished (Henrard 1993). However, this assertion deserves to be tested explicitly.

(iii) In our analysis of probabilistic capture, we have adopted the conventional approach of considering the crossing of a well-defined separatrix, and thus neglected the effects of chaos. In fact, retention of higher order terms in eccentricity and inclination will increase the number of degrees of freedom, rendering the Hamiltonian non-integrable. In a system with multiple degrees of freedom, the vicinity of the separatrix may be engulfed in a chaotic layer, which may alter the outcome of resonant encounters (attempts at characterization of crossing of stochastic layer have been previously made by Henrard & Morbidelli 1993).

(iv) At present, resonant capture theory does not account for external stochastic forces. Such effects have, however, been shown to compromise the locking and longevity of resonances (Rein & Papaloizou 2009; Paardekooper, Rein & Kley 2013). A more complete theoretical framework for this reduction in capture probability may perhaps be constructed with the aid of stochastic calculus (see e.g. Adams et al. 2008; Batygin, Morbidelli & Holman 2015, for a related discussion).

(v) A somewhat more approximate (but similar in spirit to what has been done here) integrable model for second-order resonances has recently been presented by Delisle et al. (2014). Accordingly, the existing capture theory for second-order resonances in the restricted problem (Henrard 1982; Borderies & Goldreich 1984) can be adopted for the elliptic problem as well.

Cumulatively, it is clear that the extent of theoretical expansion that can potentially be undertaken is substantial. However, developments such as those proposed above will surely contribute to the construction of a comprehensive model for planetary system formation and evolution, and are thus well-motivated.

ACKNOWLEDGEMENTS

I am thankful to Katherine Deck, Chris Spalding, Peter Goldreich, Greg Laughlin, Heather Knutson, Geoff Blake, Renu Malhotra and Mike Brown for inspirational conversations, as well as to Alessandro Morbidelli, whose careful review of the manuscript and insightful suggestions led to a substantial improvement of the paper.

REFERENCES

Sessin W., 1981, PhD thesis (University of Sao Paulo)
Tittermore W. C., Wisdom J., 1988, Icarus, 74, 172
van der Marel N. et al., 2013, Science, 340, 1199
Wisdom J., 1985, Icarus, 63, 272
Wisdom J., 1986, Celest. Mech., 38, 175
Yoder C. F., 1973, PhD thesis (University of California Santa Barbara)
Yoder C. F., 1979, Celest. Mech., 19, 3

This paper has been typeset from a TeX file prepared by the author.