THEORY OF SECULAR CHAOS AND MERCURY'S ORBIT

Yoram Lithwick[1] and Yanqin Wu[2]
[1] Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA
[2] Department of Astronomy and Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada

ABSTRACT

We study the chaotic orbital evolution of planetary systems, focusing on *secular* (i.e., orbit-averaged) interactions, which dominate on long timescales. We first focus on the evolution of a test particle that is forced by multiple planets. To linear order in eccentricity and inclination, its orbit precesses with constant frequencies. But nonlinearities modify the frequencies, and can shift them into and out of resonance with either the planets' eigenfrequencies (forming eccentricity or inclination secular resonances), or with linear combinations of those frequencies (forming mixed high-order secular resonances). The overlap of these nonlinear secular resonances drives secular chaos. We calculate the locations and widths of nonlinear secular resonances, display them together on a newly developed map (the "map of the mean momenta"), and find good agreement between analytical and numerical results. This map also graphically demonstrates how chaos emerges from overlapping secular resonances. We then apply this newfound understanding to Mercury to elucidate the origin of its orbital chaos. We find that since Mercury's two free precession frequencies (in eccentricity and inclination) lie within ~25% of two other eigenfrequencies in the solar system (those of the Jupiter-dominated eccentricity mode and the Venus-dominated inclination mode), secular resonances involving these four modes overlap and cause Mercury's chaos. We confirm this with *N*-body integrations by showing that a slew of these resonant angles alternately librate and circulate. Our new analytical understanding allows us to calculate the criterion for Mercury to become chaotic: Jupiter and Venus must have eccentricity and inclination of a few percent. The timescale for Mercury's chaotic diffusion depends sensitively on the forcing. As it is, Mercury appears to be perched on the threshold for chaos, with an instability timescale comparable to the lifetime of the solar system.

Key words: chaos – planets and satellites: dynamical evolution and stability

Online-only material: color figures

1. INTRODUCTION

The question of the stability of planetary orbits in the solar system has a long history, and has attracted the attention of some of the greatest scientists, including Newton, Laplace, Lagrange, Gauss, Poincaré, Kolmogorov, and Arnol'd. Newton thought that interplanetary perturbations are eventually destabilizing, and that divine intervention is required to restore the planets' orbits to their rightful places (Laskar 1996). Yet it is only over the last 20 years that the stability of the solar system has been definitively settled, with the aid of computer simulations (Sussman & Wisdom 1988; Laskar 1989; Quinn et al. 1991; Wisdom & Holman 1991; Lecar et al. 2001; Laskar & Gastineau 2009). We now know that Newton was not far off: the solar system is *marginally stable*; it is unstable, but on a timescale comparable to its age. In the inner solar system, the planets' eccentricities chaotically diffuse on a billion-year timescale, with the two lightest planets, Mercury and Mars, experiencing particularly large variations. In fact, Mercury has roughly a 1% chance of colliding with Venus or the Sun within the next five billion years (Laskar & Gastineau 2009). By comparison, the giant planets in the outer solar system are well spaced, and their orbital elements undergo largely quasiperiodic variations, exhibiting chaotic diffusion only on extremely long timescales (Laskar 1996; Murray & Holman 1999).

The dynamics in the inner solar system is primarily due to *secular* interactions (Laskar 2008). In general, interplanetary interactions can be decomposed into secular ones and mean motion resonances (MMRs—not to be confused with secular resonances). Secular interactions result from orbit-averaging

the equations of motion. Since averaging a Keplerian orbit produces an elliptical ring, secular evolution can be thought of as interactions between elliptical rings. Secular timescales are long—they are longer than the orbital time by at least the ratio of the star's mass to that of a planet. By contrast, interactions driven by MMRs depend on orbital phase, and typically occur on the orbital timescale, or longer if some of the planets' orbital periods are close to integer ratios. Intuitively, one would expect that the dynamics on long timescales can be treated by averaging over the fast orbital phase—i.e., they are secular in nature. This is true in the inner solar system. It is also true more generally for well-spaced planets that do not happen to lie near MMRs.[3]

Linear secular theory has been understood for hundreds of years, dating back to the famous solution of Laplace and Lagrange (see Murray & Dermott 2000). To linear order in the planets' eccentricities and inclinations, secular theory reduces to a simple eigenvalue problem, with two eigenmodes per planet—one for the eccentricity degree of freedom, and one for the inclination. Each eigenmode has a constant amplitude and a longitude that precesses uniformly in time. However, secular interactions among solar system planets, even at their very low values of eccentricities and inclinations (a few percent), lead to chaos and instabilities that are clearly beyond the power of linear secular theory.

It has been suggested that secular chaos shapes the architecture of the inner solar system (Laskar 1996), as well as the

---

[3] In the outer solar system the dynamics is not mainly secular because the giant planets lie near a number of MMRs, such as the 5:2 between Jupiter and Saturn (the "Great Inequality"), and the 2:1 between Uranus and Neptune.

architecture of extra-solar planetary systems (Wu & Lithwick 2011). Surprisingly, despite the importance of secular chaos, there has been little theoretical understanding of it (see, e.g., the review of solar system chaos by Lecar et al. 2001). By contrast, chaos due to MMRs is well understood. Standard tools have been developed to calculate the width and location of these resonances (see, e.g., Murray & Dermott 2000), and it has been firmly established that the overlap between these resonances leads to chaos (Chirikov 1979; Lecar et al. 2001). These theoretical understandings have been successfully applied to explain, e.g., the Kirkwood gaps in the asteroid belt (Wisdom 1983), the very weak chaos of the outer solar system planets, which is due to three-body MMRs (Murray & Holman 1999), and the unstable orbital zones in binary star systems (Mudryk & Wu 2006).

An equivalent theory for secular chaos is needed. While an MMR occurs when the orbital mean motions (or their linear combinations) are integer multiples of each other, a secular resonance happens when the precession frequencies (or their linear combinations) are integer multiples of each other. It has been generally supposed that chaos in the inner solar system is caused by the overlap between these secular resonances. Yet thus far there has been little quantitative calculation. To our knowledge, the only previous theoretical work toward calculating secular chaos was by Sidlichovsky (1990) for the special case of coplanar systems (Section 3). For a general secularly interacting system, such as the solar system or the newly discovered Kepler 11 system with six transiting planets (Lissauer et al. 2011), it is desirable to analytically predict at what strength of interaction (i.e., values of eccentricity and inclination) resonance overlap occurs. However, how the locations and widths of the secular resonances depend on these interaction strengths is not yet understood. We set out to derive these dependencies in this work.

In the case of the inner solar system, numerical attempts to identify the mechanism of chaos were made by Laskar (1990, 1992) and Sussman & Wisdom (1992). These authors found that the angle associated with the (secular) frequency $(g_{\text{mercury}} - g_{\text{jupiter}}) - (s_{\text{mercury}} - s_{\text{venus}})$ alternately librated and circulated in their simulations, where g is the apsidal precession rate, and s is the nodal precession rate (or, to be more precise, g and s here refer to the frequencies of the normal mode that is dominated by the corresponding planet). Laskar (1992) also found that two angles associated with Earth and Mars, corresponding to $2(g_{\text{mars}} - g_{\text{earth}}) - (s_{\text{mars}} - s_{\text{earth}})$ and $(g_{\text{mars}} - g_{\text{earth}}) - (s_{\text{mars}} - s_{\text{earth}})$, alternately librated, and conjectured that the overlap of those secular resonances was responsible for chaos. But, as Sussman & Wisdom (1992) note, Laskar's conjecture is not fully convincing because there are too many unrelated angles that alternately circulate and librate,[4] and it is not clear which are dynamically important. Furthermore, only one librating angle has been identified for Mercury, yet chaos requires the overlap of at least two resonances. So why is Mercury chaotic (Lecar et al. 2001)? Without a theory for secular chaos, the dynamics remain obscure, and we will be forever at the mercy of computer simulations.

A theory for secular chaos will guide us in searching for the relevant angles, and more usefully, would illuminate the criterion for chaos. In the second half of this work, we will use our newly developed theory to answer the following questions:

why does instability in the solar system occur at such low values of eccentricity and inclination ($\sim$ few percent)? And, what sets the timescale of the chaos?

In this paper, we construct the theory for secular chaos of a test particle, and then apply the theory to Mercury. In Section 2, we present the test particle's equations of motion. In Section 3, we describe the coplanar solution, and in Section 4 we generalize to the case when bodies have non-zero inclinations. In Section 5, we apply the theory to N-body simulations of the real Mercury. We conclude in Section 6. Readers interested in an outline of the results are referred to Section 6 for an itemized summary.

2. SECULAR EQUATIONS OF MOTION

We focus on the secular evolution of a massless test particle that is orbiting a star in the presence of multiple massive planets, assuming the planets' orbits are known.

The particle has six orbital elements, $\{a, e, i, \lambda, \varpi, \Omega\}$, using standard notation (Murray & Dermott 2000). In secular theory, one averages over λ. As a consequence, a is a constant of motion, leaving only four orbital elements to be considered. The equations of motion for the particle's eccentricity and longitude of periapse (e and ϖ) are given by Hamilton's equations for the Poincaré canonical variables $\Gamma \equiv \sqrt{GM_\odot a}(1 - \sqrt{1 - e^2})$ and $\gamma \equiv -\varpi$ (Murray & Dermott 2000). Since a is constant, it is simpler to choose the canonical momentum to be $\propto \Gamma/\sqrt{GM_\odot a}$, so we introduce the momentum

$$p_e \equiv 2\big(1 - \sqrt{1 - e^2}\big) \qquad (1)$$

$$= e^2 + \mathcal{O}(e^4) \qquad (2)$$

and its conjugate coordinate to be ϖ. Although this is a non-canonical transformation from Poincaré's variables, if we simultaneously re-scale the energy by defining as the Hamiltonian

$$H \equiv -\frac{2}{\sqrt{GM_\odot a}}E, \qquad (3)$$

where E is the particle's energy per unit mass,[5] then the equations of motion are Hamilton's equations:

$$d\varpi/dt = \partial H/\partial p_e \qquad (4)$$

$$dp_e/dt = -\partial H/\partial \varpi. \qquad (5)$$

Therefore, we may consider (p_e, ϖ) to be canonically conjugate. Similarly, for the inclination and longitude of node (i, Ω), we take the canonical variables to be related to the corresponding Poincaré variables in the same way by defining

$$p_i \equiv 4\sqrt{1 - e^2}\sin^2(i/2) \qquad (6)$$

$$= i^2 + \mathcal{O}(e^2i^2, i^4), \qquad (7)$$

and taking its conjugate coordinate to be Ω. The equations of motion for $(p_e, \varpi, p_i, \Omega)$ are Hamilton's equations generated by the scaled H. These equations are exact as long as a is constant, which is the case for secular interactions.

An alternative formulation of the equations of motion will also prove useful. For an arbitrary Hamiltonian $H(p, q)$, one

---

[4] For example, even an angle involving both Mercury and Neptune appears to alternately librate and circulate (Sussman & Wisdom 1992); yet this is almost certainly a coincidence, and of no dynamical significance.

[5] The test particle's energy per unit mass E is given by Equation (A1) in Appendix A for the case of a single external planet.

may define the *complex canonical variable*[6] $Z \equiv \sqrt{p}e^{iq}$. As may be directly verified, the equation of motion for Z is then $dZ/dt = i\partial H/\partial Z^*$ where $H(Z, Z^*) = H(p, q)$. This complex equation of motion simultaneously encodes both of Hamilton's equations. Since our real canonical variables are $(p_e, \varpi, p_i, \Omega)$, we introduce the complex ones,

$$z \equiv \sqrt{p_e}e^{i\varpi} = \left[2\left(1 - \sqrt{1 - e^2}\right)\right]^{1/2}e^{i\varpi} \approx ee^{i\varpi} \quad (8)$$

$$\zeta \equiv \sqrt{p_i}e^{i\Omega} = 2(1 - e^2)^{1/4}\sin(i/2)e^{i\Omega} \approx ie^{i\Omega}, \quad (9)$$

which, to leading order in e and i, are the usual complex eccentricity and inclination.[7] The equations of motion for (z, ζ) are

$$dz/dt = i\partial H/\partial z^* \quad (10)$$

$$d\zeta/dt = i\partial H/\partial \zeta^*. \quad (11)$$

Throughout this paper, we freely switch between the set of real canonical variables (RCV) and the set of complex canonical variables (CCV),

$$\text{RCV: } (p_e, \varpi; p_i, \Omega) \quad (12)$$

$$\text{CCV: } (z, \zeta). \quad (13)$$

Although Hamilton's equations of motion for the RCV and CCV are exact, we take an approximate form for the Hamiltonian by expanding H to fourth order in e and i, and to leading order in the ratios of semimajor axes, keeping only secular terms. The relevant terms are listed in Table 1 in Appendix A. Throughout the bulk of this paper, we focus on the $c_1, c_2, c_4, c_{11}, c_{12}, c_{14}$, and c_{15} terms in that table. As we show in Section 4.5, the remaining terms in the table have a small effect in the parameter regime we focus on.

3. COPLANAR JUPITER AND SATURN

In this section, we consider the evolution of a test particle in the presence of two massive exterior planets ("Jupiter" and "Saturn"), with all bodies having zero inclinations. We assume that the massive planets' orbits are given by their linear Laplace–Lagrange solution, and evolve the test particle's equations of motion to leading nonlinear order (terms listed in Table 1 in Appendix A). This case was worked out by Sidlichovsky (1990). We describe it here in some detail because it sets the stage for the considerably more complicated case with non-zero inclinations (Section 4).

3.1. Jupiter Only

We shall solve the coplanar case with a sequence of increasingly complicated sub-cases. Consider first the linear secular evolution of a test particle perturbed by a circular Jupiter. From Table 1 in Appendix A, the particle's Hamiltonian is

$$H(z) = \gamma|z|^2, \quad (14)$$

---

[6] Our definition of the complex canonical variable differs from Ogilvie (2007) by a minus sign, and hence our Hamilton's equation also differs by a minus sign.

[7] The symbol e denotes both the eccentricity and the exponential (Euler's constant), and the symbol i denotes both the inclination and the imaginary unit. There should be no confusion because for the remainder of this paper we use as our dynamical variables either the RCV or CCV (Equations (12) and (13)) in lieu of e and i.

where the constant

$$\gamma \equiv \frac{3}{4}\frac{m_J}{M_\odot}\alpha^3\left(\frac{GM_\odot}{a^3}\right)^{1/2} \quad (15)$$

is the particle's linear free precession rate induced by Jupiter (Equation (A4)); m_J is Jupiter's mass, and α is the ratio of the particle's semimajor axis to Jupiter's. The equation of motion is $dz/dt = i\partial H/\partial z^* = i\gamma z$, with solution $z = \text{const.} \times e^{i\gamma t}$, an orbit with constant eccentricity that precesses at frequency γ.

For the second sub-case, we consider the test particle's linear evolution when Jupiter is assigned a constant eccentricity e_J and precession rate g_J, in which case the particle's Hamiltonian is

$$H(z) = \gamma(|z|^2 - (\epsilon_J e^{ig_J t}z^* + \text{c.c.})), \quad (16)$$

where $\epsilon_J = \frac{5}{4}\alpha e_J$; we drop nonlinear terms (i.e., the fourth-order terms in Table 1). Of course, if Jupiter were the only massive planet in the system, it would not precess ($g_J = 0$). But we consider a finite g_J in anticipation of what happens when Saturn is included.

Hamilton's equation is

$$\frac{1}{i\gamma}\frac{dz}{dt} = z - \epsilon_J e^{ig_J t}. \quad (17)$$

The solution is a sum of free and forced eccentricities:

$$z = \text{const.} \times e^{i\gamma t} + \frac{\epsilon_J}{\Delta}e^{ig_J t}, \quad (18)$$

where

$$\Delta \equiv \frac{\gamma - g_J}{\gamma}. \quad (19)$$

The forced solution diverges at secular resonance ($\gamma = g_J$). But this divergence is not physical. It is a consequence of dropping nonlinear terms.

With the nonlinear term included, the test particle's Hamiltonian becomes[8]

$$H(z) = \gamma\left(|z|^2 - \frac{|z|^4}{4} - (\epsilon_J e^{ig_J t}z^* + \text{c.c.})\right), \quad (20)$$

with equation of motion

$$\frac{1}{i\gamma}\frac{dz}{dt} = z\left(1 - \frac{|z|^2}{2}\right) - \epsilon_J e^{ig_J t}. \quad (21)$$

The nonlinear term reduces the free frequency from γ to $g \equiv \gamma(1 - |z|^2/2)$. We call g the nonlinear free frequency. It can also be introduced by first re-writing the Hamiltonian in terms of the real canonical variables p_e and ϖ and defining

$$g \equiv \frac{d\varpi}{dt}\Big|_{\epsilon_J=0} = \gamma(1 - p_e/2). \quad (22)$$

Because g is a function of eccentricity, secular resonance can occur if the eccentricity is chosen so that $g \approx g_J$. This is a *nonlinear secular resonance*.

Solutions of Equation (21) are shown in Figure 1 for two cases, $\gamma < g_J$ and $\gamma > g_J$. When $\gamma < g_J$ (left panel), the free frequency at small $|z|$ is less than the forcing frequency,

---

[8] We drop the c_3 term (see Table 1) because it merely alters the coefficient of the c_1 term by a small amount.
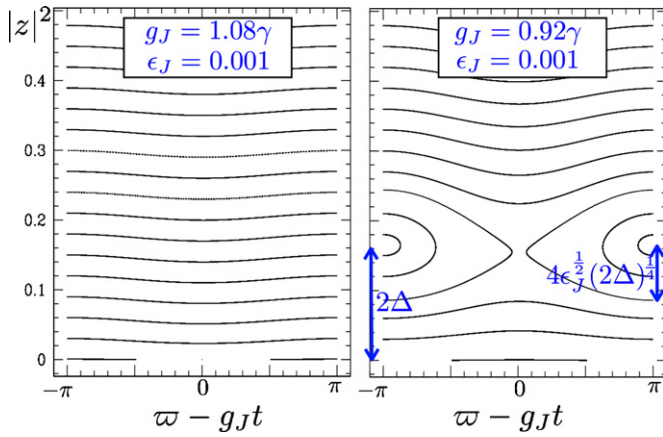
Figure 1. Trajectories of z for a test particle in the presence of a coplanar Jupiter: the curves are solutions of Hamiltonian (20), i.e., of Equation (21). Equivalently, they are level curves of Hamiltonian (24). The y-axis is approximately e^2, and the x-axis is the phase, modulo 2π. Nonlinear secular resonance occurs for the case $g_J < \gamma$.

(A color version of this figure is available in the online journal.)

Figure 2. Surfaces of section for the coplanar case (H given by Equation (27)): section taken at times when $e^{i(g_J - g_S)t} = 1$; the y-axis is approximately e^2; the x-axis is the phase, modulo 2π. In the left panel, the chosen parameters yield non-overlapping separatrices, and very little chaos is seen. In the right panel g_J is slightly larger, yielding overlapping separatrices and a sea of chaos.

(A color version of this figure is available in the online journal.)

$g|_{|z|\to 0} \approx \gamma < g_J$. Increasing $|z|$ decreases g, moving it further away from resonance. Hence, the particle's evolution is similar to the linear case, but with a smaller precession rate. When $\gamma > g_J$ (right panel), the precession rate at small values of $|z|$ exceeds g_J. Increasing $|z|$ causes the precession rate g to decrease until it nearly matches g_J. A resonant island appears, within which the angle $\varpi - g_J t$ librates. Increasing $|z|$ even further forces the free frequency to be less than g_J, taking the particle out of secular resonance.

To calculate the width of the nonlinear secular resonance, we first write Hamiltonian (20) in terms of the real canonical variables p_e and ϖ (Equation (8)), and then convert the Hamiltonian to a time-independent one by making the canonical transformation to (p_e, ϖ_-), where

$$\varpi_- \equiv \varpi - g_J t, \tag{23}$$

yielding the new Hamiltonian

$$\frac{1}{\gamma} H_-(p_e, \varpi_-) = -\frac{p_e^2}{4} + \Delta \cdot p_e - 2\epsilon_J \sqrt{p_e} \cos \varpi_- . \tag{24}$$

This form for a Hamiltonian has been called the "second fundamental model for resonance," the first being the pendulum. Its properties have been extensively cataloged because it also applies to first-order MMRs, for which the variables have a different interpretation (Henrard & Lemaitre 1983; Murray & Dermott 2000). The numerical integrations shown in Figure 1 trace level curves of this Hamiltonian. For parameters such that the resonant island both exists and is sufficiently far from $|z| = 0$ (as in the right panel of Figure 1), one may approximate the $\sqrt{p_e}$ multiplying the cosine term as constant, in which case the Hamiltonian is that of the pendulum. The center of the resonant island is where $(d/dt)\varpi_- = 0$, i.e., at

$$p_{e*} = 2\Delta \tag{25}$$

in this approximation; this is equivalent to $g = g_J$. The width of the island may be found by first completing the square in the "kinetic" part of the Hamiltonian, i.e., setting $-p_e^2/4 + \Delta \cdot p_e = -\frac{1}{4}(p_e - p_{e*})^2 + $ const. Since H_- is constant, the half-width of the island is given by

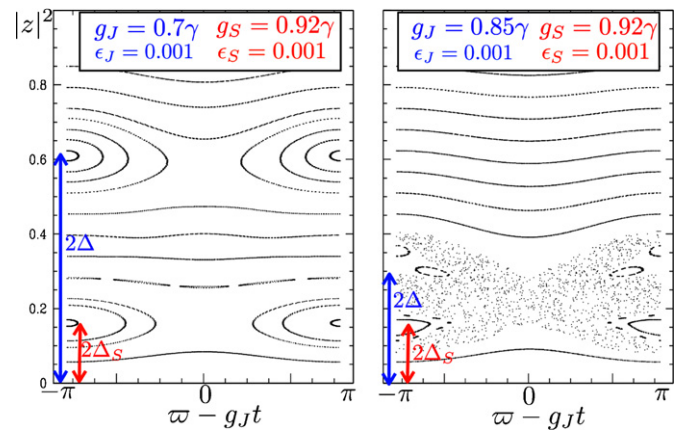$$\delta p_e = 4(\epsilon_J \sqrt{p_{e*}})^{1/2} = 4\epsilon_J^{1/2}(2\Delta)^{1/4}. \tag{26}$$

3.2. Jupiter and Saturn

We now add in the effect of a second coplanar planet, Saturn, and assume that Jupiter's and Saturn's evolution is described by their linear Laplace–Lagrange solution. In that solution, Jupiter and Saturn participate in two normal modes, which we call the Jupiter-dominated and Saturn-dominated modes. We denote the eigenfrequencies of these two modes g_J and g_S. Jupiter's (complex) eccentricity is a sum of two terms, one for each mode, and may be written as $e_{J,J}e^{ig_J t} + e_{J,S}e^{ig_S t}$. Similarly, Saturn's eccentricity has one term $\propto e^{ig_J t}$ and another $\propto e^{ig_S t}$. When all four terms are included, one arrives at the following form for the test particle's Hamiltonian (see Equation (20)):

$$H(z) = \gamma \left(|z|^2 - \frac{1}{4}|z|^4 - (\epsilon_J e^{ig_J t} z^* + \epsilon_S e^{ig_S t} z^* + \text{c.c.}) \right), \tag{27}$$

where γ is now re-interpreted to represent the test particle's precession rate due to both Jupiter and Saturn. The ϵ_J term is due to the Jupiter-dominated mode, with ϵ_J now a weighted sum of both Saturn's and Jupiter's eccentricities within that mode; similarly, ϵ_S is for the Saturn-dominated mode. For the purposes of this section, rather than solving the linear Laplace–Lagrange equations, we consider the parameters γ, ϵ_J, and ϵ_S to be adjustable constants.

Hamiltonian (27) would also describe the test particle's evolution if Jupiter's and Saturn's eccentricities and precession rates were enforced by hand to be e_J, e_S, g_J, and g_S, with $\epsilon_J = \frac{5}{4}\alpha_J e_J$ and $\epsilon_S = \frac{5}{4}\alpha_S e_S$. This re-interpretation of Hamiltonian (27), while unphysical as far as Jupiter and Saturn are concerned, can be helpful when considering the test particle's evolution under their influence.

Figure 2 shows results of numerical integrations of Hamilton's equation, plotted as surfaces of section. Changing to real canonical variables (Equation (8)), we see that Hamiltonian (27) has two cosine terms. The ϵ_J cosine term acting alone would yield a resonant island as long as $\gamma \gtrsim g_J$. Similarly, the ϵ_S term would yield an island if $\gamma \gtrsim g_S$. The location and width of the islands are quantified in Figure 1. When acting together, there may be two resonant islands. The left panel of Figure 2 shows a case when the parameters have been chosen to yield two non-overlapping islands. The result is mostly regular motion. The

right panel shows what happens when g_J is increased, so that Jupiter's Δ (Equation (19)) is reduced sufficiently that the islands overlap: the overlapping islands break up into a sea of chaos. This the well-known Chirikov resonance-overlap criterion for chaos (Chirikov 1979). From the widths and locations of the resonances as displayed in Figures 1 and 2, one deduces that the criterion for chaos is $2|g_J - g_S|/\gamma \lesssim 4\epsilon_J^{1/2}(2\Delta)^{1/4} + 4\epsilon_S^{1/2}(2\Delta_S)^{1/4}$, where $\Delta_S \equiv (\gamma - g_S)/\gamma$ (Sidlichovsky 1990).

It is instructive to consider the form of the surface of section shown in Figure 2 in more detail. This reasoning will also be helpful when we include a second degree of freedom (i.e., inclination) below. When the motion of z is regular, it can be written as a Fourier sum of terms with frequencies equal to the three fundamental frequencies of the problem (i.e., g, g_J, g_S, where g is the nonlinear free frequency), as well as integer combinations of these frequencies. But since only relative frequencies are physically meaningful, there are really only two fundamental frequencies, which may be chosen to be $g - g_J$ and $g_S - g_J$, i.e., relative to g_J. In other words, if z is regular, then $\tilde{z} \equiv z e^{-ig_J t}$ is a doubly periodic function with periods $2\pi/(g - g_J)$ and $2\pi/(g_S - g_J)$. In Figure 2, we choose to plot the amplitude versus phase of $\tilde{z}$ whenever the second period completed an integer number of cycles. As long as z is regular, the value of $\tilde{z}$ at those times is a singly periodic function, and hence appears in the plot as a connected curve. By contrast, when z is chaotic, it appears as scattered points.

4. AN INCLINED AND ECCENTRIC JUPITER

4.1. Equations of Motion

In this section, we consider the evolution of a test particle that comes under the influence of a single planet ("Jupiter") that has fixed values of eccentricity e_J, inclination i_J, apsidal precession rate $g_J \equiv \dot{\varpi}_J$, and nodal precession rate $s_J \equiv \dot{\Omega}_J$.

This is a model for the case when a particle comes under the influence of two planetary Laplace–Lagrange modes, one eccentric and one inclined (see the discussion of Hamiltonian (27)). As we show below, in the real solar system the main modes affecting Mercury are the Jupiter-dominated eccentricity mode and the Venus-dominated inclination mode, with the Venus-dominated eccentricity mode also playing a role. Therefore, for application to Mercury, e_J and g_J refer to the amplitude and frequency of the Jupiter-dominated eccentricity mode, while i_J and s_J refer to those of the *Venus*-dominated inclination mode. Nonetheless, for the purposes of the present section it is simplest to assume that Jupiter is the only planet, and that both of its precession rates are enforced by divine intervention.

We evolve the secular equations for the test particle to leading nonlinear order. Because the test particle now has two coupled degrees of freedom, its evolution is more complicated than before, and there are many more terms to include in its Hamiltonian. At first (Sections 4.1–4.4), we include only the following terms from Table 1:

$$\frac{1}{\gamma} H(z, \zeta) = |z|^2 - |\zeta|^2 - \frac{|z|^4 - |\zeta|^4}{4} - 2|z|^2|\zeta|^2$$
$$- (\epsilon_J e^{ig_J t} z^* - i_J e^{is_J t} \zeta^* + \text{c.c.}), \quad (28)$$

where $\epsilon_J = \frac{5}{4}\alpha e_J$. Note that the effect of Jupiter's eccentricity is diluted by a factor $\sim \alpha \ll 1$, whereas the effect of its inclination is undiluted by any such factor. In Section 4.5 we add in all the remaining terms from Table 1, and show that these additional terms have little effect in the parameter regime of interest. The

terms in Hamiltonian (28) that are second order in eccentricity or inclination (i.e., the first two terms, and the bracketed terms on the second line) are responsible for linear evolution. We keep only three nonlinear terms, $\propto |z|^4$, $|\zeta|^4$, and $|z|^2|\zeta|^2$. As we show in this subsection, these are responsible for nonlinear frequency shifts. And as we show in subsequent subsections, nonlinear frequency shifts are crucial for resonance overlap and chaos. Even though the frequency shifts might be small (second order in eccentricity and inclination), they can still be sufficient to shift the frequency into and out of secular resonance. One of the terms we drop is the Kozai resonance, i.e., the term $c_{26} \times (z^{*2}\zeta^2 + \text{c.c.})$ in Table 1. That term has little effect on Mercury's evolution, because its phase is rapidly varying, and hence the term nearly averages to zero for parameters similar to Mercury's (Section 4.5). By contrast, the frequency-changing terms can never average to zero.[9]

The equations of motion are

$$\frac{1}{i\gamma}\frac{dz}{dt} = z\left(1 - \frac{|z|^2}{2} - 2|\zeta|^2\right) - \epsilon_J e^{ig_J t} \quad (29)$$

$$\frac{1}{i\gamma}\frac{d\zeta}{dt} = \zeta\left(-1 + \frac{|\zeta|^2}{2} - 2|z|^2\right) + i_J e^{is_J t}. \quad (30)$$

Expressed in terms of the real canonical variables (Equation (12)), Hamiltonian (28) has two cosine terms (i.e., primary resonances), which become important when their arguments vary slowly. We first focus on the term $\propto \cos(\varpi - g_J t)$. Defining the free nonlinear apsidal frequency

$$g \equiv \frac{d\varpi}{dt}\Big|_{\epsilon_J = i_J = 0} = \gamma\left(1 - \frac{1}{2}p_e - 2p_i\right), \quad (31)$$

one would expect this resonance to be important near where

$$g(p_e, p_i) = g_J, \quad (32)$$

as may also be inferred by an inspection of Equation (29). To nonlinear order, g is a function of eccentricity and inclination, because of the nonlinear frequency-changing terms included in Hamiltonian (28). Hence by varying e and i, one can alter the free precession frequency and bring it into secular resonance. In the p_e–p_i plane, the resonance traces out a one-dimensional curve—or, in fact, a straight line to leading nonlinear order.

The second cosine term in the above Hamiltonian, $\cos(\Omega - s_J t)$ behaves similarly. Defining

$$s \equiv \frac{d\Omega}{dt}\Big|_{\epsilon_J = i_J = 0} = \gamma\left(-1 + \frac{1}{2}p_i - 2p_e\right), \quad (33)$$

one would expect it to be important near where

$$s(p_e, p_i) = s_J. \quad (34)$$

We shall make these considerations more precise in Section 4.3, where we also work out the resonant widths, and show that in addition to the primary resonances are a multitude of secondary resonances.

---

[9] We do drop some frequency-changing terms, specifically ones that are given by const. $\times |z|^2$, and const. $\times |\zeta|^2$, where the constant is $\mathcal{O}(e^2, i^2)$. Even though these terms do not average to zero, they merely shift the linear frequencies, and hence do not change the behavior qualitatively. In the absence of these terms, the linear apsidal and nodal frequencies are equal and opposite; we rectify this shortcoming in our κ-model Hamiltonian (Equation (53)).
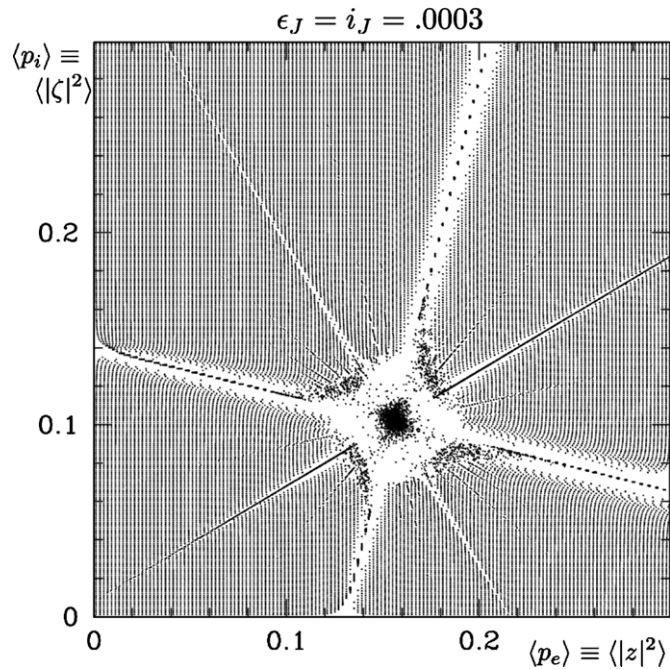
Figure 3. Map of the mean momenta (MMM) with low excitation: each point is the result of a single integration of Hamiltonian (28), with parameters $\epsilon_J = i_J = 0.0003$, $g_J = 0.72\gamma$, and $s_J = -1.26\gamma$. The y-axis is the time-averaged $|\zeta|^2 \approx i^2$, and the x-axis is the time-averaged $|z|^2 \approx e^2$. The initial conditions are on a uniform grid in $|\zeta|^2$ and $|z|^2$. Resonant bands are clearly visible, as is the zone of chaos where the bands overlap. Because of the averaging, chaotic orbits that explore the full extent of the overlap zone give rise to points at the center of the zone. Chaotic orbits that only partially explore the overlap zone (as is true of the real Mercury today; see below) give rise to chaotic points at the edges of the overlap zone. The orbits are very regular at small p_e, p_i; for comparison, the real Mercury currently has $\langle p_e \rangle \sim 0.05$ and $\langle p_i \rangle \sim 0.02$. Its motion would be regular under this weak forcing.

4.2. Simulations: Maps of the Mean Momenta

We run suites of simulations of Hamiltonian (28), i.e., of Equations (29) and (30). There are only five parameters, γ, g_J, s_J, ϵ_J, and i_J. The linear frequency γ sets the overall timescale, and can be scaled out. We choose $g_J/\gamma = 0.72$ and $s_J/\gamma = -1.26$, since these are close to the true values in the solar system for the Jupiter-dominated eccentricity mode (relative to Mercury's free precession frequency), and for the Venus-dominated inclination mode (see Section 5). We also set the excitation amplitudes equal to each other, $\epsilon_J = i_J$, and present sequences of simulations with various amplitudes. The true value in the solar system for the corresponding modes is, very roughly, $\epsilon_J \sim i_J \sim 0.01$ (see Section 5 for more precise values).

The dynamics is that of two nonlinearly coupled harmonic oscillators, each of which is also nonlinear and is forced periodically. We have attempted many different methods for visualizing the integration results, such as using catalogs of surfaces of section or Fourier transforms. However, most methods were too complicated, and obscured the underlying simplicity of the dynamics, i.e., they obscured that it is the overlapping of resonances that drive chaos. In the end, we invented a new method, the map of the mean momenta (MMM). This method has many advantages over the usual surfaces of section. It is somewhat similar to the frequency map analysis (FMA) of Laskar (1990; see below).

Figure 3 maps the results from around 50,000 numerical integrations at very small excitation, $\epsilon_J = i_J = 0.0003$, using
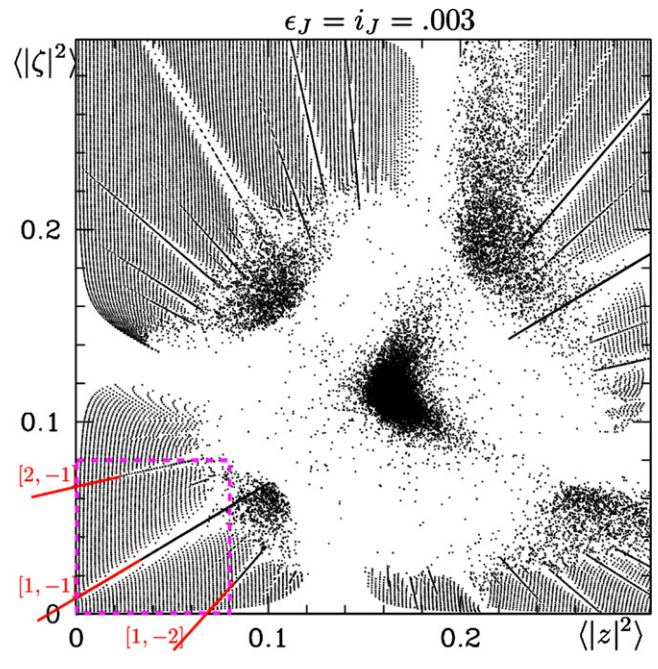

Figure 4. MMM with medium excitation: similar to Figure 3, but with ϵ_J and i_J increased by a factor of 10. The resonant bands are larger, as is the chaotic zone, which has encroached much closer to where the real Mercury lies, $\sim(0.05, 0.02)$. The dashed magenta square is for comparison with the axes of Figure 5.

(A color version of this figure is available in the online journal.)

the MMM. Each point in the plot is the time-averaged value of $|z|^2 \approx e^2$ and $|\zeta|^2 \approx i^2$ from a single simulation, averaged over a time span of $3 \times 10^4/\gamma$. Before taking the time average, we filter with a Hanning filter (Laskar 1993), which leads to a faster convergence of the averages (when they do converge). The simulations were initialized with values of $|z|^2$ and $|\zeta|^2$ that were equally spaced on a grid, with the spacing in $|z|^2$ twice that in $|\zeta|^2$; the initial phases were $\varpi = \pi/2$ and $\Omega = -\pi/2$.

Three kinds of motion are readily apparent in the MMM: (1) regular and non-resonant, (2) regular and resonant, and (3) chaotic. Most of the figure is covered with a regular grid of points that nearly traces the initial conditions. Here, the values of z and ζ remain regular and non-resonant throughout the simulation. A few resonant bands also appear in the figure, where the motion is also regular. Note that if the initial conditions span a resonant island, and if the motion remains regular, then the time-averaged momenta ($|z|^2$ and $|\zeta|^2$) exhibit a sharp discontinuity. Inside the island, they average to their values near the island center, whereas outside the island's separatrix they average to a value offset from the center by a finite amount, of order the resonance width (or, more accurately, around $1/4$ of the resonance full-width). This can be seen clearly in Figure 3, where the regular points at the center of the resonant bands represent librating particles. Also apparent in the figure are the chaotic trajectories. These show up as the cluster of irregular points near where resonant bands intersect. We have checked the Lyapunov exponent, as well as surfaces of section (see below), to verify that points that appear on the figure to be chaotic are truly chaotic.

Figures 4 and 5 show the MMM for simulations with the same parameters as in Figure 3, but with ϵ_J and i_J increased first to 0.003, and then to 0.01. With increasing forcing, the locations of the resonant bands do not change, but they get wider, and higher order resonances become visible. As a result, the zone of chaos
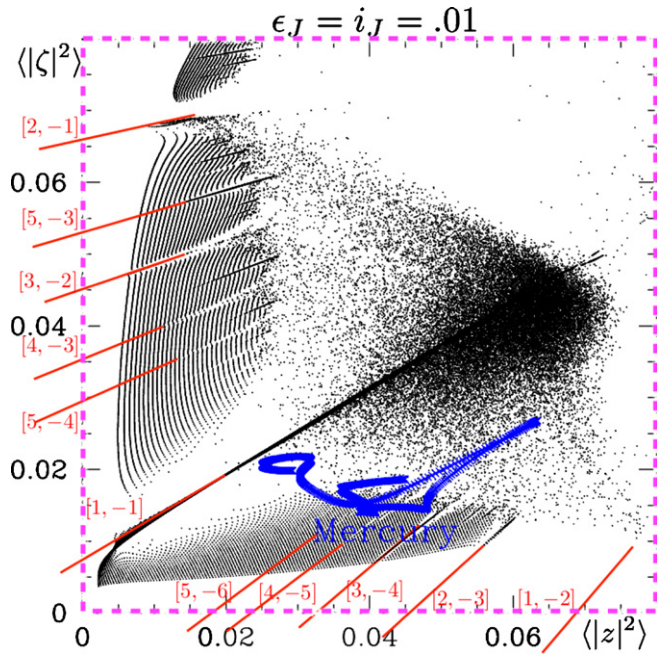
Figure 5. MMM with high excitation: similar to Figures 3 and 4, but with $\epsilon_J = i_J = 0.01$. Note the expanded scale. Many high-order resonances are visible. The zone of chaos approaches the origin, even though ϵ_J and i_J are $\ll 1$. The points labeled Mercury are the result of an N-body simulation of the full solar system; each point is Mercury's p_e and p_i averaged over a time span of 100 Myr, for the first 600 Myr of the simulation shown in Figure 16. Mercury's true orbit lies near the boundary between regular motion and chaos in the MMM of the simplified model. Parameters used for this MMM are within $\sim 20\%$ of the true solar system values. The true solar system is more chaotic due to other forcings.

(A color version of this figure is available in the online journal.)

where the bands overlap expands. Surprisingly, even with the seemingly modest forcing of $\epsilon_J = i_J = 0.01$—values that are comparable to those in the real solar system (see below)—the zone of chaos approaches very low values of e and i, and close to the values for the real Mercury.

Our method for displaying results, the MMM, is similar in philosophy to FMA (Laskar 1990). But whereas in FMA one plots the frequencies of the coordinates, here we plot the averages of the momenta. We have also performed the FMA (not shown); and when we convert from frequencies to momenta via the inverse of Equations (31) and (33), the resulting maps are almost identical to the MMM. For the purposes of the present paper, we prefer the MMM because its axes are approximately $\langle e^2 \rangle$ and $\langle i^2 \rangle$, which are simpler to interpret than the precession frequencies.

4.3. Theory: Resonance Locations and Widths, and Zone of Chaos

To develop understanding of the behavior seen in the MMMs, we first re-write Hamiltonian (28) in terms of the real canonical variables (Equations (8) and (9)), and then transform from (p_e, ϖ) to $(p_e, \varpi_- \equiv \varpi - g_J t)$ and from (p_i, Ω) to $(p_i, \Omega_- \equiv \Omega - s_J t)$, which transforms the Hamiltonian to a time-independent one:

$$\frac{1}{\gamma} H_-(p_e, \varpi_-; p_i, \Omega_-) = \Delta \cdot p_e + \Delta_s \cdot p_i - \frac{p_e^2 - p_i^2}{4}$$
$$-2 p_e p_i - 2\epsilon_J \sqrt{p_e} \cos \varpi_- + 2i_J \sqrt{p_i} \cos \Omega_-, \quad (35)$$
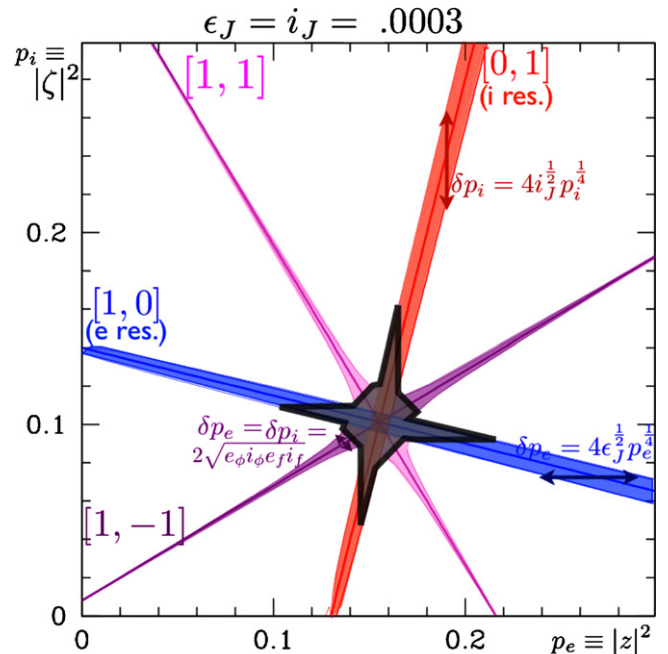

Figure 6. Four strongest resonances and their widths and region of overlap (analytic calculation): parameters are as in Figure 3. The center of each $[m, n]$ resonance is the line determined by $mg_- + ns_- = 0$ (Equations (41) and (42)). The arrows show the half-widths, with the orientation aligned with the direction of motion in the corresponding resonances. The central gray shaded region shows the region of overlapping separatrices, where chaotic motion is expected. Comparing with Figure 3 shows agreement between theory and simulation.

(A color version of this figure is available in the online journal.)

where Δ is the linear apsidal frequency mismatch (Equation (19)), and

$$\Delta_s \equiv \frac{-\gamma - s_J}{\gamma} \quad (36)$$

is the mismatch for the nodal frequencies. In the absence of the coupling term ($\propto p_e p_i$) both degrees of freedom would evolve independently according to the equations of the second fundamental model (Section 3.1).

In the following, we determine the location and width of the two primary resonances: the eccentricity resonance ([1,0]) and the inclination resonance ([0,1]). To do so, we ignore inclination forcing (setting $i_J = 0$) when studying the eccentricity resonance, and vice versa. The system is trivially integrable if either $i_J = 0$ or $\epsilon_J = 0$. In the former case, p_i is constant because the Hamiltonian does not depend on Ω_-. Therefore, H_- is equivalent to the coplanar Hamiltonian (Equation (24)), but with $\Delta \rightarrow \Delta - 2p_i$. This implies that the center of the eccentricity resonance is located at

$$p_{e*} = 2(\Delta - 2p_i) \quad \rightarrow [1, 0] \quad (37)$$

(Equation (25)) and the island has half-width at fixed p_i given by

$$\delta p_e = 4(\epsilon_J \sqrt{p_{e*}})^{1/2} \quad \rightarrow [1, 0] \quad (38)$$

(Equation (26)). The interior of this island is plotted in Figure 6 as a blue band, with parameter values as chosen for the simulations of Figure 3. Comparing the two figures shows that the above analytic expressions agree well with the result of the numerical integrations. Note that we plot the half-width in Figure 6 rather than the full-width because the average

momentum of an orbit that lies just outside of a separatrix is approximately half-way to the edge of the resonance. Figure 6 also shows the inclination resonance as a red band. Reasoning as before, its center and half-width (at fixed p_e) are

$$p_{i*} = -2(\Delta_s - 2p_e) \quad \to [0, 1] \tag{39}$$

$$\delta p_i = 4(i_J \sqrt{p_{i*}})^{1/2} \quad \to [0, 1] \tag{40}$$

In addition to these primary resonances are an infinite number of secondary ones. Far from resonances, we may ignore the cosine forcing terms since they tend to average to zero. The test particle's free precession frequencies relative to Jupiter are then

$$g_- \equiv g - g_J \equiv \frac{d\varpi_-}{dt}\Big|_{\epsilon_J = i_J = 0} = \gamma \left(\Delta - \frac{1}{2}p_e - 2p_i\right) \tag{41}$$

$$s_- \equiv s - s_J \equiv \frac{d\Omega_-}{dt}\Big|_{\epsilon_J = i_J = 0} = \gamma \left(\Delta_s + \frac{1}{2}p_i - 2p_e\right). \tag{42}$$

Resonances are important near where

$$mg_- + ns_- = 0 \quad \to [m, n] \tag{43}$$

for integer pair $[m, n]$. Therefore, the center of each $[m, n]$ resonance traces out a line in the p_e–p_i plane. For instance, the centers of the $[1, 0]$ and $[0, 1]$ resonances are as worked out above, and the center of the $[1, -1]$ resonance is the line

$$p_{e*} - \frac{5}{3}p_{i*} = \frac{2}{3}(\Delta_s - \Delta) \quad \to [1, -1] \tag{44}$$

which passes close to the origin for our choices of Δ and Δ_s. In general, the slope of a resonance line in the p_e–p_i plane is $(4n + m)/(n - 4m)$, and all $[m, n]$ resonant lines intersect at the point in the p_e–p_i plane where $g_- = s_- = 0$, i.e., at (p_{e**}, p_{i**}), where

$$p_{e**} \equiv \frac{2}{17}(\Delta + 4\Delta_s), \quad p_{i**} \equiv \frac{2}{17}(4\Delta - \Delta_s). \tag{45}$$

Although there are an infinite number of secondary resonances, most are very weak, i.e., their widths are small. The most prominent resonances in Figure 3 are the primary resonances $[1, 0]$ and $[0, 1]$, whose widths have been worked out above. Next most prominent are the $[1, -1]$ and $[1, 1]$ resonances, whose widths may be understood qualitatively as follows (see Appendix B for a quantitative calculation). The linear solution for z is a sum of two terms, the free and forced complex eccentricities (Equation (18)). Similarly, to linear order ζ is a sum of free and forced complex inclinations. Therefore, to leading nonlinear order, the coupling term in the Hamiltonian ($p_e p_i = |z|^2|\zeta|^2$) can be written as a sum of terms, one of which has the form $z_\phi z_f^* \zeta_\phi^* \zeta_f \approx e_\phi e_f i_\phi i_f \exp(i(g_\phi - g_J - s_\phi + s_J))$, where the subscript f denotes forced and ϕ denotes free. This term has frequency corresponding to the $[1, -1]$ resonance; hence the width of this resonance is $\sim \sqrt{|e_\phi i_\phi e_f i_f|}$. The $[1, 1]$ resonance behaves similarly.

The quantitative calculation in Appendix B shows that, in agreement with the above estimate, the $[1, \pm 1]$ resonances have half-widths (Equation (B10)),

$$\delta p_e = \delta p_i = 2\sqrt{e_\phi i_\phi e_f i_f} \quad \to [1, \pm 1] \tag{46}$$

after defining the free values as

$$e_\phi \equiv \sqrt{p_{e*}} \quad i_\phi \equiv \sqrt{p_{i*}} \tag{47}$$

and the forced values as

$$e_f \equiv \frac{\epsilon_J}{|\Delta - p_{e*}/2 - 2p_{i*}|}, \quad i_f \equiv \frac{i_J}{|\Delta_s + p_{i*}/2 - 2p_{e*}|}, \tag{48}$$

where the asterisk denotes values at resonant center, and we neglect here the small difference between the lower-case momenta and the upper-case ones used in Appendix B. As in linear theory, the forced eccentricity scales as the inverse of the frequency detuning (Equation (18)), although now it is the nonlinear frequency detuning g_- (Equation (41)) that is relevant; similarly, i_f is inversely proportional to s_-.

The above widths for the $[1, \pm 1]$ resonances are shown in Figure 6. The $[1, -1]$ resonance leaves $p_e + p_i$ nearly constant, which produces trajectories in the p_e–p_i plane that have a slope of -1 (see Appendix B; we again neglect the small difference between lower- and upper-case momenta). Therefore, for each value of (p_{e*}, p_{i*}) at resonant center (i.e., where $g_- - s_- = 0$), the upper and lower envelopes of the resonant band are at $(p_{e*} \pm \delta p_e, p_{i*} \mp \delta p_e)$. Comparing with Figure 3 shows excellent agreement between theory and simulations.

The central gray region in Figure 6 is the overlap zone of the above four resonances. Its shape is peculiar because each resonance induces a trajectory with a particular orientation in the p_e–p_i plane, and we take the overlapping region to be wherever one resonance can induce motion into a second resonance. The zone of chaos seen in Figure 3 is about twice as large as that predicted in Figure 6. This difference is not surprising because we plot the half-widths in Figure 6, whereas one might expect that chaos would begin where the full-widths overlap. Furthermore, we shall show that higher order combinations of the four strongest resonances also play a role in the chaos. Nonetheless, the gray zone provides a reasonable estimate of the zone of chaos.

Figure 7 plots the widths calculated above when the forcing ϵ_J and i_J take on the values of Figure 4. The latter figure is overlayed on Figure 7, showing excellent agreement between theory and simulation in the zones of regular motion. And, as before, the zone of chaos from the simulations is around twice as large as the region where the resonant half-widths overlap.

The dimensions of the zone of chaos may be estimated analytically. These estimates will be used later to infer the threshold for chaos in the solar system. In Figures 6 and 7, the horizontal and vertical spikes of the chaotic zone are due to the $[1, 0]$ and $[0, 1]$ resonances; the extent of the spikes is simply estimated by the resonance width near zone center,

$$\delta p_{e*} \simeq 4\epsilon_J^{1/2} p_{e**}^{1/4} \quad \to [1, 0] \times [0, 1] \tag{49}$$

$$\delta p_{i*} \simeq 4i_J^{1/2} p_{i**}^{1/4} \quad \to [1, 0] \times [0, 1], \tag{50}$$

for the horizontal and vertical spikes, respectively, where p_{e**} and p_{i**} are the coordinates of the zone center (Equation (45)). The extent of the zone caused by the overlap between the $[1, -1]$ separatrix with the $[1, 0]$ resonance can be estimated similarly as

$$\delta p_{e*} = k(\epsilon_J i_J)^{1/4}(p_{e**} p_{i**})^{1/8} \quad \to [1, -1] \times [1, 0], \tag{51}$$
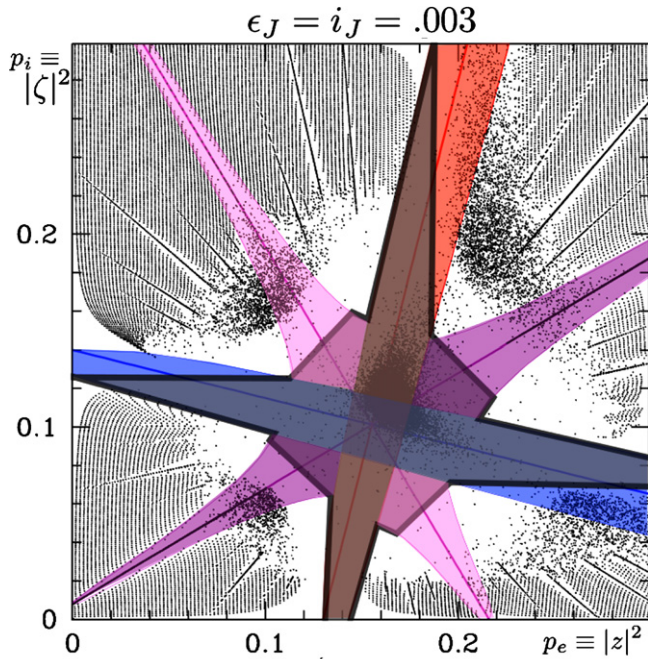
Figure 7. Resonant bands and their region of overlap, overlayed on MMM: the colored bands show the theoretically predicted widths of the four strongest resonances, as in Figure 6. The gray region is the zone of resonance overlap for these four resonances. Also shown is the MMM of Figure 4, showing excellent agreement between the theoretical and numerical resonant widths, and satisfactory agreement for the zone of chaos.

(A color version of this figure is available in the online journal.)

Figure 8. Surfaces of section from integrations of Hamiltonian in Equation (28): the parameters are $g_J = 0.72\gamma$ and $s_J = -1.26\gamma$, and other parameters as shown. The three upper left panels show surfaces of section with increasing forcing, showing how the $[1, -1]$ resonance gets wider, and its separatrix breaks up into a sea of chaos. The (blue) chaotic trajectory in the upper right panel has, very roughly, parameters comparable to the real Mercury. The lower right panel shows a "double section" of this blue chaotic trajectory. In the double section, it traces out a branch of a hyperbola in momentum space. Comparing with Figure 5 shows where this trajectory lies relative to the MMM.

(A color version of this figure is available in the online journal.)

where $k \simeq (20/17)\sqrt{3/4}$ is an order-unity constant.[10] The expression for δp_{i*} is the same, as are the extents due to the overlap between either $[1, 1]$ or $[1, -1]$ with either $[1, 0]$ or $[0, 1]$, albeit all have different order-unity values for k.

It might appear surprising that the width of the chaotic zone due to the overlap of a primary resonance with a secondary one (Equation (51)) is not much smaller than that due to the overlap between two primary resonances, despite the fact that the width of a primary resonance is first order in eccentricity or inclination (Equations (38) and (40)), whereas the width of the $[1, -1]$ is second order (Equation (46)). The reason for this is that the $[1, -1]$ resonance is enhanced by the denominator that appears in the forced eccentricity and inclination (Equation (48)). Because of this, the $[1, \pm1]$ resonances play an important role in setting the extent of the chaotic zone. For example, with the forcing frequencies that we have chosen for the MMMs (which are comparable to those for Mercury), the $[1, -1]$ resonance overlap allows the region of chaos to encroach upon the origin ($e \sim 0$, $i \sim 0$) at lower values of the forcing than would have been expected based solely on the overlap of the $[1, 0]$ and $[0, 1]$ resonances. In Section 5, we shall show explicitly that the $[1, -1]$ resonance plays a dominant role in driving chaos for the real Mercury.

One could also proceed to calculate the widths of higher order resonances. Extending our reasoning from above, the width of

an $[m, n]$ resonance in the p_e–p_i plane should scale as

$$\delta p \sim O(\epsilon^{|m|} i^{|n|}) \quad \rightarrow [m, n], \qquad (52)$$

where ϵ is comparable to the typical eccentricity, and i is comparable to the typical inclination. However, as we have seen for the $[1, \pm1]$ resonances, near-resonant denominators can make the widths significantly larger than this naive estimate.

4.4. Surfaces of Section and Libration of Resonant Angles

Here, instead of the global map (MMM), we investigate a few particular trajectories in detail to demonstrate the chaotic behavior, using both surfaces of section and resonant angles.

Figure 8 shows a number of surfaces of section from Hamiltonian (28). Since the transformed Hamiltonian (Equation (35)) is time independent with two degrees of freedom, we follow the usual practice of taking a section whenever the phase of one of the degrees of freedom (here, ϖ_-) executes an integer number of cycles. At those times, we plot the amplitude versus phase of the second degree of freedom (p_i versus Ω_-). We may also understand this form for the surface of section as follows. Hamiltonian (28) has four fundamental frequencies, g_J, s_J, g, and s where g and s are the nonlinear free frequencies of z and ζ, respectively. Only relative frequencies are physically meaningful, and there are three of these, which we may choose to be $g-g_J$, $s-s_J$, and $g-s$. But Hamiltonian (28) does not depend on $\varpi-\Omega$, and hence the frequency $g-s$ does not enter.[11] Therefore,

---

[10] Focusing on the region to the lower left of the zone center in Figure 6 or 7, the center of the $[1, -1]$ resonance is displaced from zone center in the p_e–p_i plane by the vector $-(1, 3/5)x$ (Equation (44)); we take $x > 0$. The orbit given by the half-width of the $[1, -1]$ is therefore displaced from zone center by the vector $-(1, 3/5)x + (-1, 1)\delta p_e$, where $\delta p_e \simeq 2(p_{e**}p_{i**})^{1/4}\sqrt{\epsilon_J i_J}/(17x/10)$ (Equation (46)). Equating that vector to the displacement of the center of the $[1, 0]$ resonance, i.e., to $(-1, 1/4)x'$ (Equation (37)) yields Equation (51).
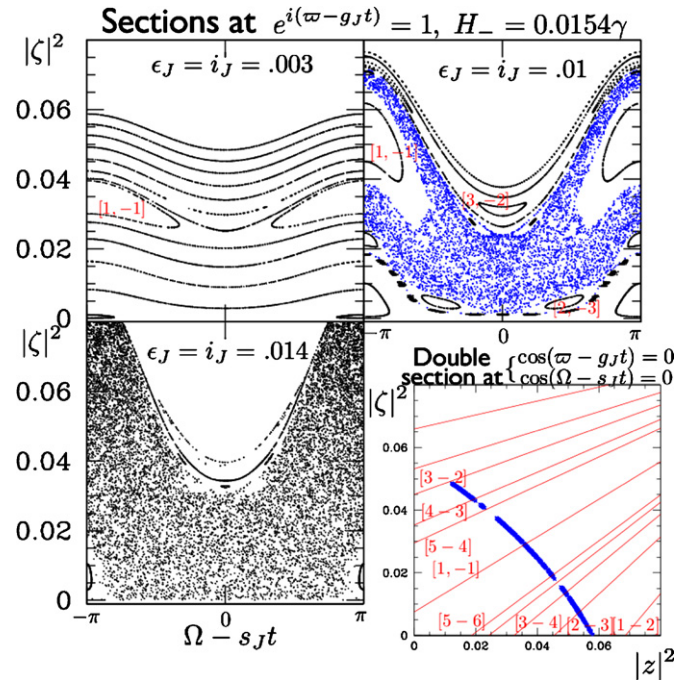
[11] The full fourth-order Hamiltonian does depend on $\varpi - \Omega$ because of the Kozai term; see Table 1. Therefore, one cannot take surfaces of section of the full fourth-order Hamiltonian, but one can still plot its MMM (Figure 10).
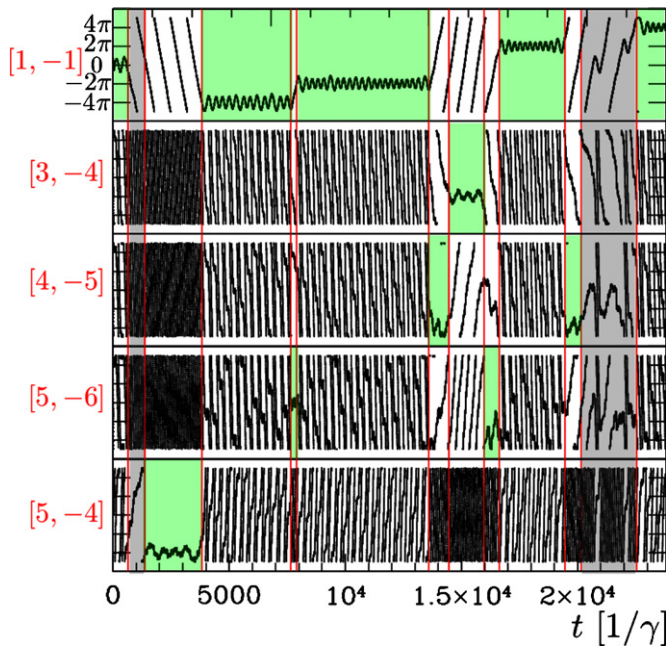
Figure 9. Chain of librating angles: each panel shows resonant angles $m\varpi_- + n\Omega_-$ (modulo 10π) with various values of $[m, n]$, for the blue chaotic trajectory of Figure 8. The green shaded zones show librating angles. Different resonant angle combinations librate in turn. The first gray strip is when $[6, -5]$ librates (not shown). The second gray shaded strip shows a time when the $[1, -1]$ and $[4, -5]$ alternately librate in rapid succession, and no other angles are clearly librating.

(A color version of this figure is available in the online journal.)

Figure 10. MMM with full fourth-order Hamiltonian: similar to Figure 5, but integrations have been performed with the full Hamiltonian (Table 1 in Appendix A), rather than the truncated Hamiltonian (28). From the fact that the two figures are broadly similar, one can infer that the terms dropped from Hamiltonian (28) are of small importance in the regime of interest.

(A color version of this figure is available in the online journal.)

there are two remaining fundamental frequencies, as for the coplanar case (Section 3.2). And, as described there, to examine the characteristics of the motion, one may take a section whenever the phase corresponding to one of the fundamental frequencies (here, the phase $\varpi - g_J t$ which corresponds to $g - g_J$) executes an integer number of cycles.

The three upper left panels of Figure 8 show surfaces of section with values of g_J and s_J as before, and with various values of ϵ_J and i_J. All the surfaces of section shown have the same (constant) value of energy, $H_- = 0.0154\gamma$. To map out phase space would require many different energy values, but for that purpose the MMM is more useful. The lower right panel in Figure 8 shows a "double section" of the (blue) chaotic trajectory in the upper right panel, i.e., it shows the two momenta wherever both ϖ_- and Ω_- have executed a half-integer number of cycles. At these times, the cosine terms in the transformed Hamiltonian vanish, and all trajectories with a fixed energy H_- fall along the same branch of a hyperbola. Of course, double sections from all the trajectories shown in Figure 8 would lie along the same branch of the hyperbola because they all have the same value of H_-.

The surface of section in the upper left panel has a relatively low ϵ_J and i_J, and the motion is mostly regular for the value of energy chosen. This can also be seen in the MMM (Figure 4) near the relevant hyperbola branch. The $[1, -1]$ resonance is clearly evident in the top left panel of Figure 8. Its half-width is $\delta p_i = 0.0055$, as compared to the prediction of 0.006 from Equation (46). The upper right panel of Figure 8 shows the case with higher forcing. With this higher forcing, the $[1, -1]$ resonance is wider, and the region near its separatrix has broken up into a wide zone of chaos. Some of the higher order resonances are visible in this section. In the lower left panel, the
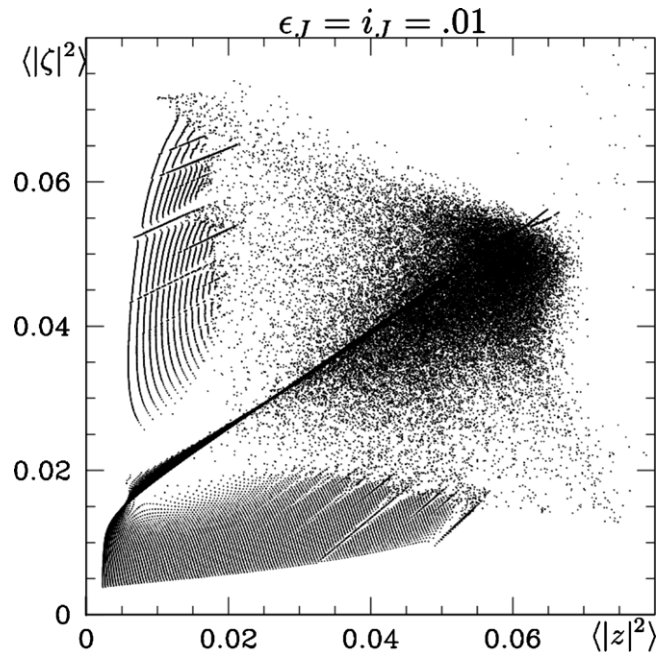
forcing has been raised further. Even though ϵ_J and i_J are still relatively small compared to unity, the zone of chaos is vast.

The blue chaotic trajectory in the upper right panel of Figure 8 behaves qualitatively like the real Mercury, and the parameters are also similar (Section 5). Therefore, we investigate it in more detail. From its surface of section, we see that the separatrix of the $[1, -1]$ resonance is largely responsible for driving the chaos for this orbit, together with overlapping higher order resonances. This orbit remains bounded by the $[3, -2]$ and $[2, -3]$ resonances, and hence can never come under the direct influence of the primary resonances ($[1, 0]$ and $[0, 1]$). The bound on the chaotic zone is a consequence of using a truncated Hamiltonian that can be written in a time-independent form with two degrees of freedom. For the full Hamiltonian, one might expect that diffusion could act on long timescales (Arnold diffusion), ultimately allowing the trajectory to cross into other regions of phase space.

Figure 9 shows explicitly that the chaos is due to the overlapping of high-order resonances. The resonant angles $m\varpi_- + n\Omega_-$ are plotted for various values of $[m, n]$. Different resonant angles librate in turn, showing that this orbit first comes under the influence of the $[1, -1]$, then the $[5, -4]$, then the $[1, -1]$, etc.

4.5. Full Fourth-order Hamiltonian

Thus far we have focused on the truncated fourth-order Hamiltonian (Equation (28)). Figure 10 shows the MMM of the full fourth-order Hamiltonian, expanded to leading order in α (i.e., including all terms in Table 1 in Appendix A). From the similarity of Figure 10 to the truncated integrations of Figure 5, we conclude that the terms dropped in the truncated Hamiltonian have little effect on the dynamics, particularly in the region of small e and i (lower left corner of the MMM).

The dropped terms have little effect because the only new fundamental relative frequency they introduce is $g-s$ (see first paragraph of Section 4.4). This "Kozai frequency" differs significantly from zero in the domain of Figure 10, and hence it can only combine with the other two relative frequencies ($g - g_J$ and $s - s_J$) to give resonances at high order. To be quantitative, the resonant line of the Kozai frequency is at $g - s \approx \gamma(2 + (3/2)p_e - (5/2)p_i) \approx 0$, i.e., it traces the line $p_i \approx (4 + 3p_e)/5$. Hence, the Kozai resonance is at much larger p_i than shown in Figure 10.

Aside from the Kozai term (c_{26} in Table 1), all other terms dropped from the truncated Hamiltonian depend on Jupiter's eccentricity or inclination. These do not introduce new forcing frequencies because Jupiter's frequencies already appear in the test particle's orbit at linear order—in its forced e and i. While the dropped terms do change the amplitudes of the forcing terms, the change is small as long as Jupiter's e and i is smaller than the e and i it linearly forces in the test particle, as is true of Figure 10.

We suspect that the terms dropped from Hamiltonian (28) are quite often of secondary importance. This is largely true for the real Mercury (Section 5). And we suspect that it is true more generally because if secular interactions between two planets are strong, then the forced e's and i's will typically (though not always) be larger than the forcing ones. Nonetheless, the dropped terms can be important in certain circumstances; for example, the Kozai term will play a role if a planet has a high inclination, and MMRs will be important for planets whose orbital periods are near integer ratios.

4.6. Fourier Transforms

In Section 5 we shall make the connection to the real Mercury. For that purpose, it will prove instructive to examine trajectories in Fourier space.

For a more exact comparison to Mercury, we consider here the Hamiltonian

$$\frac{1}{\gamma}H(z, \zeta) = |z|^2 - \hat{\gamma}|\zeta|^2 - \frac{|z|^4 - |\zeta|^4}{4} - 2|z|^2|\zeta|^2$$
$$- (\epsilon_J e^{ig_J t} z^* - i_V e^{is_V t}\zeta^* + \text{c.c.}), \quad (53)$$

which differs from Hamiltonian (28) by the inclusion of a constant $\hat{\gamma}$ to allow the linear apsidal and nodal precession rates to differ from each other (see footnote 9). Note that we also change notation so that i_V and s_V are the amplitude and precession rate of the Venus mode. We focus on a one-dimensional family of systems parameterized by κ, which scales all eccentricities and inclinations. More precisely, in this "κ-model" we choose the parameters $\epsilon_J = i_V = 0.008\kappa$, and initial conditions $|z| = 0.16\kappa, |\zeta| = 0.07\kappa, \varpi = \Omega = \pi/2$. The remaining parameters are $\hat{\gamma} = 0.9, g_J = 0.72\gamma, s_V = -1.14\gamma$. With these parameters, the center of the [1, 0] resonance is at $p_{e*} = 2(0.28 - 2p_i)$, as before; and the center of the [0, 1] resonance is at $p_{i*} = -2(0.24 - 2p_e)$, whereas before the constant was 0.26 rather than 0.24. This difference is of little consequence. For displaying the results of the integration, we shall choose $\gamma = 5\rlap{.}''87 \text{ yr}^{-1}$. Our rationale for choosing these particular numerical values will be explained in Section 5.

Figure 11 shows the Fourier transforms of z and ζ for the κ-model trajectory that has $\kappa = 0.75$. We normalize the Fourier transform of z as

$$\hat{z}(\omega) \equiv \frac{1}{T} \int_0^T z(t)e^{-i\omega t}dt, \quad (54)$$
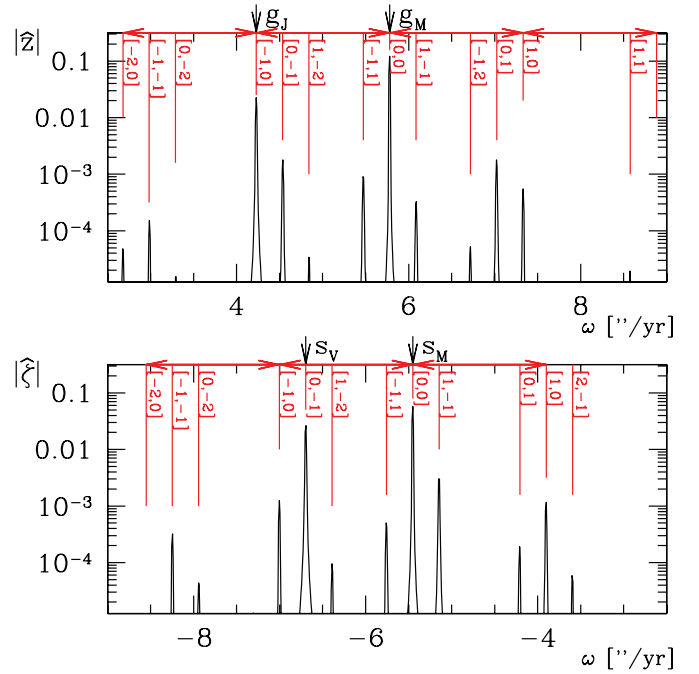

Figure 11. Fourier transforms of the test particle's z and ζ for the κ-model with $\kappa = 0.75$ (Equation (53)): with this relatively small value of κ, the trajectory is quasiperiodic, as indicated by narrow spikes in the Fourier transform. The free and forced z are peaks at g_M and g_J in the top panel. The other peaks are due to nonlinear couplings, and are at frequencies $g_M + m(g_M - g_J) + n(s_M - s_V)$, labeled $[m, n]$. The horizontal red arrows in both panels denote frequencies spaced by $g_M - g_J$. The bottom panel shows that $|\hat{\zeta}|$ is similar, with free and forced ζ at frequencies s_M and s_V, and nonlinearly generated peaks at $s_M + m(g_M - g_J) + n(s_M - s_V)$.

(A color version of this figure is available in the online journal.)

and similarly for ζ, where T is the duration of the Fourier transform, which we choose in the present subsection to be $T = 4400/\gamma$. With this normalization, if z has constant amplitude and frequency, i.e., if $z(t) = k_0 e^{i\omega_0 t}$, then $|\hat{z}| = |k_0|$ at frequencies close to ω_0.

The trajectory used for Figure 11 is quasiperiodic—the peaks in the Fourier transform are simply spikes, whose widths become narrower for larger T. We call the two largest peaks in the top panel the forced and free z. The forced z is at frequency $g_J = 0.72\gamma = 4\rlap{.}''23 \text{ yr}^{-1}$. The free z is at frequency $g_M = 5\rlap{.}''78 \text{ yr}^{-1}$. Because of nonlinearities, the free frequency (g_M) differs from the linear free frequency (γ) by a small but non-negligible amount. Similarly, in the bottom panel the largest two peaks are the forced ζ at frequency $s_V = -1.14\gamma$, and the free ζ at frequency $s_M = -5\rlap{.}''45 \text{ yr}^{-1}$, which differs from the linear free frequency $-\hat{\gamma}\gamma$.

In addition to the free and forced z and ζ, there are a multitude of peaks in Figure 11 that are generated by nonlinear couplings. The peaks in $\hat{z}$ all fall at frequencies $g_M + m(g_M - g_J) + n(s_M - s_V)$ for integers m, n. Roughly speaking, the peak amplitudes become smaller for larger values of $|m|$ and $|n|$. These amplitudes can be calculated perturbatively, as is outlined in the following. As before, we define the free and forced components as $(z_\phi, z_f, \zeta_\phi, \zeta_f)$, which have phases that rotate with frequencies (g_M, g_J, s_M, s_V), respectively. To leading nonlinear order, the nonlinear terms in the equation for dz/dt are proportional to $z|z|^2 = (z_\phi + z_f)|z_\phi + z_f|^2$ and $z|\zeta|^2 = (z_\phi + z_f)|\zeta_\phi + \zeta_f|^2$ (Equation (29)). These generate six new frequencies in z: $(2g_M - g_J), (2g_J - g_M), g_M \pm$
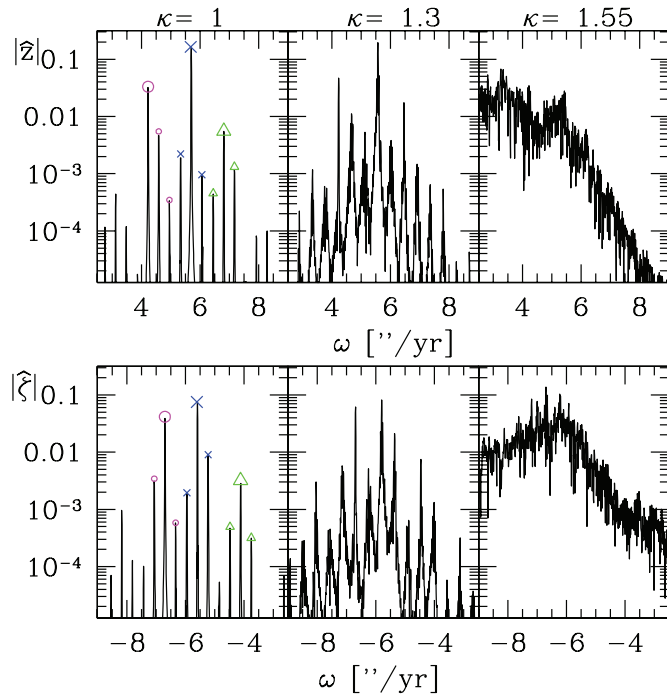
Figure 12. Fourier transforms of results from three κ-model simulations, with $\kappa = 1, 1.3, 1.55$. At $\kappa = 1$ (leftmost panels), the motion is still largely quasiperiodic. The amplitudes of the nonlinearly generated peaks have risen significantly relative to Figure 11 ($\kappa = 0.75$), even though the forced and free z and ζ have changed by a modest amount. At $\kappa = 1.3$, there is weak chaos—the peaks have widened, and neighboring peaks overlap. At $\kappa = 1.55$, the trajectory is highly chaotic.

(A color version of this figure is available in the online journal.)

Figure 13. Fourier transform of Mercury's z and ζ (roughly, its complex eccentricity and inclination) in a SWIFT N-body simulation, $\kappa_{\rm nbody} = 0.75$: the initial e's and i's of all planets were pre-multiplied by the factor $\kappa_{\rm nbody}$. The peaks here are broadly similar to those of the κ-model. As in Figure 11, they are marked by vertical lines that are displaced from g_M and s_M by $m(g_M - g_J) + n(s_M - s_V)$. The agreement between the two figures shows that the κ-model captures much of the physics of the real Mercury. Nonetheless, there are a number of differences. See the text.

(A color version of this figure is available in the online journal.)

$(s_M - s_V)$, $g_J \pm (s_M - s_V)$. Each frequency-generating term acts as a linear forcing on z. Together with the free and forced z, these account for eight of the peaks marked in the top panel of Figure 11; specifically, they account for the two highest peaks in each of the four leftmost triplets. The other peaks are accounted for by higher order nonlinear terms. One of these other peaks—the one labeled $[1, -1]$—is quite large, even though one might naively have expected that it would be smaller because it enters at a higher nonlinear order. The reason for this is that its forcing frequency differs from g_M by $\sigma \equiv (g_M - g_J) - (s_M - s_V)$ which is quite small. Hence, this near resonance amplifies the peak by $g_M/\sigma \sim 20$. We note parenthetically that the width of the $[1, -1]$ resonance (as described in Section 4.3) is directly related to the amplitudes of the three peaks at g_M and $g_M \pm \sigma$. The Fourier transform of ζ behaves similarly to that of z, with the frequency peaks at $s_M + m(g_M - g_J) + n(s_M - s_V)$.

Figure 12 shows the Fourier transforms for the κ-model at higher values of κ. The leftmost panels show the case $\kappa = 1$. The motion is largely quasiperiodic, but the nonlinearly generated peaks have increased significantly relative to the $\kappa = 0.75$ case. At $\kappa = 1.3$ the motion is chaotic, and at $\kappa = 1.55$ it is highly chaotic.

5. MERCURY

We integrate the eight solar system planets with the SWIFT symplectic integrator (Levison & Duncan 1994), supplemented with a routine for Mercury's relativistic precession (see Wu & Lithwick 2011, for code details). We initialize the planets with their current orbits and use their actual masses. The integration time step is 8 days.
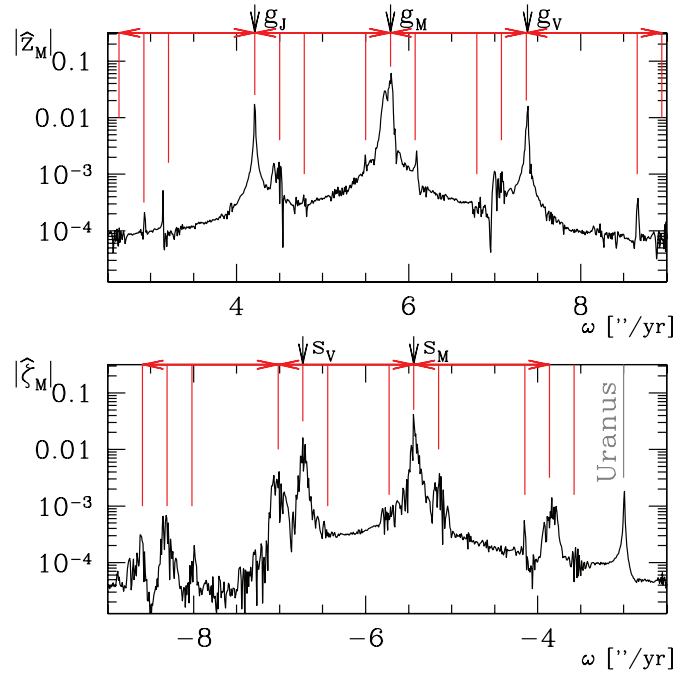
One might suspect that Mercury's orbital evolution is more complicated than our toy model for a variety of reasons: its e and i are not too small, and hence the fourth-order expansion is approximate; it is not massless, and hence backreacts onto the other planets (especially Venus); there are seven other planets that do not have constant orbital elements and frequencies but participate in the overall chaos of the solar system; and Mercury can be affected by resonant terms. Despite these complications, we show that the chaotic behavior of Mercury is qualitatively similar to the Hamiltonian model. This is perhaps not too surprising, since nonlinear dynamics are largely driven by resonances and their overlap. Hence, as long as a model roughly captures the locations and widths of the principal resonances, it should produce qualitatively correct behavior.

5.1. Fourier Transforms

To compare with the κ-model, we first consider cleaner cases by pre-multiplying the current eccentricities and inclinations of all planets by the reduction factor $\kappa_{\rm nbody}$. Figure 13 shows Mercury's Fourier transforms in a $\kappa_{\rm nbody} = 0.75$ integration lasting $T = 150$ Myr. Comparing this with the κ-model at $\kappa = 0.75$ (Figure 11) shows broad agreement. In truth, the parameters for the κ-model were chosen to match the free and forced z and ζ seen in Figure 13, i.e., the frequencies and heights of the four peaks marked g_J, g_M, s_V, s_M. Since there were eight quantities to match, we could do this by adjusting eight parameters in the κ-model: $\gamma, \hat{\gamma}, g_J, s_V, \epsilon_J, i_V$, as well as the initial values of $|z|$ and $|\zeta|$. Therefore, it is not significant that the forced and free peaks in the two figures agree. What is significant is that the other peaks that are generated by nonlinear
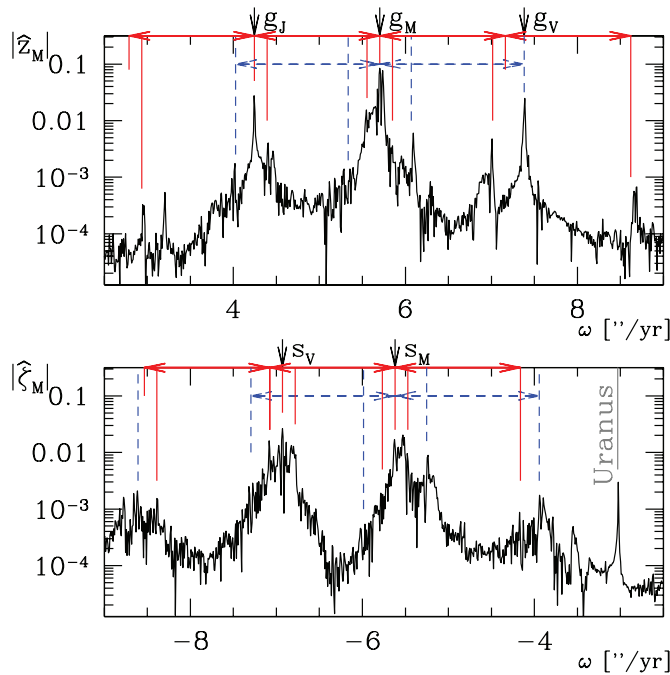
Figure 14. Same as Figure 13, but with $\kappa_{\rm nbody}$ increased from 0.75 to 0.95. The motion is more chaotic here, and Venus's eccentricity forcing at frequency g_V is distinct. The blue dashed arrows denote frequency spacings of $g_V - g_M$, and the blue dashed vertical lines denote peaks due to the g_V mode and nonlinear couplings generated by that mode.

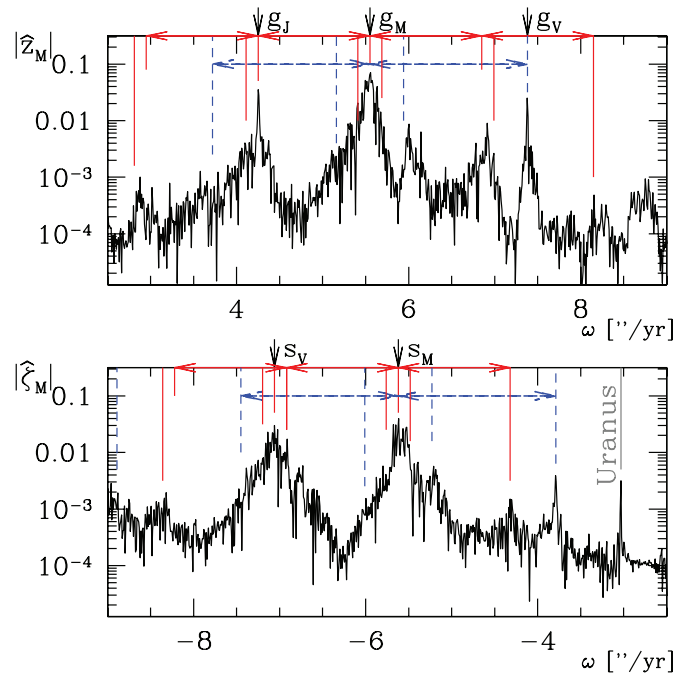(A color version of this figure is available in the online journal.)

Figure 15. Same as Figures 13 and 14, but planets are initialized with their true values ($\kappa_{\rm nbody} = 1$): the motion is significantly more chaotic, and the peaks are less easily identifiable. Nonetheless, we conclude that the primary drivers of Mercury's chaos are Jupiter's eccentricity mode and Venus's inclination mode, with Venus's eccentricity mode playing a supporting role.

(A color version of this figure is available in the online journal.)

couplings of the forced and free peaks also largely agree. This indicates that the κ-model captures much of the nonlinearity as seen in the real Mercury.

There are, however, at least three differences of note. First, the peaks in Figure 13 are broader than those in Figure 11. This is because Figure 13 suffers from weak chaos. But it is remarkable how sharp the largest peaks are: even though the e's and i's of the solar system have only been reduced by 25%, the resulting chaos is surprisingly weak. Note that the integration intervals in the two figures are the same, $T = 4400/\gamma = 150$ Myr, and hence the finite width of the peaks in Figure 13 is not due to the finite T. A second difference between the two figures is that the $\hat{z}_M$ peak at frequency $g_M + (g_M - g_J)$ is significantly larger in the $\kappa_{\rm nbody}$ integration. That peak is so large because it is overlapped by a peak forced by Venus's *eccentricity* mode, which has precession frequency $g_V \approx 2g_M - g_J$. In other words, for $\kappa_{\rm nbody} = 0.75$, Mercury is in a secular resonance with a librating angle that corresponds to frequency $2g_M - g_V - g_J$, and this largely hides the effect of g_V in the Fourier transform of Figure 13. The third difference between the two figures is the peak in $\hat{\zeta}_M$ that is caused by Uranus's inclination mode. But this peak appears to have little dynamical consequence for Mercury.

In Figure 14, the factor multiplying the initial e's and i's has been raised to $\kappa_{\rm nbody} = 0.95$. The resulting motion is more chaotic, as the widths of the peaks are wider than before, especially for $\hat{\zeta}_M$. In addition, the frequencies have been shifted sufficiently to break Mercury from the $2g_M - g_V - g_J$ resonance, and the effect of Venus's eccentricity forcing is distinct. Even though the motion is more chaotic, the principal forcing peaks and their harmonics are still identifiable.

Figure 15 shows the Fourier transform of the real Mercury ($\kappa_{\rm nbody} = 1$). Even though the initial e's and i's have been increased by only 5% relative to Figure 14, the motion is sig-

nificantly more chaotic, and the nonlinearly generated peaks are less easily identifiable, especially those near Mercury's free frequencies g_M and s_M. Nonetheless, we conclude from the progression of Figures 13–15 that the κ-model captures much of the physics. In particular, two modes—the Jupiter eccentricity mode and the Venus inclination mode—are primarily responsible for driving Mercury's chaos. The most important element lacking from the κ-model appears to be the extra forcing by the Venus eccentricity mode. The fact that the κ-model becomes chaotic at a higher threshold than the real Mercury ($\kappa \sim 1.3$ versus $\kappa_{\rm nbody} \sim 1$) is likely partly due to that extra forcing (i.e., extra forcing makes the real Mercury more chaotic). An additional contributor to the discrepancy between the two critical κ's is that the simple κ-model does not accurately capture nonlinear frequency shifts, while the precise values of the frequencies are important for where the resonances overlap. Despite this, the difference between the critical κ's is not large, and this lends support to our claimed origin for Mercury's chaos.

We note parenthetically that while we only focus on a narrow range of frequencies in Figure 15, Mercury also has peaks at $|\omega| \sim 20'' \, {\rm yr}^{-1}$, due to forcing by Earth and Mars. However, these peaks have amplitudes $\lesssim 10^{-3}$, and appear to have little influence on Mercury's chaotic motion. Had they been important, one would have expected to see their influence in Figure 13, whereas all the main peaks in that figure have already been identified.

5.2. Resonant Angles

Figure 16 shows results from a 2 Gyr SWIFT integration of the full solar system (with no reduction of the initial e's and i's). The black curve in the top panel is Mercury's total $|z_M|$, which is very nearly equal to its total eccentricity (Equation (8)), and illustrates the chaotic behavior of Mercury's orbit. The
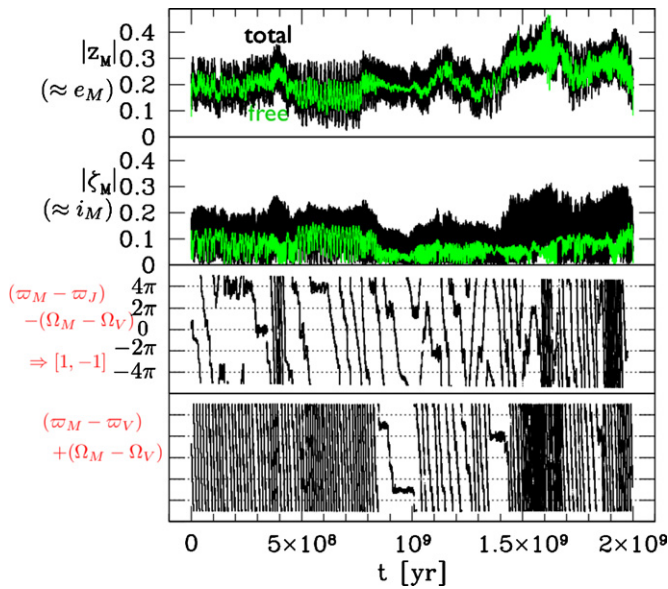
Figure 16. Mercury in an *N*-body simulation of the solar system: The top panel shows Mercury's $|z_M|$ (approximately its eccentricity) as a black curve, for the duration of a 2 Gyr SWIFT simulation. The overplotted green curve is Mercury's free $|z_M|$. The second panel shows the same, but for ζ_M (approximately its inclination). The bottom two panels show the two four-angle combinations involving Mercury that were found to undergo libration episodes over the course of this simulation. The plotted angles are the phases of the *free* orbital elements (see the main text). The angles' transitions between libration and circulation are reflected in the behavior of e_M and i_M.

(A color version of this figure is available in the online journal.)
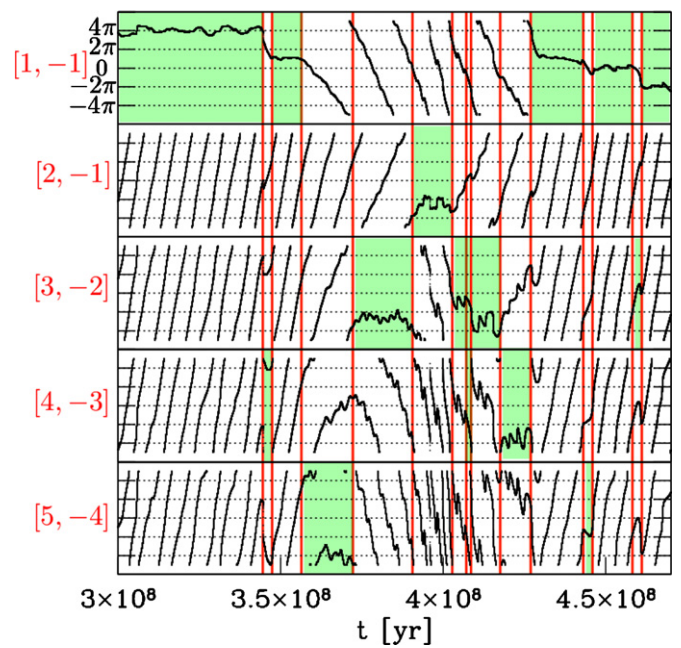

Figure 17. Chain of librating angles for the real Mercury: this is the same *N*-body simulation as in Figure 16, focusing on the time when the $[1, -1]$ angle transitions to its first extended period of circulation and then back to libration. Each panel shows the angle that has frequency $[m, n] = m(g_M - g_J) + n(s_M - s_V)$, for various values of $[m, n]$. The angles alternately librate, showing that the chaos is at least partly caused by the overlap of these resonances, as in the κ-model (Figure 9). The plotted angles are the phases of the free orbital elements.

(A color version of this figure is available in the online journal.)

overplotted green curve is the absolute value of Mercury's free z_M, which we define to be the part of its total z_M that comes from the main peak in Figure 15, i.e., we first take the Fourier transform of z_M, then set to zero all frequencies except those satisfying $4\farcs9\ \mathrm{yr}^{-1} < \omega < 6\farcs5\ \mathrm{yr}^{-1}$, and then take the inverse Fourier transform. By plotting the free z_M, the short-term variations are reduced, and long-term diffusion is clearer. The second panel in Figure 16 is the same as the top but for ζ_M; for the free ζ_M, we filter out frequencies outside of the domain $-6\farcs3\ \mathrm{yr}^{-1} < \omega < -4\farcs7\ \mathrm{yr}^{-1}$.

The bottom two panels of Figure 16 show the two four-angle combinations involving Mercury that were found to undergo libration episodes. The third panel shows the angle $(\varpi_M - \varpi_J) - (\Omega_M - \Omega_V)$, which is the angle that has frequency

$$\sigma \equiv (g_M - g_J) - (s_M - s_V), \tag{55}$$

i.e., the $[1, -1]$ angle (Equation (43)). (More precisely, we use the angles of the free elements; see below.) And the bottom panel shows the angle associated with the frequency

$$\sigma' \equiv (g_M - g_V) + (s_M - s_V). \tag{56}$$

Laskar (1992) has shown that the σ angle can change from libration to circulation. But our finding that the σ' can as well is new.[12] We note that even though both the σ and σ' angles undergo libration episodes, the four-angle combination that is the sum of the two, i.e., the angle associated with $2g_M - g_V - g_J$ does not. That angle was found to librate in the $\kappa_{\mathrm{nbody}} = 0.75$ simulation (Figure 13), and we suspect that it will eventually librate in a long enough integration of the full solar system.

---

[12] Since the motion is chaotic, it is possible that σ' did not librate at all in the simulations of Laskar (1992) and Sussman & Wisdom (1992).

The angles displayed in the bottom panels of Figure 16 were those of the *free* elements. For example, for ϖ_M we first filtered the Fourier transform of z_M as described above, and took the phase of the free part of z_M. This filtering procedure is especially important for Ω_V because Venus's ζ_V variations are dominated by forcings due to other modes, including the Mercury-, Earth-, and Mars-dominated modes (Laskar 1990), whereas we wish Ω_V to denote the phase of the Venus-dominated mode. Therefore, we first filter the Fourier transform of ζ_V, keeping only frequencies $-7\farcs7\ \mathrm{yr}^{-1} < \omega < -6\farcs3\ \mathrm{yr}^{-1}$, and use for Ω_V the phase of the filtered ζ_V. Similarly, we filter z_M, ζ_M, z_V, and z_J with appropriate windows to obtain the other angles of interest. Our method of filtering for extracting mode angles differs from that of Laskar (1990), who extracts mode angles by projecting onto the numerically determined nonlinear "proper modes." We have experimented with a number of different methods, and also with changing the size of the filter window, and the duration of the integration, and found that our filtering method is simple to implement, is computationally efficient, and gives reliable results.

In addition to the two four-angle combinations of Figure 16, one might suspect that there are many more higher order combinations that librate when both of those angles simultaneously circulate, as in the Hamiltonian model. In Figure 17, we zoom into the episode when the $[1, -1]$ angle first undergoes an extended period of circulation, and plot some higher order combinations associated with the frequencies $[m, n] \equiv m(g_M - g_J) + n(s_M - s_V)$. It can be seen that these angles librate in turn, just as in the Hamiltonian model (compare with Figure 9). This provides another demonstration that the physics of the Hamiltonian model is similar to that of the real Mercury.

6. SUMMARY AND DISCUSSION

The general picture for how secular chaos occurs is simple. For secularly interacting planets, the secular frequencies are constant to linear order in e and i. But nonlinearities can shift planets into and out of secular resonance with each other, and when two resonances overlap, chaos results. However, despite the importance of secular chaos in the solar system, and potentially in extra-solar systems as well, a nonlinear secular theory has not yet been attempted. In this work we set out to develop such a theory and we accomplish that goal in a number of steps.

In Sections 2–4, we focus on the dynamics of a test particle in the presence of multiple massive planets. The test particle is evolved to leading nonlinear order, and the e's, i's, and precession rates of the massive planets (or more properly of the planet modes) are taken to be constant. We first consider the simple case of coplanar perturbers, as worked out by Sidlichovsky (1990), and find the following basic results.

1. A single eccentric precessing perturber (with frequency g_J) does not lead to chaos.[13] The Hamiltonian resembles the one that describes a first-order MMR. The precession frequency of the test particle (with frequency γ in linear theory) is modified nonlinearly, and when it coincides with the perturber's precession frequency, a nonlinear secular resonance occurs: the center is at $p_{e*} = 2\Delta$ (Equation (25)) and the width is $\delta p_e = 4\left(\epsilon_J\sqrt{p_{e*}}\right)^{1/2} = 4\epsilon_J^{1/2}(2\Delta)^{1/4}$ (Equation (26)).

2. Two eccentric, precessing perturbers produce chaos when their respective resonance islands overlap (Chirikov 1979). One can write down the overlap criterion using the above information on centroid and width (Sidlichovsky 1990).

When both eccentricities and inclinations are present, the dynamics are more complex, and nonlinear effects become important at much lower amplitudes. Such a general case has not been studied analytically before. In Section 4, we consider the case when a test particle comes under the influence of an eccentric, inclined perturber that precesses (with frequencies g_J and s_J for the apsidal and nodal precessions, respectively). Our principal findings are as follows.

1. The particle's free precession frequencies (apsidal g and nodal s) are both altered by its eccentricity and inclination. Therefore, in the particle's e–i plane, each resonance traces out a curve. (Equivalently, it traces a line in the p_e–p_i plane, where $p_e \approx e^2$ and $p_i \approx i^2$.)

2. Because of this behavior, a particularly illuminating way to map out the nonlinear dynamics is the use of a novel "mean momentum map" (MMM), in which each trajectory is represented by a point on the map, its location being the time-averaged momenta (p_e and p_i). In the MMM, regular and chaotic trajectories are clearly distinguished, individual secular resonances are identified (even high-order ones), and their widths easily measurable. This map also graphically illustrates how the overlap of secular resonances leads to chaos (Figures 3–5). The MMM is similar to the well-known FMA (Laskar 1990), but with the axes representing time-averaged actions rather than average angular frequencies.

3. Secular resonances occur where $m(g - g_J) + n(s - s_J) = 0$, for integers m, n (Equation (43)). We derive the locations of all such $[m, n]$ resonances in the p_e–p_i plane (Equations (37), (39), (42)–(45)), as well as the widths of the four strongest resonances: [1, 0] (eccentricity only), [0, 1] (inclination only), and [1, ±1] (Equations (38), (40), and (46)). We show that our analytical results agree remarkably well with the numerical MMM results (Figures 3–7).

4. The MMM allows us to identify the resonances responsible for chaos. For parameters relevant for Mercury, we find overlap between $[m, m \pm 1]$ resonances (for small integer m's) account for the observed chaos (Figures 3–5). We also confirm this using the more conventional approaches of examining surfaces of section (Figure 8) and explicitly tracing the chain of librating angles (Figure 9).

5. When e_J and i_J are only of order a few percent, a test particle on a Mercury-like trajectory becomes chaotic. Furthermore, the area of the chaotic region expands rapidly with increasing e_J and i_J, as we show both numerically and analytically (Equations (49)–(51)).

In Section 5, we apply our secular theory for a test particle to understand the orbital evolution of Mercury in N-body simulations. We show that despite all the simplifications we made in the Hamiltonian model of Section 4 (one forcing body, lowest order nonlinearity, no feedback), the real Mercury behaves in a similar manner. To show this, we must abandon the MMM because the solar system has so many degrees of freedom that the resonances are washed out in the real Mercury's MMM. Instead, we demonstrate the dynamics using the Fourier transforms (Figures 11–15), as well as by tracking the resonant angles directly (Figures 16 and 17). For the real Mercury, we find the following.

1. *Culpable resonances.* Mercury's chaos is primarily driven by the s_V and g_J modes (i.e., the Venus- and Jupiter-dominated i and e secular modes), although the g_V mode also plays a role. Similar to the test particle case, we find that a slew of resonances involving those modes and Mercury's own free modes (with frequencies g_M and s_M) are important; in particular, the relevant frequency combinations are of the form $m(g_M - g_J) + n(s_M - s_V)$, where $n = m \pm 1$ and m is a small integer (Figures 13–15). The corresponding angles sequentially librate (Figure 17). We identify a new resonance $\sigma' = (g_M - g_V) + (s_M - s_V)$ (Equation (56)) that can also undergo libration episodes (Figure 16) and may contribute to Mercury's chaos. Whether it is an unnecessary coincidence for Mercury's chaos is unknown. Figures 13–15 show that the aforementioned nonlinear secular resonances account for virtually all of the features seen in Mercury's Fourier transform. Furthermore, Mercury's Fourier transform agrees with the Fourier transform of our simplified κ-model, both for real forcings and when the forcings are artificially reduced (Figures 11 and 12). This shows that our model quantitatively captures the principal effects responsible for Mercury's chaotic orbit. Other effects not included in the model, such as MMRs and other secular resonances play an unimportant role.[14] This agrees with previous findings that N-body integrations of Mercury's orbital evolution yield similar results to integrations that use the secular equations (e.g., Laskar 2008).

---

[13] At very high eccentricity and inclination, i.e., beyond the leading nonlinear order, a single perturber can lead to chaos (Naoz et al. 2011; Lithwick & Naoz 2011).

[14] We ignore throughout this paper the effect of Mercury's spin on its orbital evolution because there is negligible angular momentum in its spin relative to its orbit. Nonetheless, its chaotic orbit could have been responsible for forcing Mercury into the 3:2 spin-orbit resonance (Correia & Laskar 2004).

Table 1
Terms in Scaled Disturbing Function $\frac{8R}{3\alpha^2} = \sum_i c_i h_i$

i	1	2	3	4	11	12	13	14	15	16	17
c_i	1	$-\frac{5}{4}\alpha$	$\frac{3}{2}$	$-\frac{1}{4}$	-1	1	$\frac{3}{2}$	$\frac{1}{4}$	$-\frac{5}{8}$	$-\frac{5}{8}$	$\frac{1}{4}$
h_i	$|z|^2$	z^*z' + c.c.	$|z|^2|z'|^2$	$|z|^4$	$|\zeta|^2$	$\zeta^*\zeta'$ + c.c.	$|\zeta|^2|\zeta'|^2$	$|\zeta|^4$	$|\zeta|^2\zeta^*\zeta'$ + c.c.	$|\zeta'|^2\zeta^*\zeta'$ + c.c.	$\zeta^{*2}\zeta'^2$ + c.c.

i	21	22	23	24	25	26	27	28
c_i	-2	$-\frac{3}{2}$	$-\frac{3}{2}$	$\frac{7}{4}$	$\frac{7}{4}$	$\frac{5}{4}$	$\frac{5}{4}$	$-\frac{5}{2}$
h_i	$|z|^2|\zeta|^2$	$|z'|^2|\zeta|^2$	$|z|^2|\zeta'|^2$	$|z|^2\zeta^*\zeta'$ + c.c.	$|z'|^2\zeta^*\zeta'$ + c.c.	$z^{*2}\zeta^2$ + c.c.	$z^{*2}\zeta'^2$ + c.c.	$z^{*2}\zeta\zeta'$ + c.c.

Notes. The Hamiltonian is $H = \gamma \sum_i c_i h_i$ where γ is defined in Equation (A4). The variables z and ζ (defined in Equations (8) and (9)) are approximately the complex eccentricity and inclination. Terms 1–4 are z-only. Terms 11–17 are ζ only. Terms 21+ are mixed. c.c. denotes complex conjugate.

2. *Amplitude threshold for Mercury's chaos.* Having identified the secular resonances responsible for Mercury's chaos, we can calculate the threshold for chaos. We do that here in an approximate way. First, since Mercury's apsidal and nodal frequencies differ from g_J and s_V by $\sim 25\%$ (i.e., $\Delta \sim \Delta_s \sim 0.25$, Equations (19) and (36)), the location in the momenta plane (p_e–p_i) where all $[m, n]$ resonances intersect are around half that, or $p_{e**} \sim p_{i**} \sim 0.12$ (Equation (45)). These values are determined by the linear precession frequencies of the system. Second, given the values of Δ and Δ_s, the $[1,-1]$ resonance reaches closest to the origin.[15] The width of the chaotic overlap zone between the $[1, -1]$ and the $[1, 0]$ (or the $[0, 1]$) resonances is $\sim 2(\epsilon_J i_J)^{1/4}(p_{e**}p_{i**})^{1/8}$ (Equation (51)).[16] As a result, when $\epsilon_J \sim i_J \gtrsim p_{e**}^{3/2}/4 \sim 0.01$, the region of chaos will encroach upon the origin in the p_e–p_i plane and even trajectories that are initially circular/coplanar will be chaotic. This quantitatively explains why Mercury can be chaotic even though the eccentricities and inclinations in the solar system are at the level of a few percent. Alternatively, our analytical theory alone would have allowed us to predict that Mercury's orbit is chaotic, were we to live in a world without digital computers.

3. *Mercury's precarious state.* We discover that Mercury is perched on the threshold of chaos. Although the diffusion of Mercury's orbit takes place on a timescale comparable to the lifetime of the solar system, when we reduce the e's and i's of all the planets by only 25%, Mercury's motion becomes nearly regular (Figure 13). By contrast, when we increase the planets' e's and i's by 20% we observe violent instability, with Mercury ejected in ~ 100 Myr. This behavior is also apparent in our test particle model (Figure 12) and is explained by our results on the location and extent of the chaotic zone (Equations (45), (49)–(51)). This observation, however, raises an interesting question as to what puts the solar system in such a delicate balance, with the instability time comparable to the system lifetime.

With the origin of Mercury's chaos elucidated, we look ahead to directions along which our work can be extended and areas where it can be applied. The weakly nonlinear secular theory developed here can be extended to order-unity eccentricities and inclinations. Although that case will be more complicated, we suspect that the basic structure will remain, with resonant zones in the e–i plane whose overlap leads to chaos. Another important direction is to incorporate MMRs, which may be

common among planetary systems and can interact with secular chaos in interesting and yet unsuspected ways.

As mentioned above, Mercury, the most unstable planet in our system, appears to be perched at the threshold for chaos. This is certainly a clue for understanding how the solar system arrived at its current marginally stable state. But much more work is needed.

Our nonlinear secular theory can also be applied to Earth and Mars, which also have chaotic orbits and for which librating angles have been identified that are similar to those we found for Mercury, i.e., angles of the form $m(\varpi_{\rm mars} - \varpi_{\rm earth}) + n(\Omega_{\rm mars} - \Omega_{\rm earth})$, with $[m, n] = [1, -1], [2, -1]$, and $[3, -2]$ (Laskar 1992; Sussman & Wisdom 1992). We now have the right tools to study in detail the origin of chaos in these two planets. We have also proposed that secular chaos can play a role in shaping extra-solar planetary systems (Wu & Lithwick 2011), and hence the theory of secular chaos might be applicable to extra-solar planets as well.

APPENDIX A

FOURTH-ORDER SECULAR HAMILTONIAN

In this appendix, we give the expression for the secular Hamiltonian of a test particle perturbed by an external planet, where both particle and planet are orbiting a star. The Hamiltonian is expanded to fourth order in the particle's eccentricity and inclination, and to leading order in the ratio of semimajor axes. The energy per unit mass of the test particle is

$$E = -\frac{GM_\odot}{2a} - \frac{Gm'}{a'}R, \tag{A1}$$

where $M_\odot$ is the mass of the star, a and a' are, respectively, the test particle's and planet's semimajor axes, m' is the planet's mass, and R is the disturbing function. We approximate R by only retaining the secular terms up to fourth order in e and $s \equiv \sin(i/2)$ and second order in $\alpha \equiv a/a'$ (except for the f_{10} term whose leading contribution is $O(\alpha^3 e^2)$):

$$\begin{aligned} R \approx\ & f_2 e^2 + f_3 s^2 + f_5 e^2 e'^2 + f_7(e^2 s^2 + e^2 s'^2 + e'^2 s^2) \\ & + f_8 s^4 + f_9 s^2 s'^2 + f_{10} ee' \cos(\varpi - \varpi') \\ & + \big(f_{14} ss' + f_{15} ss'(e^2 + e'^2) + f_{16} ss'(s^2 + s'^2)\big) \cos(\Omega - \Omega') \\ & + f_{18} e^2 s^2 \cos(2\varpi - 2\Omega) + f_{21} e^2 ss' \cos(2\varpi - \Omega' - \Omega) \\ & + f_{18} e^2 s'^2 \cos(2\varpi - 2\Omega') + f_{26} s^2 s'^2 \cos(2\Omega - 2\Omega'), \end{aligned} \tag{A2}$$

in the notation of the Appendix of Murray & Dermott (2000). The f_i are functions of α that may be expressed as sums of Laplace coefficients and their derivatives. We drop terms that are independent of the test particle's orbital elements.

---

[15] Other $[m, n]$ resonances may be important in different systems. For instance, the $[1,1]$ resonance will dominate if $\Delta \approx -\Delta_s$.

[16] We include here an extra factor of two to account for the difference between the half- and full-width, as described in Section 4.3.

In this paper, we work with a scaled Hamiltonian, $H \equiv -2E/\sqrt{GM_\odot a}$ (Equation (3)), and hence

$$H = \gamma \frac{8R}{3\alpha^2}, \qquad (A3)$$

dropping the Keplerian term in E because it is irrelevant for secular dynamics, and defining

$$\gamma \equiv \frac{3}{4} \frac{m'}{M_\odot} \alpha^3 \left(\frac{GM_\odot}{a^3}\right)^{1/2}, \qquad (A4)$$

which is the test particle's secular free precession frequency based on linear theory. The scaled disturbing function $8R/(3\alpha^2)$ is a sum of terms that are listed in Table 1, after expanding the f_i to $O(\alpha^2)$, and f_{10} to $O(\alpha^3)$.

APPENDIX B

WIDTH OF THE [1, −1] AND [1, 1] RESONANCES FROM VON ZEIPEL TRANSFORMATION

We start from Hamiltonian (35), which we reproduce here as

$$H(p_e, q_e; p_i, q_i) = -\frac{p_e^2 - p_i^2}{4} + \Delta p_e + \Delta_s p_i - 2 p_e p_i$$
$$- 2\epsilon_J \sqrt{p_e} \cos q_e + 2 i_J \sqrt{p_i} \cos q_i, \quad (B1)$$

setting $\gamma = 1$, $q_e \equiv \varpi_- = \varpi - g_J t$, and $q_i \equiv \Omega_- = \Omega - s_J t$. We solve this Hamiltonian perturbatively, treating ϵ_J and i_J as the small parameters. This is equivalent to expanding in the test particle's forced eccentricity and inclination, assumed to be much smaller than the free e and i. We transform to capitalized variables with the von Zeipel generating function

$$F(P_e, q_e; P_i, q_i) = P_e q_e + P_i q_i + k_e(P_e, P_i) \sin q_e$$
$$+ k_i(P_e, P_i) \sin q_i, \qquad (B2)$$

where the first two terms generate the identity transformation, and the functions k_e and k_i are first order in ϵ_J and i_J; their form will be chosen to "kill" the cosine terms in the Hamiltonian to leading order. The von Zeipel generating function transforms variables as follows:

$$p_e = P_e + k_e \cos q_e \qquad (B3)$$

$$p_i = P_i + k_i \cos q_i \qquad (B4)$$

$$Q_e = q_e + \partial_{P_e} k_e \sin q_e \qquad (B5)$$

$$Q_i = q_i + \partial_{P_i} k_i \sin q_i. \qquad (B6)$$

Inserting into the Hamiltonian yields

$$H(P_e, Q_e; P_i, Q_i) = -\frac{P_e^2 - P_i^2}{4} + \Delta P_e + \Delta_s P_i$$
$$- 2 P_e P_i - k_e k_i \cos(Q_e - Q_i)$$
$$- k_e k_i \cos(Q_e + Q_i) \qquad (B7)$$

to second order, after setting

$$k_e = \frac{2\epsilon_J \sqrt{P_e}}{\Delta - P_e/2 - 2P_i} \qquad (B8)$$

$$k_i = \frac{-2 i_J \sqrt{P_i}}{\Delta_s + P_i/2 - 2P_e}, \qquad (B9)$$

to eliminate the first-order terms. The two cosine terms in this Hamiltonian are the [1, −1] and [1, 1] resonances, respectively (see Equations (41) and (42) and following). Note that we have dropped second-order terms in the Hamiltonian that are proportional to $\cos^2 Q_e$, $\sin^2 Q_e$, $\cos^2 Q_i$, and $\sin^2 Q_i$ because these have little effect on the [1, 1] and [1, −1] resonances; but they generate new frequency components which are important for higher order resonances.

To leading order, k_e is twice the product of the free eccentricity ($\sqrt{P_e}$) with the forced eccentricity, where the forced eccentricity differs from the linear expression (ϵ_J/Δ; see Equation (18)) by the terms $P_e/2 + 2P_i$ in the denominator, which arise from the nonlinear shift of the frequency (Equations (41) and (42)). Similarly k_i is twice the product of the free and forced inclinations. Hence, the strengths of the [1, −1] and [1, 1] resonances are proportional to the products of the free and forced eccentricities and inclinations, as argued qualitatively in Section 4.

To determine the width of the [1, −1] resonance, we drop the last cosine term in the above Hamiltonian. Since $P_+ \equiv P_e + P_i$ is an integral of the motion, we may re-write the Hamiltonian as $H(P_e, Q_-) = 2(P_e - P_*)^2 - k_e k_i \cos Q_-$, dropping a constant and defining $Q_- \equiv Q_e - Q_i$ and $P_* \equiv (5/8)P_+ - (1/4)(\Delta_g - \Delta_s)$. Therefore, the half-width of the resonance is

$$\delta P_e = \sqrt{|k_e k_i|}. \qquad (B10)$$

We take the amplitude of the cosine term to be fixed at its value at resonance center (e.g., Chirikov 1979). Since $P_e + P_i$ is constant, the half-width in P_i is the same, $\delta P_i = \sqrt{|k_e k_i|}$.

REFERENCES

Chirikov, B. V. 1979, Phys. Rep., 52, 263
Correia, A. C. M., & Laskar, J. 2004, Nature, 429, 848
Henrard, J., & Lemaitre, A. 1983, Celest. Mech., 30, 197
Laskar, J. 1989, Nature, 338, 237
Laskar, J. 1990, Icarus, 88, 266
Laskar, J. 1992, in IAU Symp. 152, Chaos, Resonance, and Collective Dynamical Phenomena in the Solar System, ed. S. Ferraz-Mello (Cambridge: Cambridge Univ. Press), 1
Laskar, J. 1993, Celest. Mech. Dyn. Astron., 56, 191
Laskar, J. 1996, Celest. Mech. Dyn. Astron., 64, 115
Laskar, J. 2008, Icarus, 196, 1
Laskar, J., & Gastineau, M. 2009, Nature, 459, 817
Lecar, M., Franklin, F. A., Holman, M. J., & Murray, N. J. 2001, ARA&A, 39, 581
Levison, H. F., & Duncan, M. J. 1994, Icarus, 108, 18
Lissauer, J. J., Fabrycky, D. C., Ford, E. B., et al. 2011, Nature, 470, 53
Lithwick, Y., & Naoz, S. 2011, ApJ, submitted (arXiv:1106.3329)
Mudryk, L. R., & Wu, Y. 2006, ApJ, 639, 423
Murray, C. D., & Dermott, S. F. 2000, Solar System Dynamics (Cambridge: Cambridge Univ. Press)
Murray, N., & Holman, M. 1999, Science, 283, 1877
Naoz, S., Farr, W. M., Lithwick, Y., Rasio, F. A., & Teyssandier, J. 2011, Nature, 473, 187
Ogilvie, G. I. 2007, MNRAS, 374, 131
Quinn, T. R., Tremaine, S., & Duncan, M. 1991, AJ, 101, 2287
Sidlichovsky, M. 1990, Celest. Mech. Dyn. Astron., 49, 177
Sussman, G. J., & Wisdom, J. 1988, Science, 241, 433
Sussman, G. J., & Wisdom, J. 1992, Science, 257, 56
Wisdom, J. 1983, Icarus, 56, 51
Wisdom, J., & Holman, M. 1991, AJ, 102, 1528
Wu, Y., & Lithwick, Y. 2011, ApJ, 735, 109